

CapTransU-Net Model: Accurate Diagnosis of Kidney Tumor Segmentation Using Convolutional Capsule U-net Transformer

Dr. Velagapudi Sreenivas

Department of Computer Science and Engineering

Dhanekula Institute of Engineering and Technology Vijayawada, Andhra Pradesh- 521139, India.

velagapudisreenivas@gmail.com

Abstract: Globally, kidney cancer is a public health issue that affects thousands of individuals. Accurate segmentation of kidney tumors is a critical task that aids doctors to reduce their time to interpret images, making diagnoses and formulating treatment plans. However, manual segmentation is an intensive procedure that varies throughout professionals due to the heterogeneity of tumors. Convolutional neural networks (CNNs) have become more and more popular in recent years for the automatic segmentation of medical images. Although three-dimensional (3D) networks are complex and need a lot of computing power. This research proposes a CapTransU-net model for kidney tumour segmentation with pre-processing module. The pre-processing stage is employed for contrast enhancement and image denoising purpose. The segmentation model used here is convolutional capsule transformer based U-net architecture. The findings of the proposed model are evaluated in terms of dice similarity coefficient and Jaccard index and also, the cross fold validation analysis is conducted to examine the effectiveness of segmentation model.

Key words: Kidney tumor, segmentation, jaccard, dice, capsule transformer, convolutional, U-net model, cross validation.

1. Introduction

Nowadays, kidney cancer ranks among the top 10 cancers that affect both men and women. For a precise diagnosis of kidney-related illness, Computed Tomography (CT) imaging is a vital tool, which produces images with great spatial resolution and contrast [1]. Segmenting kidney tumors is a crucial procedure that preserves medical professional effort when analyzing images, making diagnoses, and formulating treatment plans [2]. Yet, manual segmentation is a laborious procedure that varies among experts due to the heterogeneity of tumors. Kidney cancer is currently a very severe disease and early diagnosis has been much needed, which can be enhanced by computational techniques [3, 4]. So that some researchers are focused on machine learning (ML) based an effective automatic kidney cancer segmentation (KCS). The support vector machine (SVM) model based KCS technique is used for distinguishing the cancer of renal cells from the CT scans but it attains less effectiveness [5].

Recently, the world's obsession with deep learning algorithms (DL) has produced amazing advancements in the kidney segmentation field. The utilization of convolutional neural network (CNNs) to develop segmentation schemes with reduced reliance on previous shape statistics and increased automation [6]. However, CNN-based techniques typically neglected the particular shape-related data. For the purpose of semantic segmentation in CT images, a number of weakly-supervised CNNs have been created recently. Furthermore, another weakly-supervised instance involved the segmentation of kidney CT images by a fully convolutional network (FCN) that was trained with super pixel labels [7, 8]. Although weakly learned-based techniques demonstrate equivalent efficiency in other organs but they haven't been used on accurate KCS. For the Kidney Tumor Segmentation Challenge, few researchers suggest a cascaded segmentation approach utilizing volumetric FCNs. However, the cascaded segmentation techniques always lead to high computational cost and large time consumption [9, 10].

According to recent developments, Deep CNNs (DCNN) are shown to perform better in learning-based segmentation tasks. Numerous researches built upon the U-Net model have excelled in a KCS challenge but they have weak capacity for generalization [11]. Few of authors give efforts were concentrated on improved 3D U-Net and making greater use of the little training dataset. However, it is not suitable for large amount of dataset. To address these conventional issues, this research we focused on an efficient u-net transformer based tumor segmentation model.

2. Related Work

Luana et al. [12] had been developed an innovative Kidney-tumor segmentation method with the help of 2.5D-CNN model. Here, the suggest method can be integrated with three stages such as pre-processing, initial segmentation and reduction. During pre-processing stage, windowing, proportional dataset distribution and data augmentation real time is utilized. To perform the starting stage of detection and segmentation of kidney tumor

the DeepLabv3 and 2.5-D model with DPN-131 encoder is utilized. In post processing, it removes the unwanted elements and derived the final segmentation.

