

Semantic Segmentation of Pancreas in Computed Tomography Images using Convolutional Neural Network

Geethanjali T M¹, Dr. Minavathi² and Dr. Dinesh M S³

¹PES College of Engineering/Information Science Engineering Department, Mandya, India

²PES College of Engineering/Computer Science Engineering Department, Mandya, India

³PES College of Engineering/PET Research Center, Mandya, India

Email: geethanjalitm@pesce.ac.in, minavathi@pesce.ac.in, dineshmys@gmail.com

Abstract— Diagnosis of pancreatic cancer is challenging due to complex background and fade boundaries in Computed Tomography (CT) images. Radiologists segment the pancreatic area manually by using semantics, which takes more time. This research aims to achieve a desirable performance on pancreatic segmentation with less computational power and effort through semantic segmentation using convolutional neural network. The proposed DeepLab-ResNet-SE-Net (DRSE-Net) is evaluated on National institutes of Health (NIH) pancreas CT dataset combining DeepLabV3 with ResNet101 and Squeeze-Excitation Blocks. Accordingly, obtained average Dice-Score, Jaccard Index, Precision and Recall are 0.8890, 0.7996, 0.8871, 0.8926 respectively compared to state of art approaches by reducing the parameters of the proposed model.

Index Terms— Semantic Segmentation, Computed Tomography (CT), Pancreas and Convolutional Neural Networks.

I. INTRODUCTION

One of the rarest cancer considering about 7% of all cancer deaths in U S is a pancreatic cancer with high mortality rate [1]. Pancreatic cancer during its early stages will be asymptomatic in nature, therefore it is very difficult to diagnose. CT images plays a very important role to capture detailed pancreas which in turn enhances the probability of radiologists detecting the tumors in the pancreatic organ. Diseases are being diagnosed by the radiologists and scientists by the help of medical images [2].

Few researchers have revealed the significance of semantic segmentation of abdominal organs considering CT images [3, 4]. Some authors have also focused on semantic segmentation taking into account healthy pancreas [5, 6]. Clinicians cannot manually perform semantic segmentation of CT/MRI images due to lot of time consumption. Pancreas segmentation is challenging because of huge variations in shape, its position along with complex background and fade boundaries in images.

Few contributions of the study are summarized as follows:

- Firstly, data alumentation (Horizontal flip, vertical flip, Rotate (limit=15, 30, 45, 60) are carried on NIH dataset.

- The DeeplabV3Plus ResNet-101 architecture includes Squeeze and Excitation block which enhances the performance metrics like Dice-Score, Jaccard-Index, Recall and precision used to segment pancreas.
- Model parameters (9.8 M) are less when compared with state-of-the art method.
- Prediction time taken by the proposed method is 64.81seconds.

Firstly, recent reviews are discussed in section 2. Section 3 introduces dataset preparation, proposed methodology, performance measures and loss function used to segment pancreas organ. Results are being rendered with their related work in section 4. Finally, conclusion is noted in section 5.

II. RELATED WORKS

Mei-Ling Huang et.al [7] designed MBU-Net model which is a combination of U-Net and MobileNet-V2 networks to segment pancreas and achieved the dice-coefficient of 82.87% and jaccard similarity coefficient of 70.97%. MBU-Net decreases the computational parameters, but in turn not all the measures used for performance were best compared to the related works. Deep Convolutional neural network model designed by Dingwen Zhang et.al [8] to segment pancreas consisting of localization sub-network (Lenet) and segmentation sub-network (senet). Lenet is a light weighted fully convolutional network which adds prior propagation module (PPM) which denotes pixel value probability from its context. Senet is also a light weighted encoder decoder architecture with PPM and Scale transferable feature fusion module (STFFM) obtaining greater level of feature fusion. Lenet+Senet+PPM has obtained the mean dice score of 84.90% and has lost some spatial information to explore from 3D images.

