IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 1, February, 2023, pp. 217-225
Journal homepage: http://iieta.org/journals/ts

Depth Invariant 3D-CU-Net Model with Completely Connected Dense Skip Networks for MRI Kidney Tumor Segmentation

Sitanaboina S.L. Parvathi*, Bolem Sai Chandana, Jonnadula Harikiran

School of CSE, VIT-AP University, Amaravathi 522237, Andhra Pradesh, India

Corresponding Author Email: parvathi.19phd7017@vitap.ac.in

https://doi.org/10.18280/ts.400120

Received: 23 November 2022
Accepted: 6 February 2023

Keywords:
3D-CU-Net model, U-Net model, kidney tumor segmentation, loss function and deep learning

ABSTRACT

Due to the impact and importance of the kidney objects in human body, the kidney tumor analysis from three dimensional CT and MRI medical images becomes a pivotal research topic, which helps in diagnosing the kidney diseases like kidney stones, polycystic and kidney tumors etc. In deep learning, U-Net became a prominent and reliable solution for kidney image analysis and objects segmentation process. Although several research works were focused on kidney object detection and tumor segmentation from medical images, they are suffering from some intrinsic limitations due to: variance in network depths, enforced feature fusion, segmentation errors and inaccuracy. In order to address these limitations in kidney tumor segmentation process, in this paper we proposed the 3D-CU-Net model for kidney tumor segmentation, which is a custom variant of the U-Net. In 3D-CU-Net, the encoder-decoder network model is unified to tolerate the depth invariance issues, while training various input images with the same model. Completely connected dense skip connections are designed at each layer of 3D-CU-Net, to control the enforced feature fusion and to extract the crucial features. An integrated loss function is designed with Binary Cross Entropy (BCE) and Soft-Dice Coefficient (SDC) to mitigate the segmentation errors and inaccuracy. Experiments on TCGA-KIRC dataset with 3D-CU-NET recorded the high accuracy in kidney tumor segmentation with mIoU (91.21%) and mDSC (92.69%).

1. INTRODUCTION

Since a decade, the kidney disease patients are growing up worldwide and today the 15% of US adult population is suffering from various kidney diseases. Among these kidney diseases, carcinoma tumors are the most dangerous and cause to high mortality in diseased patients. Abnormal or chaotic growths on kidneys are considered as tumors or renal mass, which may be cancerous or non-malignant. Tumors are scattered over various regions of the kidneys with different shapes and sizes.

In medical science, disease diagnosis is the primary activity, which happens by thoroughly analyzing the patient's clinical and radiological (medical image) information. X-rays, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) and Ultrasound are the popular imaging models in radiology. As the MRI contrast agents are safer than other imaging methodologies, physicians use to prefer the 3D-MRI scan for kidney disease diagnosis. According to Glaser et al. [1] the 3D-MRI images supports the multi planner reformatting and provides the equal spatial resolution for X, Y planes and Z direction. This isotropic spatial resolution helps in constructing the high definition kernels and sharpened filters while processing the 3D images for disease diagnosis.

In order to detect and lineate the tumors from kidney diseases, the 3D-MRI images should be analyzed thoroughly at pixel level. Base Region of Interest (RoI) selection, objects detection, boundaries segmentation and Tumors lineation are the main phases of the 3D- MRI kidney object segmentation and tumors selection process. As it is widely utilizing in disease diagnosis, severity estimation and surgeries planning, automated segmentation of the tumors from 3D-MRI kidney images became a popular research topic in recent years.

Most of the former research works concentrated on kidney MRI 2D image segmentation over the 3D image segmentation, due to the segmentation process complexity and resource deficiency. In imaging modalities, the 3D images are composed with the isotropic voxels, which provides the high spatial resolution and improvises the visibility of the image objects at low noise rate. Indeed the 3D images are very accurate in disease diagnosis than 2D images, due to the multi-dimensional planes acquisition and higher signal to noise ratio while scanning. In general, the 3D MRI segmentation is having more process complexity than 2D, but 3D provides the better visualization of the small objects and also helps in estimating the objects size, shape, convexity, inertia and location. The advantages of MRI 3D images made them prominent and widely using in clinical diagnosis.

Inspired from the advantages of MRI 3D images and to facilitate the in depth analysis with limited resources, our research is aimed to design the accurate and reliable 3D MRI kidney tumor segmentation model. As part of the design of automated kidney tumor segmentation model, literature review is conducted and a set of relevant former research works were reviewed for problem definition is presented in section-2. As on several research works proposed various deep learning models (i.e. ResNet, VGG-16, GoogLeNet, U-Net etc.) for segmentation discussed by Parvathi et al. [2], among them the U-Net become more popular and reliable due to its wide variety of characteristics.