In order to segment the large-sized kidney tumor, Ehwa et al. [13] had been suggested a 3D-Multi-Scale Residual Fully Convolutional Neural Network (MS-RFCNN). The performance of the suggested method is validated in KiTS dataset. Comparing with the existing method, 3D-MS-RFCNN showed greater accuracy. Hsiao et al. [14] had been suggested a way to calculate the volume of kidney tumor. And segmentation are also performed on CT images with the help of Deep Learning. During pre-processing stage, statistical HU windowing, image screening and alternative process labeling is utilized. The results attained an automatic volume calculator which yielded a volume difference of 1.43% respectively.

In order to accurately diagnosing Kidney tumor, Abu Baker [15] had been suggested a Feature Pyramid Network and Squeeze-and-Excitation-ResNet (FPN-SE-ResNet) Model. Here, the suggested model's accuracy and effectiveness are evaluated by Intersection over Union (IoU) and F1-score metrics.

3. Proposed Methodology

In order to perform the accurate segmentation process, this research can employs an effective model name as CapTransU-net. Initially, the Kidney cancer images are fed into the preprocessing stage in order to perform a denoising process. After that, the cancer affected tissues in the kidney is segmented using the CapTransU-net model. The workflow of the proposed model design is demonstrated in Figure 1 and the in-depth description of the both preprocessing and CapTransU-net model is illustrated in below.

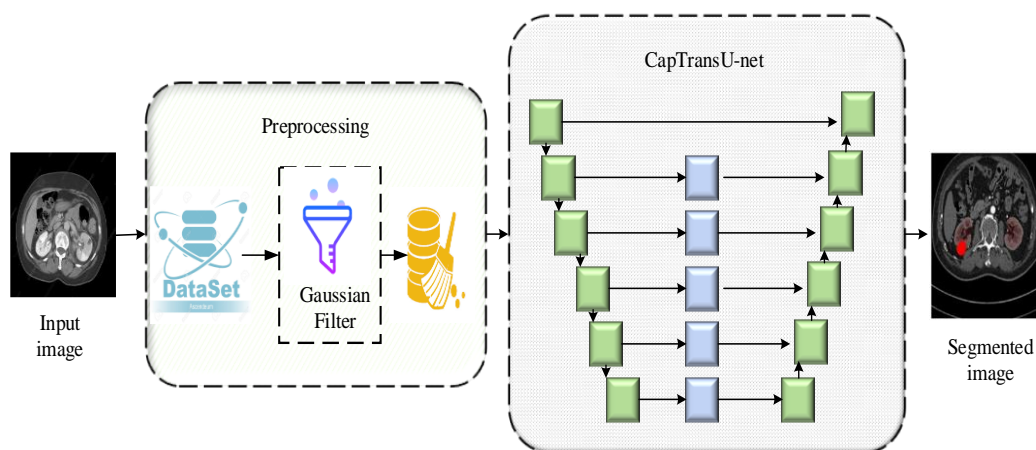


Figure 1. Block diagram of proposed method

3.1 Preprocessing

In this research, the goal of the pre-processing stage is to identify and minimize the defects of the Kidney CT images. The removal of undesirable artifacts and noise from CT images, this stage is important for effective KCS. There are three phases are utilized in this stage such as image scaling, noise reduction, and local contrast enhancement. Initially, the input images are reshaped as same dimension. After that, the Gaussian smoothing filter is employs in order to eliminating the noises. Eventually, localized contrast of the images was enhanced using a logarithm operator, which increases the low-intensity pixel value. Following preprocessing, the fresh augmented X-ray pictures are typically produced at the end by applying the geometric augmentations. Finally, the preprocessed images are fed into the segmentation model in order to segment the cancer region.