Location to segmentation strategy was followed by yan Wang et.al [9] to segment pancreas using view adapting 3DU-Net (VA-3DUNet) with average dice similarity co efficient of 86.19%. Author uses two 3DU-Net, one to identify the volume of interest another 3DU-Net which is been trained in a view adaptive way to segment pancreas. Accurate segmentation of pancreas is carried out by Meixiang Huang et.al [10] using deformable convolutional module to enhance feature representation in U-net architecture (DU-Net). They have carried experiment using NIH dataset and achieved average DSC of 0.8725. A combination of deep U-net, Mixup and random image cropping & patching (RICAP) proposed by Mizuho Nishio et. al [11] used for pancreas segmentation using NIH database has secured DSC of 0.789 ± 0.083 and Jaccard index of 0.658 ± 0.103 respectively.

The aim of this study is to evaluate and validate the effectiveness of DRSE-Net model with respect to computational parameters and performance measures.

III. MATERIALS AND METHODS

A. Dataset

For segmenting pancreas, the study utilized 82 contrast enhanced CT scans of the abdomen from the NIH clinical centre. Each CT scan was evaluated by a radiologist for the annotation of pancreatic regions. The NIH dataset details are provided in Table I. CT images in Neuroimaging Informatics Technology Initiative (NIFTI) file format is considered in the work. All CT volumes in the dataset are of dimension 512 x 512 pixels, number of slices varying between 181-466 and slice thickness varying from 1.5mm-2.5mm. To achieve fairness in the comparison between different methods, CT scans that contained pancreas were the only ones used. For training 80% of the dataset was used remaining were used equally for validating and testing the model.

TABLE I. NIHPANCREAS-CT DATASET

Organization	National Institute of Health(NIH)
Modality	Computed Tomography
File Format	NIFTI
No of Volumes	82
Dimension	512 x 512

B. Image Preparation

Pancreas occupies a smallest region in the complete abdomen CT image and possesses a multiple variation in shape and angles as illustrated in Fig.3.1. An example of CT image slice with its annotated mask is represented in Fig.3.1 (a) and Fig.3.1 (b) respectively. Boundaries are blurred between pancreas and other tissues making the segmentation task more difficult. To train the model effectively only the slices with pancreas were selected for experimentation. The number of image slices and annotated masks also known as ground truths were finally reduced to 7,054. These slices and their ground truths were further converted into portable network graphics (PNG) file format. To overcome the issues of training time and overloading the memory, size of the image slices and masks were reduced from 512 X 512 to 256 X 256.

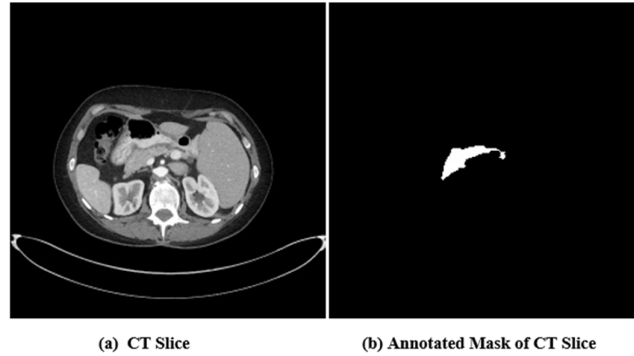


Figure 3.1: Example of Pancreatic slice

Data albumentation is used in the study to boost the performance of the model which consists of many image transform operations that are been used to optimize the performance. Albumentation works fine for medical images and can be used in many tasks namely classification, semantic segmentation, instance segmentation and object detection. Horizontal Flip, Vertical flip and rotation are the operations used in the study. Fig.3.2 displays the different data albumentation operations performed only on training set.