Although the former researcher works concentrated on U-Net for designing the 3D MRI kidney tumor segmentation, they encountered a set of the considerable limitations are: i) variance in network depths, ii) enforced feature fusion, iii) segmentation results inaccuracy and errors.

To overcome these limitations in 3D MRI kidney tumors diagnosis process, an accurate and reliable 3D-CU-Net model is proposed in this paper, for kidney object and tumor boundaries segmentation. Unified encoder-decoder networks are designed to counter the variance in network depth and completely connected dense skip connections are designed to counter the enforced feature fusion. Efficient integration of the loss functions reduced the errors in training and improved the results accuracy.

The main objective of this study is designing a deep learning model for automated segmentation of the kidney object and tumor from 3d MRI images, to help the physicians in disease diagnosis, severity estimation, treatment planning etc. As part of the research experiments, the proposed model is developed using the python supporting image processing libraries. TCGA-KIRC kidney MR images are selected to conduct the experiments on tumor diagnosis and the comparative analysis also conducted with 3D-CU-Net counterparts. Comparative analysis proven that; the proposed 3D-CU-Net achieved the high accuracy in kidney tumor diagnosis.

Rest of this paper is organized as follows: section-2 presents the literature analysis on U-Net and the other deep learning models with their limitations in segmentation process. Section-3 describes the proposed 3D-CU-Net architecture, functionality and process flow information. Section-4 presents the experimental results and comparative analysis over the proposed model. Section-5 presents the conclusions and future research directions on proposed model.

2. RELATED WORK

In this section, the literature on basic U-Net architecture and its limitations in MRI medical image segmentation process are described in brief.

U-Net: In recent times, U-Net became a popular deep learning model for 3D medical image (i.e., CT or MRI) seamless segmentation process, was introduced by Ronneberger et al. [3]. Basic U-Net architecture is an encoder-decoder networking model, designed with parallel processing paths are: contracting path (encoder part) and the expanding path (decoder part). In U-Net the encoder sub-network part performs the convolutions (down-sampling and pooling operations) on low-level fine-grained features of training samples for semantic segmentation. Similarly, the decoder sub-network part performs the de-convolutions (up-sampling and concatenation) on coarse-grained semantic features of convolved samples for target object instance detection and segmentation. The input data images are encoded first to extract the feature maps and then decoded with feature propagation to classify the target output with same pixel resolution. Each level of U-Net encoders and decoders are symmetrically connected through the skip connections and bottleneck layers to concatenate the encoder and decoder paths. The processing encoder node is connected with all the other nodes of that level including decoder node to forward the knowledge obtained in training to the others.

The encoder path performs 3x3 convolutions with Rectified Linear Unit (ReLU) activation functions to reduce the input size and to highlight the target features Ronneberger et al. [3]. The highlighted target features are then down sampled with 2x2 max-pooling operations. The same set of operations (Conv+ReLU+Max-Pool) evaluated at each level of encoders and the resultant feature maps are forwarded to the decoders using the cascading convolutional operations. At decoder path, the extracted features are up-sampled (2x2 Conv+ReLU) to separate the foreground and background information. After training the model with stochastic gradient descent, the pixel-wise soft-max evaluates the energy function over the loss function is as follows:

$$\Delta(E)=\sum_{q\in\Omega}\left(W_c(q)+W_0*\exp\left(-\frac{(r_1(q)+r_2(q))^2}{2\sigma^2}\right)\right)\log(P_{l(q)}*(q) \qquad (1)$$

The energy function $\Delta(E)$ is evaluated in above equation with the weighted map W_c, softmax function $P_{l(q)}$, distances from border to the first nearest cell r_1 and second nearest cell r_2, pixel $q \in \Omega$ and the σ is a static value.

U-Net Limitations: By impressing from these segmentation features, some former researchers utilized this 'U' shaped convolutional neural network (U-Net) for medical image segmentation and identified some limitations in segmentation process are: variance in network depths, enforced feature fusion, semantic segmentation inaccuracy and segmentation errors.