3.2 Segmentation using CapTransU-net

This research employs a deeply layered U-shaped architecture termed as CapTransU-net for KCS process and the diagrammatic representation of the proposed model is depicted in Figure 2. Although, a highly layered framework can use a lot of resources during training. As can be seen in Figure 2, the TransU2-Net networking architecture has a two-level layered U-shaped framework. The TransU2-Net takes in $4 \times 160 \times 160$ as its input size. The proposed backbone CapTransU-net network framework is composed with 17 blocks, including 5 transformer and 12 convolution blocks respectively. In order to gain deeper global data, this research additionally employing a capsule transformer in the U-Net hierarchy. Furthermore, this research can incorporate feature map information at multiple levels and the jump fusion technique fuses low-level features to high-level features, resulting in more powerful linguistic data in the final KCS outcome. The detailed implementation of each block is described in depth in the subsequent sections.

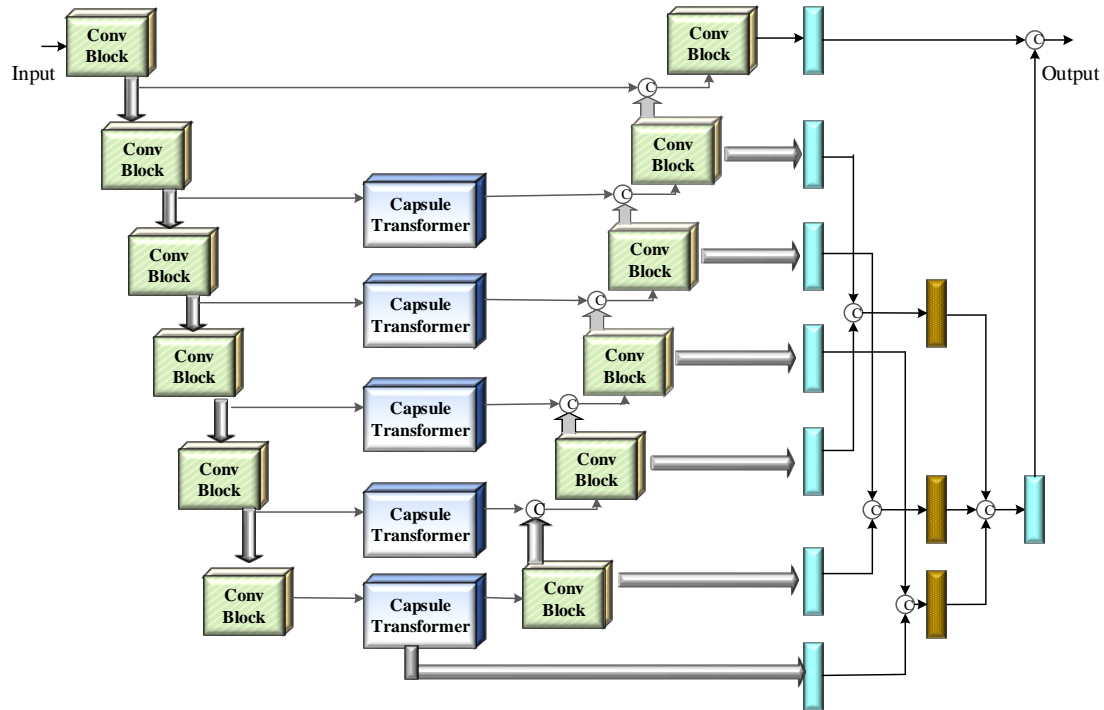


Figure 2. Overall Architecture of the proposed CapTransU-net model

3.2.1 Convolution Layer

The proposed model employs the Conv block over every single layer of the decoding and encoding processes to perform difficult KCS tasks. The proposed model can able to extract more information and finer features because the module resembles an expanded U-Net framework. Furthermore, the dilated convolution is utilized in the intermediate layer in every conv block to reduce overfitting and improve the perception area. Every Conv Block framework is shown in Figure 3 with one step size and padding respectively and the size of each convolutional kernel is 3×3 . The model can reduce the amount of convolution layers when each Conv Block as the phase goes higher, and in phases 5 and 6, the model switches out the pooling and up sampling processes by expanded convolutions. This means that in stages 5 and 6, every intermediary feature maps have identical resolution as their input feature maps.