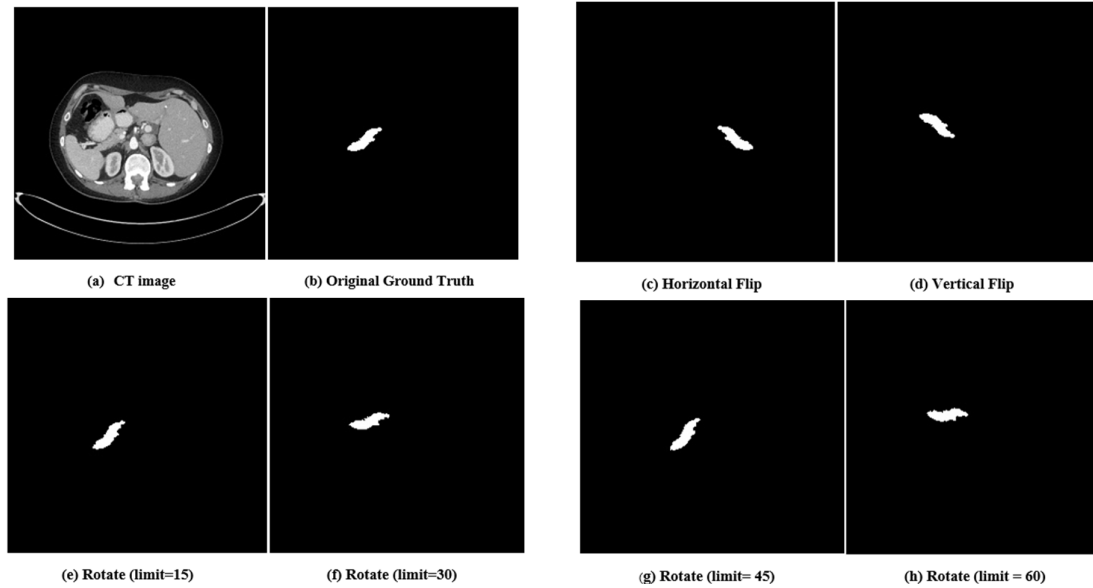


Figure3:2 outcomes of Data albumentation

C. Proposed Model

The proposed DeepLabv3-ResNet-SE-Net (DRSE-Net) represented in Fig.3.3 is a convolutional neural network consisting of encoder and decoder parts. DeeplabV3[14,15] is a semantic segmentation [19,20,21,22,23,24] encoder-decoder architecture considering spatial pyramid pooling[16,17,18] used to capture the image

information at different scales. The DeeplabV3 also uses Astrous convolution which is a dilated convolution used to get the image information from larger field of view. ResNets [25] promotes residual learning which enhances the training and yields better performance in image recognition task. ResNet-101 is used as a part of encoder, which is being trained on ImageNet weights. Encoder part of the model takes the output of ResNet-101 and passes through Astrous Spatial Pyramid Pooling (ASPP). ASPP consists of 5 blocks. First block is image pooling block depicted in Fig.3.4 (b), second block consists of 1x1 convolution, and rest of the three blocks are dilated convolution with the rate of 6, 18, and 24. All the five feature maps are concatenated and 1x1 convolution with dilation rate 1 is applied. The output of the encoder is taken into the decoder part and it is up sampled by 4 and concatenated with low level features of ResNet-101. Later 3x3 Convolution with up sampling by 4 is performed to obtain the output mask. Squeeze-Excitation (SE) [26] block represented in Fig.3.4 (a) is being included to enhance the performance of the model.