In general, the depth of the U-Net architecture is decided by the input dataset factors are: class labels, image size and process complexity etc. Based on this training data decision making factors, the U-Net architecture depth is selected at runtime, which may be different from application to application. Ciompi et al. [4] highlighted that, the variance in network depth requires a separate U-Net model for each depth level. Some applications may need to utilize various U-Net architectures (with depth variance) to train the same model with the underlying dataset images. In medical image datasets, the images are collected from various patients and devices are having the variance in their properties like size, brightness and contrast etc. Due to the variance in medical image properties, various U-Net models are required to train various images with depth variance. In order to obtain the comprehensive knowledge in this way of training, the multiple trained models needed to be consolidated. Dietterich et al. [5] specified that, the consolidation (ensemble) of various trained networks with depth variance will increases the detection ambiguity and segmentation inaccuracy, due to the uncommon encoders. This unconventional model of training and consolidation will also cause to loss the benefits of multi-tasking in model training.

While training a deep learning model, sometimes the succession decrease of training loss may not improvise the accuracy of finding the ROI, is considered as "vanishing gradient problem". To solve the vanishing gradient problem and to speed up the learning process, the back propagation method was proposed in deep learning models with cost functions. Back propagation method optimizes the partial derivatives by adjusting the hyper parameter weights and bias values at each layer. Although the back propagation optimizes the model parameters iteratively, the frequent backward moves in layers will decrease the gradient value, which causes to the "training instability" [6]. Recent deep learning architectures were introduced the skip connections among the model layers, to alleviate the instability in training and to keep the adequate gradient value [5]. In case of the low gradient values in training, the skip connections will activate the gradient [7] to skip the immediate layers and feed the current gradient as input to the other layers. This method provides the additional connections among the layers of neurons for model convergence and instigates the stability in training with adequate gradients.

U-Net also contains the skip connections, which are simply connected using the feature concatenation process as shown in Figure 1. Even though the U-Net skip connections are assuring the feature maps with gradient as a non-zero, in deep segmentation tasks these simple skip connections are partially extracting the crucial features [8]. On other hand, the U-Net skip connections are designed for the same layer encoder decoder feature maps, which restrict and enforce the feature fusion at same layer. Feature fusion at same layer encoder-decoder networks (via skip connections) may get back the foreground features available in that layer, but not from the previous layers. Same layer features restoration with skip connections is facing the semantic dissimilarity among feature maps due to the less homogeneity in feature properties.

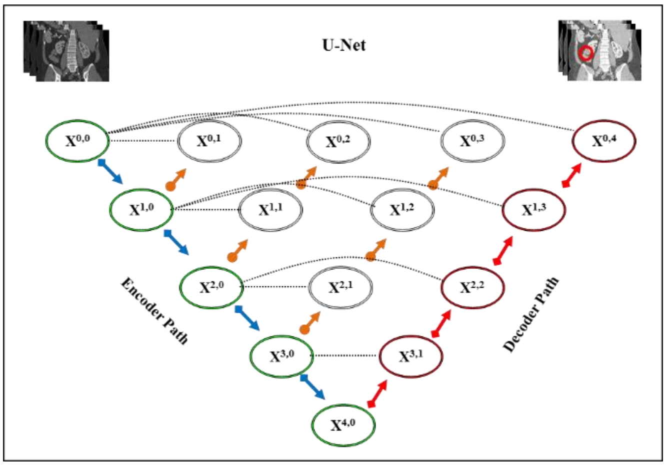

Figure 1. Graphical representation of the U-Net process layers

3. 3D- CU-NET AND KIDNEY TUMOR SEGMENTATION

Deep learning process models emerged as the reliable solutions for medical image data processing and object segmentation. Advancements in deep learning process models gained the capability to diagnose the disease information (i.e., Carcinoma Tumors, Stones, Polycystic, glomeruli etc.) from 3D MRI kidney medical images.

To overcome the challenges and limitations in kidney tumor segmentation process, deep learning-based 3D-CU-Net kidney tumor segmentation model is proposed (shown in Figure 2) in this paper. This model describes the components, connectivity, actions and results of the proposed 3D-CU-Net in an integrated manner with its modules and flow in action. The coherence view of this model presents the organization of the feasible solutions at each module to address the limitations discussed in section-2. At every step of processing, this model is designed with the fine gained solutions, which are simplified, feasible, compatible and reliable.