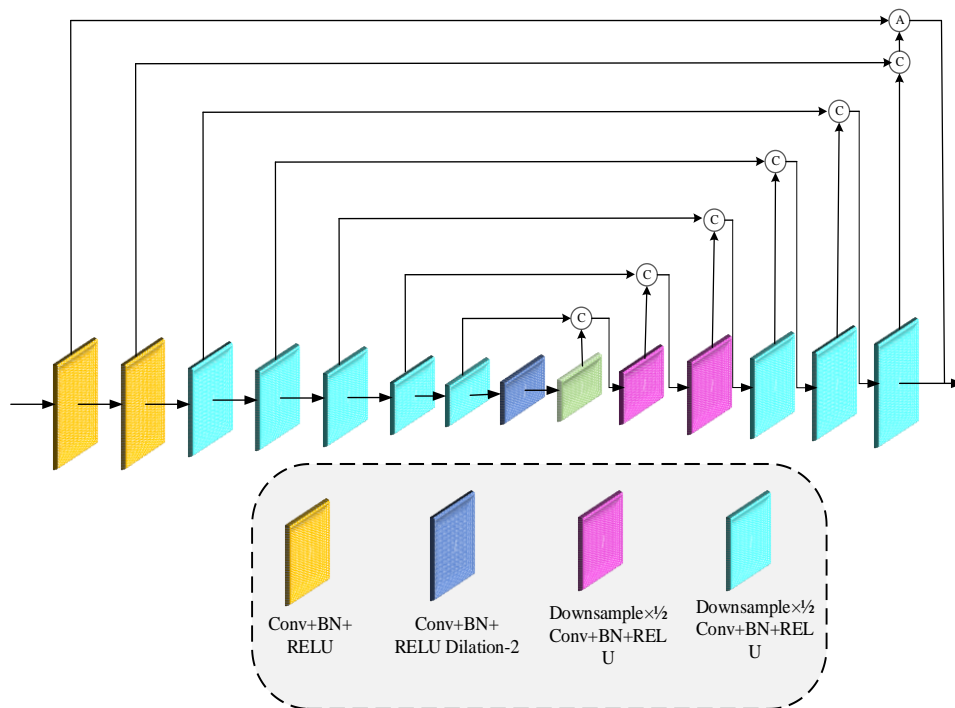


Figure 3. Conv block layout

3.2.2 Capsule Transformer

Using a self-attention process, the transformer encoder's role is to take benefit of the extended global feature relationships of its input tokens. A series of capsule transformer (CT) units layer up, exporting the input tokens' aggregate depiction of features. This proposed model can use the same quantity of CT blocks. It should be noted that throughout each of the CTs, the transformer encoder keeps the identical token quantity, feature channel number, and capsule dimension of D . In other words, every CT's result format and input token sizes match exactly. Typically, two simultaneous multi-head self-attention (MSA) and one Multi-layer perception (MLP) is utilized to capitalize on global and local features respectively, that two modules are used to form the CT block. The diagrammatic illustration of the CT block is represented in Figure 4.