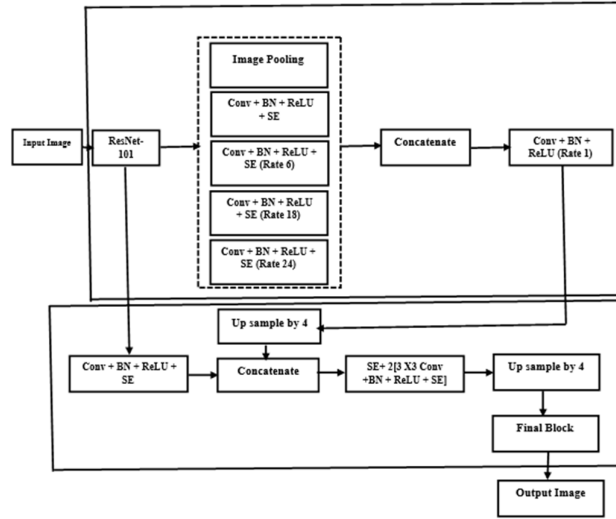


Figure3.3: proposed DRSE-Net Model

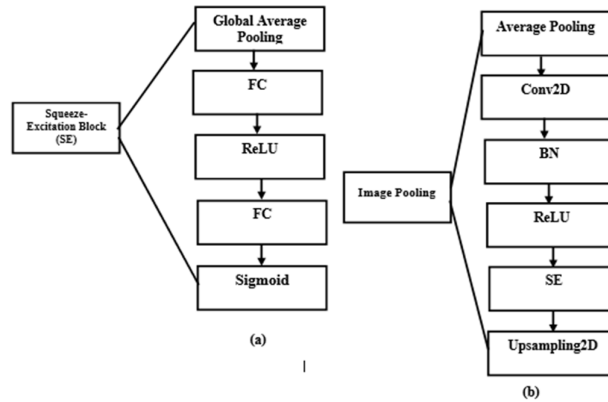


Figure 3.4: SE and Image pooling Block

D. Loss Function

To address the issue of class imbalance problem in semantic segmentation of pancreas organ, dice loss is being selected. Dice loss [12, 13] tries to address the imbalance problem which occurs between foreground and background of the images. The dice loss is represented in Eq.1 where y , p represents ground truth and predicted labels respectively.

$$Dice\ Loss(y, p) = 1 - \frac{2yp + 1}{y + p + 1} \quad (1)$$

E. Performance Evaluation Metrics

The four essential metrics in Eq.2, Eq.3, Eq.4, Eq.5 used to know the performance of the proposed DRSE-Net model are Dice-score, Jaccard-Index (Intersection over Union), Precision and Recall.

- (1) Dice – Score: Is one of the basic and most important metric used for semantic segmentation which measures the overlapping between the ground truth and predicted mask.

$$Dice - Score = \frac{2 * true_{positives}}{2 * true_{positives} + false_{negatives} + false_{positives}} \quad (2)$$

- (2) Jaccard-Index: Metric used majorly for semantic segmentation which is used to measure the number of pixels which are in common between the ground truth and predicted mask divided by the total number of pixels which are present.

$$Jaccard - index = \frac{true_{positives}}{(true_{positives} + false_{positives} + false_{negatives})} \quad (3)$$

- (3) Precision: Metric which is effectively used to describe true positives which are relative to the ground truth.

$$Precision = \frac{true_{positives}}{(true_{positives} + false_{positives})} \quad (4)$$

- (4) Recall: This is another metric which is prominently used to describe the completeness of true positives which are relative to ground truth.

$$Recall = \frac{true_{positives}}{(true_{positives} + false_{negatives})} \quad (5)$$

IV. RESULTS AND DISCUSSIONS

In this section the proposed DRSE-Net model is been trained on the original NIH dataset after performing data albumentation. The hyper parameters used in training the model are learning rate (1e-4) and batch size of 8. Firstly, we compare different state of art methods [7-11] which are worked using the same dataset. Later comparison of proposed method both using SE block and without SE block. Finally, DRSE-Net Model with SE Block achieved better results compared with DRSE-Net Model without SE block and other methods in literatures by not increasing much of trainable parameter. Table II signifies the quantitative comparison results of the proposed model with other state of art studies carried using NIH dataset.

TABLE II. COMPARISON TO THE STATE OF THE ART STUDIES ON THE NIH DATASET

Literatures	Method	Dice-Score	Jaccard-Index	Precision	Recall	Model parameters
Mei-Ling Huang et al.(2022) [7]	MBU-Net	82.87 ± 1.00	70.97% ± 1.00	89.29 ± 0.98	77.37 ± 1.41	6.30 M
Dingwen Zhang et al.(2021) [8]	DCNN based FCNN	84.90				25.13 M
Van Wang et al. (2021) [9]	VA 3DU-Net	86.19				
Meixiang Huang et al.(2021) [10]	DU-Net	87.25				
Mizuho Nishio et al. (2020) [11]	U-Net + Mixup + RICAP	78.9 ± 8.3	65.8 ± 10.3			
This Study (without SE Block)	DRSE-Net	88.04	78.65	88.01	88.38	9.7M
This Study (with SE Block)	DRSE-Net	88.90	79.96	88.71	89.26	9.8M