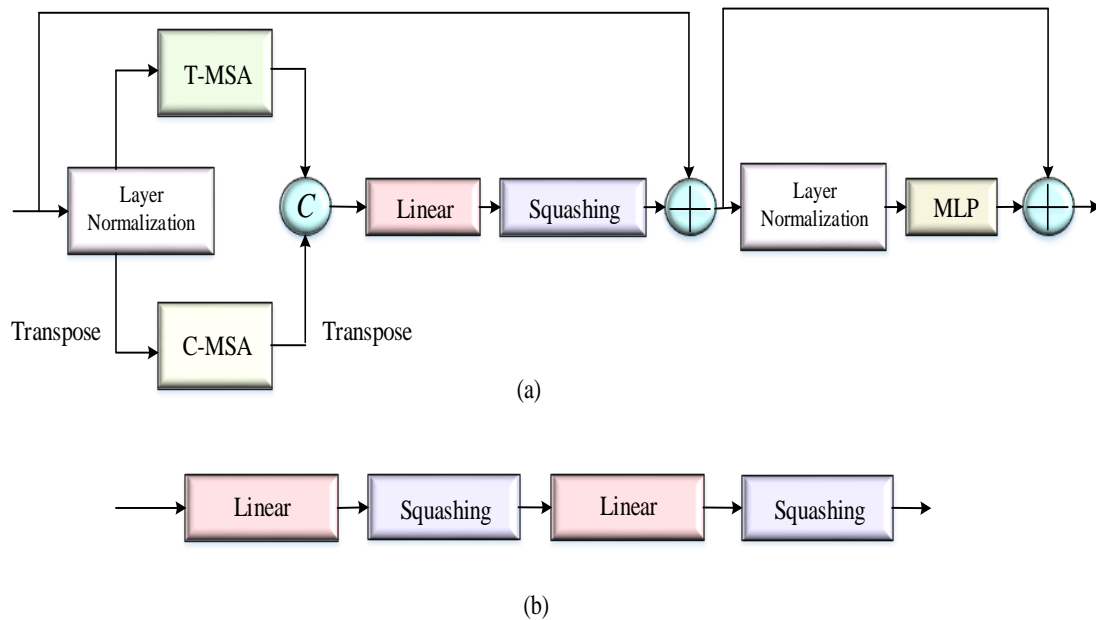


Figure 4. (a) General architecture of the CT block (b) Systematic view of MLP

From the Figure 4 (a), it can seem that the normalized layer is attached between the MLP and MSA modules. Model augmentation is then carried out via a residual connection, which appears following each of them. Initially, the input sequences are normalized with help of normalized layer. After that, the normalized data are copied into two set and fed to token-wise MSA (T-MSA) and channel-wise MSA(C-MSA) in order to perform a global feature self-attention from token' and channel' perspectives respectively. In this case, the tokens are changed between the spatial domain to the channel domain using the transpose process. The input token is specified as $T^{tS \times FC \times CS}$, here, the input token size, capsule size and number of feature channel is represented as tS , CS and FC correspondingly. Furthermore, $T^{tS \times FC \times CS}$ and $T^{FC \times tS \times CS}$ is denoted as input of T-MSA and C-MSA respectively. Eventually, the results from the T-MSA and C-MSA modules are combined and transformed to the linear layer for further fusion and then a squashing nonlinearity. The MLP make up with two linear and Squashing layers correspondingly and the systematic view of the MLP is illustrate in Figure 4 (b).

Following a capsule-based MSA structure, the T-MSA and C-MSA units have a similar design, which can demonstrate in Figure 5 (a). Initially, three distinct sets of linear layers such as query, key, value set respectively are organized from the input embedded tokens $T^{tS \times FC \times CS}$. Each of them includes n distinct parallel linear layers with various parameters and the mathematical representation of the Q_s , K_s , V_s are depicted in Eqn (1),(2) and(3) correspondingly.

$$Q_s = \{Q_1, Q_2, \dots, Q_N\}, Q_i \in T^{tS \times FC} \quad (1)$$

$$K_s = \{K_1, K_2, \dots, K_N\}, K_i \in T^{tS \times FC \times CS} \quad (2)$$

$$V_s = \{V_1, V_2, \dots, V_N\}, V_i \in T^{tS \times FC} \quad (3)$$

Here, $i = 1, 2, \dots, N$. At last, every n head's results are successively combined then merged via a linear layer and a squashing nonlinearity. Figure 5(b) provides an illustration of the single-head self-attention module's elaborate construction.

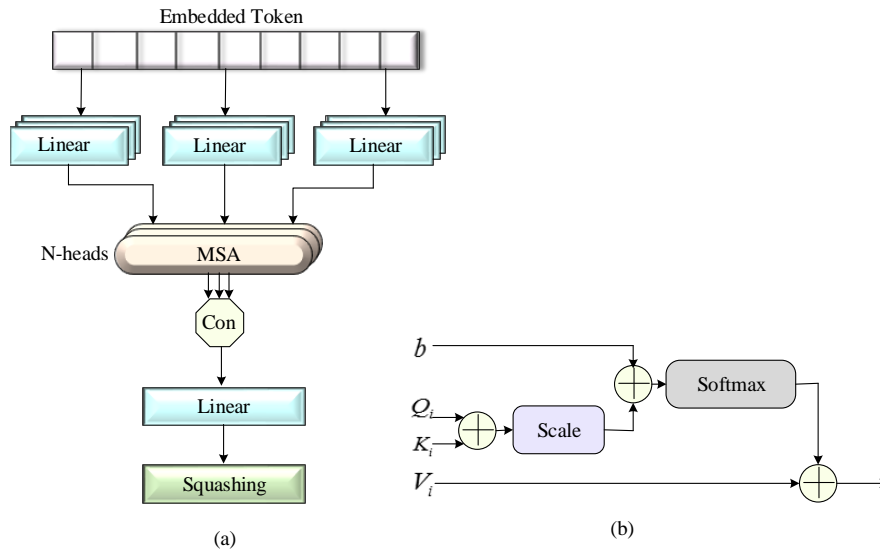


Figure 5. (a) Architectural diagram of MSA (b) layout of single head self-attention

4. Result and Discussion

The experimental results of the proposed method is presented in this section. The proposed method is implemented in python programming language and the database used here is KiTS19. To verify the effectiveness of the proposed segmentation model, several measures are examined which includes, dice similarity coefficient and Jaccard index (IoU). Intersection over Union (IoU) employs the ratio of common pixels to quantify the similarity between the ground truth and predicted masks. The balanced measure of segmentation accuracy is termed as Dice and it is also named as F score. The mathematical formulation of two metrics are stated below:

$$IoU = \frac{T_p}{T_p + F_p + F_N} \tag{4}$$

$$DSC = \frac{2T_p}{2T_p + 2F_p + F_N} \tag{5}$$

Here, the true positive, false negative and false positive are indicated as T_p , F_N and F_p respectively.

4.1 Cross validation analysis

The cross validation is conducted to examine the performance of the proposed segmentation model and here the fivefold cross validation analysis takes place by the consideration of with and without preprocessing stages. Therefore, the cross validation analysis of dice score under several existing methods are shown in table 1 and 2 respectively.

Table 1. K-Fold summarization of with preprocessing

Kidney Dice Score	ResNet-41 (%)	EfficientNet-B4 Unfrozen encoder (%)	Proposed model (%)
First fold	94.55	94.91	99.43
Second fold	96.18	96.86	98.67
Third fold	94.95	95.10	98.71
Fourth fold	95.26	97.05	99.03
Fifth fold	95.20	95.94	99.65
Average	95.23	95.97	99.09

Table 2. K-Fold summarization of without preprocessing

Kidney Dice Score	ResNet-41 (%)	EfficientNet-B4 Unfrozen encoder (%)	Proposed model (%)
First fold	93.41	94.66	95.23
Second fold	93.63	96.44	96.23
Third fold	93.53	93.79	96.78

Fourth fold	94.13	96.50	96.99
Fifth fold	92.18	95.41	95.34
Average	93.38	95.36	96.31

4.2 Comparison with State of Art Techniques

In this section, the effectiveness KCS of the proposed model is compared with various techniques such as CNN [16], U-Net [17-20], V-Net [21], Cascaded [22], RAU-Net [23], FPN-Resnet50 and 3D-U-Net [24]. Typically, the CNNs utilized for multiple hyper parameters extraction but sometimes method for adjusting hyper parameters can reduce the segmentation accuracy. The U-Net and V-Net are mostly employed for segmentation problems yet its need preceding outcomes for accurate KCS. Recently, Cascaded, RAU-Net and 3D –Net are introduced for accurate KCS, however its takes lots of time for computation. Furthermore, FPN-Resnet50 model is suggested by some authors for KCS process but it’s require high energy. When compared to other model, proposed method can attains high outcomes such as Kidney dice, tumor dice, kidney LoU score and Tumor LOU score as 99.76 %, 98.78 %, 99.12% and 98.99 % correspondingly. The comparative analysis of various existing and proposed model is illustrated in figure 6.