Fig.4.1 qualitatively illustrates three randomly selected CT slices by combining Squeeze and Excitation (SE) block into Deeplab v3 + ResNet101 to improve the performance measures. We observe that the proposed method is better able to extract information about the pancreas than related models. A clear visual inspection of the results reveals that the model proposed significantly improves pancreas Dice score and Jaccard-Index when compared to state of the art related models. Fig 4.1 (a), (b) and (c) indicates the CT image slice, ground truth and predicted masks obtained by the proposed model with SE block. Values below the picture represents the predicted mask dice-score with Jaccard index scores for individual slices.

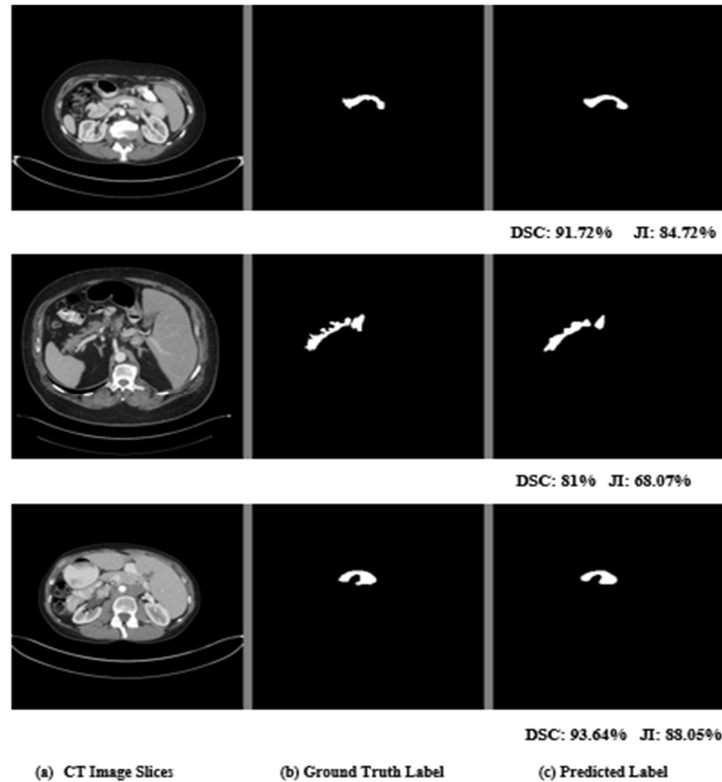


Figure 4.1: Visualization of 2D pancreas segmentation from the DRSE-Net Model

V. CONCLUSIONS

In the study DRSE-Net model is evaluated on NIH dataset with data albumentation. The results obtained by experimentation are compared with recent state of the art methods presented in related works. The proposed model with SE block reveals the average dice-score of 88.90% and Jaccard index of 79.96% compared with state of the art studies with less computational parameters of 9.8M by obtaining good results for segmenting small tissues. Few limitations in our work are not all the performance measures are best compared with other methods. But as an overall, proposed approach is more simple and generalized to segment organs.

VI. ACKNOWLEDGEMENT

This work was carried out using the equipment funded by the AICTE [Grant No. F.No.9-29/IDC/MODROB/Policy-1/2019-20].

REFERENCES

- [1] The American Cancer Society, <https://www.cancer.org/cancer/pancreatic-cancer/>.
- [2] L.M. Prevedello, S.S. Halabi, G. Shih, C.C. Wu, M.D. Kohli, F.H. Chokshi, B. J. Erickson, J. Kalpathy-Cramer, K.P. Andriole, A.E. Flanders, Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions, Radiology: Artificial Intelligence (2019) e180031, <https://doi.org/10.1148/ryai.2019180031>.

- [3] E. Gibson, F. Giganti, Y. Hu et al., Automatic Multi-Organ Segmentation on Abdominal CT with Dense V-Networks IEEE Transactions on Medical Imaging PP 2018.
- [4] M. Larsson, Y. Zhang, F. Kahl, Robust abdominal organ segmentation using regional convolutional neural networks, *Appl. Soft Comput.* 70 (2018) 465–471.
- [5] A.E. Kavur, N.S. Gezer, M. Bans., S. Aslan, P.H. Conze, et al., CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation, *Med. Image Anal.* 69 (2021), 101950.
- [6] Y. Wang, G. Gong, D. Kong, Q. Li, J. Dai, H. Zhang, J. Qu, X. Liu, J. Xue, Pancreas segmentation using a dual-input v-mesh network, *Med. Image Anal.* 69 (2021), 101958.
- [7] Mei-Ling Huang , Yi-Zhen Wu, Semantic segmentation of pancreatic medical images by using convolutional neural network, *Biomedical Signal Processing and Control* 73 (2022) 103458, <https://doi.org/10.1016/j.bspc.2021.103458>.
- [8] Dingwen Zhang , Jiajia Zhang , Qiang Zhang, , Jungong Han, ,Shu Zhang , Junwei Han, Automatic pancreas segmentation based on lightweight DCNN modules and spatial prior propagation, *Pattern Recognition* 114 (2021) 107762, <https://doi.org/10.1016/j.patcog.2020.107762>.
- [9] Yan Wang , Jianpeng Zhang , Hengfei Cui , Yanning Zhang , Yong Xia, View adaptive learning for pancreas segmentation, *Biomedical Signal Processing and Control* 66 (2021) 102347, <https://doi.org/10.1016/j.bspc.2020.102347>.
- [10] Meixiang Huang , Chongfei Huang, Jing Yuan,Dexing Kong , A Semiautomated Deep Learning Approach for Pancreas Segmentation, *Hindawi, Journal of Healthcare Engineering Volume 2021, Article ID 3284493*, <https://doi.org/10.1155/2021/3284493>.
- [11] Mizuho Nishio , Shunjiro Noguchi , Koji Fujimoto, Automatic Pancreas Segmentation Using Coarse-Scaled 2D Model of Deep Learning: Usefulness of Data Augmentation and Deep U-Net, *Appl. Sci.* 2020, 10, 3360; doi:10.3390/app10103360.
- [12] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 240–248. Springer, 2017.
- [13] Shruti Jadon, A survey of loss functions for semantic segmentation, 978-1-7281-9468-4/20, IEEE, arXiv: 2006.14822v4 [eess.IV], 2020.
- [14] Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking astrous convolution for semantic image segmentation. ArXiv: 1706.05587 (2017)
- [15] Liang-Chieh Chen, George Papandreou, Senior Member Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille, DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Astrous Convolution, and Fully Connected CRFs, arXiv: 1606.00915v2, 2017.
- [16] 1 Grauman, K., Darrell, and T.: The pyramid match kernel: Discriminative classification with sets of image features. In: ICCV. (2005).
- [17] Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: CVPR. (2006)
- [18] He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: ECCV. (2014).
- [19] Long, J., Shelhamer, E., Darrell, and T.: Fully convolutional networks for semantic segmentation. In: CVPR. (2015)
- [20] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. (2015)
- [21] Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. PAMI (2017).
- [22] Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: ICCV. (2015).
- [23] Lin, G., Milan, A., Shen, C., Reid, I.: Refinenet: Multi-path refinement networks with identity mappings for high-resolution semantic segmentation. In: CVPR. (2017)
- [24] Pohlen, T., Hermans, A., Mathias, M., Leibe, B.: Full-resolution residual networks for semantic segmentation in street scenes. In: CVPR. (2017).
- [25] Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large kernel matters–improve semantic segmentation by global convolutional network. In: CVPR. (2017).
- [26] Kaiming He; XiangyuZhang; Shaoqing Ren; Jian Sun. (2016). Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
- [27] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Enhua Wu, Squeeze-and-Excitation Networks,arXiv:1709.01507v4,2019.