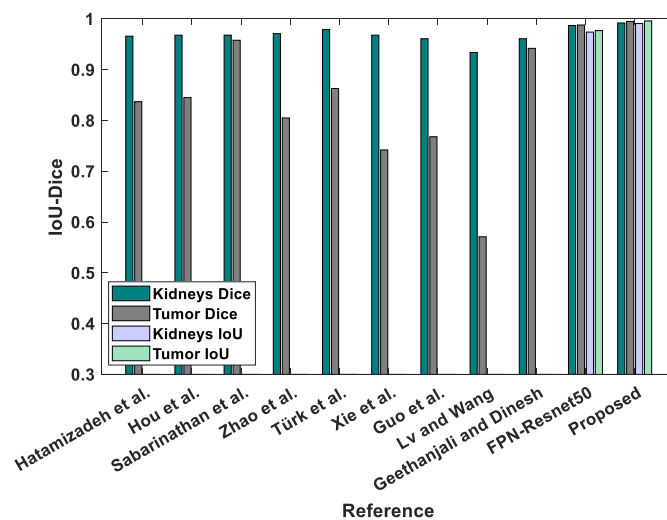


Figure 6. Comparative analysis of segmentation model

5. Conclusion

In the past, medical professionals used their expertise in diagnosing kidney disorders to interpret computed tomography (CT) images. However, due to the quick rise in CT scans, these interpretations took a lot of time and effort, and the outcomes were not always reliable. To address this issue, a number of innovative neural network models have been carried together to automatically recognize kidney or tumor regions on CT scans. Therefore, this research proposes a CapTransU-net model for kidney tumor segmentation. In future, we recommend looking at alternatives to parameter optimization methods because defined models typically include a large number of experimentally determined hyperparameters.

References

- [1] How QM, Lee HL (2022) Kidney and Kidney Tumour Segmentation in CT Images. arXiv preprint arXiv:2212.13034.
- [2] Srinivasan M, Soji ES (2022) Kidney tumour segmentation and classification using deep learning. Central Asian Journal of Medical and Natural Science vol. 3, no. 6, pp. 247-69.
- [3] Thomas NR, Anitha J (2022) An automated kidney tumour detection technique from computer tomography images. In 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS) pp. 1-6. IEEE.
- [4] Yu Q, Shi Y, Sun J, Gao Y, Zhu J, Dai Y (2019) Crossbar-net: A novel convolutional neural network for kidney tumor segmentation in ct images. IEEE transactions on image processing vol. 28, no. 8, pp. 4060-74.
- [5] Abdelrahman A, Viriri S (2022) Kidney tumor semantic segmentation using deep learning: A survey of state-of-the-art. Journal of imaging vol. 8, no. 3, pp. 55.
- [6] Türk F, Lüy M, Barışçı N (2020) Kidney and renal tumor segmentation using a hybrid V-Net-Based model. Mathematics vol. 8, no. 10, pp. 1772.
- [7] Kang L, Zhou Z, Huang J, Han W, Member IE (2022) Renal tumors segmentation in abdomen CT Images using 3D-CNN and ConvLSTM. Biomedical Signal Processing and Control vol. 72, pp. 103334.

- [8] Marie F, Corbat L, Chaussy Y, Delavelle T, Henriot J, Lapayre JC (2019) Segmentation of deformed kidneys and nephroblastoma using case-based reasoning and convolutional neural network. *Expert Systems with Applications* vol. 127, pp. 282-94.
- [9] da Cruz LB, Araújo JD, Ferreira JL, Diniz JO, Silva AC, de Almeida JD, de Paiva AC, Gattass M (2020) Kidney segmentation from computed tomography images using deep neural network. *Computers in Biology and Medicine* vol. 123, pp. 103906.
- [10] Santini G, Moreau N, Rubeaux M (2019) Kidney tumor segmentation using an ensembling multi-stage deep learning approach. A contribution to the KiTS19 challenge. arXiv preprint arXiv:1909.00735.
- [11] Choudhari K, Sharma R, Halarnkar P (2020) Kidney and tumor segmentation using U-Net deep learning model. In5th International Conference on Next Generation Computing Technologies (NGCT-2019) 2020.
- [12] Hsiao CH, Sun TL, Lin PC, Peng TY, Chen YH, Cheng CY, Yang FJ, Yang SY, Wu CH, Lin FY, Huang Y (2022) A deep learning-based precision volume calculation approach for kidney and tumor segmentation on computed tomography images. *Computer methods and programs in biomedicine* vol. 221, 106861.
- [13] da Cruz LB, Júnior DA, Diniz JO, Silva AC, de Almeida JD, de Paiva AC, Gattass M (2022) Kidney tumor segmentation from computed tomography images using DeepLabv3+ 2.5 D model. *Expert Systems with Applications* vol. 192, pp. 116270.
- [14] Yang E, Kim CK, Guan Y, Koo BB, Kim JH (2022) 3D multi-scale residual fully convolutional neural network for segmentation of extremely large-sized kidney tumor. *Computer Methods and Programs in Biomedicine* vol. 215, pp. 106616.
- [15] Abdelrahman A, Viriri S (2023) FPN-SE-ResNet model for accurate diagnosis of kidney tumors using CT images. *Applied Sciences* vol. 13, no. 17, pp. 9802.
- [16] Hatamizadeh A, Terzopoulos D, Myronenko A (2020) Edge-gated CNNs for volumetric semantic segmentation of medical images. arXiv preprint arXiv:2002.04207.
- [17] Hou X, Xie C, Li F, Wang J, Lv C, Xie G, Nan Y (2020) A triple-stage self-guided network for kidney tumor segmentation. In2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI) pp. 341-344. IEEE.
- [18] Sabarinathan D, Parisa Beham M, Mansoor Roomi SM (2020) Hyper vision net: kidney tumor segmentation using coordinate convolutional layer and attention unit. InComputer Vision, Pattern Recognition, Image Processing, and Graphics: 7th National Conference, NCVPRIPG 2019, Hubballi, India, December 22–24, 2019, Revised Selected Papers 7 pp. 609-618. Springer Singapore.
- [19] Zhao W, Jiang D, Queralt JP, Westerlund T (2020) Multi-scale supervised 3D U-Net for kidneys and kidney tumor segmentation. arXiv preprint arXiv:2004.08108.
- [20] Geethanjali TM, Dinesh MS (2021) Semantic segmentation of tumors in kidneys using attention U-Net models. In2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICECCOT) pp. 286-290. IEEE.
- [21] Türk F, Lüy M, Barışçı N (2020) Kidney and renal tumor segmentation using a hybrid V-Net-Based model. *Mathematics* vol. 8, no. 10, pp. 1772.
- [22] Xie X, Li L, Lian S, Chen S, Luo Z (2020) SERU: A cascaded SE-ResNeXT U-Net for kidney and tumor segmentation. *Concurrency and Computation: Practice and Experience* vol. 32, no. 14, p. e5738.
- [23] Guo J, Zeng W, Yu S, Xiao J (2021) RAU-Net: U-Net model based on residual and attention for kidney and kidney tumor segmentation. In2021 IEEE international conference on consumer electronics and computer engineering (ICCECE) pp. 353-356. IEEE.
- [24] Lv Y, Wang J (2021) Three uses of one neural network: automatic segmentation of kidney tumor and cysts based on 3D U-Net. InInternational Challenge on Kidney and Kidney Tumor Segmentation pp. 40-45. Cham: Springer International Publishing.

Authors Profile



Dr. Velagapudi Sreenivas is currently working as an Associate Professor in Department of Computer Science and Engineering in SRK Institute of Technology. He Received ph.d in SRK University, Bhopal and M.S from University of Hertfordshire, U.K. He has Published 30 International Journals which include SCI and Scopus Papers. He has published two textbooks and two patents. He has applied for Project Funding to various agencies like SERB, DST, UGS and many more. He is currently active in the area of research are Machine Learning, AI, Big data and other areas.