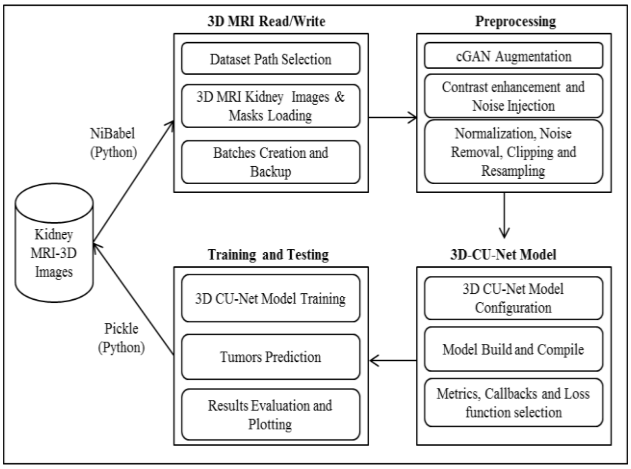
3D MRI Read/Write
Dataset Path Selection
3D MRI Kidney Images & Masks Loading
Batches Creation and Backup
Preprocessing
cGAN Augmentation
Contrast enhancement and Noise Injection
Normalization, Noise Removal, Clipping and Resampling
NiBabel (Python)
Kidney MRI-3D Images
Pickle (Python)
Training and Testing
3D CU-Net Model Training
Tumors Prediction
Results Evaluation and Plotting
3D-CU-Net Model
3D CU-Net Model Configuration
Model Build and Compile
Metrics, Callbacks and Loss function selection

Figure 2. Architecture diagram of the proposed 3D-CU-Net kidney tumor segmentation model

Proposed 3D-CU-Net kidney tumor segmentation model contains 4 main phases in processing are i) 3D MRI Read/Write Phase ii) 3D MRI Data Preprocessing Phase iii) 3D-CU-Net Model Configuration Phase and iv) 3D-CU-Net model Training and Testing phase. Each phase of this model is connected with the next phase and the output of one phase is sent as input to the next phase for further processing.

3D MRI Read/Write Phase: This section explores the process of reading and writing of the kidney 3D MRI images from dataset using the python I/O libraries. Initially a set of kidney 3D MRI images from TCGA-KIRC dataset are collected from the web and stored in a specific system path. As the collected MRI images contain the neuroimaging file format (NifTI1), the NiBabel library functions are selected for read/write operations. After selecting the dataset path, the NiBabel library read functions are used to load the 3D MRI kidney images and its associated masks (ground truths) from path and transforms the pixel data into the three-dimensional numeric vectors for processing. By the time of transformation, each image metadata is collected and stored in headers. With help of this metadata and the vector affine information and position of the image can be referenced while processing. To facilitate the processing model activities (train, validate and test), each 3D kidney image and its relevant mask is stored with same index values at same path.

After preprocessing of the input dataset, this read/write module functions are used to create the batches to initiate the parallel processing and to back up the intermediate models while processing. Once the deep CU-Net model is created, it’s the responsibility of the learning algorithm (gradient descent) to update model from the training examples iteratively. The batch size specifies the number of samples to be considered from the available training set to update the model. At the end of the iterations, the batch elements are shuffled from the training dataset to generate the new batch for next iteration. Similarly, the intermediate batches with high relevance need to be serialized in local space for further processing. In order to serialize those batches temporarily, the “python pickle library” functions are selected for batches serialization and deserialization process.

3D MRI Data Preprocessing Phase: Data preprocessing is a common and prominent phase in medical data analysis, which transforms the raw input data into model acceptable input using various preprocessing techniques. As part of the data preprocessing, a set of image preprocessing techniques (i.e. image synthesis, noise reduction, normalization and resampling) are employed in this study to make the training data sufficient and error free.

In general, the medical data is available in limited quantity, which is insufficient for deep learning models to train and test. To overcome this limitation, image synthesis techniques was implemented using the 3D Conditional Generative Adversarial Networks (3D-cGAN) [9] to generate the plausible dataset images for efficient learning. Among the available data synthesis techniques, the 3D-cGAN is selected because it contains both generator and discriminator networks, which helps in balancing the augmentation process and produces the indistinguishable images from actual dataset, is proven by Wang et al. [10]. 3D-cGAN is having the ability to generate the high contrast synthetic 3D MRI images from the images that are generated under the low contrast agents.

In contrary to the other generator models, the 3D-cGAN model controls the generator (G) and discriminator (D) functionalities with the metadata(m) based conditions as input to them. Generator combines the image noise $I_q(Q)$ and the metadata 'm', whereas the discriminator adjoins the metadata ‘m’ and discrimination methods (n) to generate the synthetic images at the range between minimum and maximum count is as follows:

$$V[\min(G) * \max(D)] = \log D(m|n) * E_{m\approx I(q)} + \left[\log\left(1 - D\big(G(q|n)\big)\right) * E_{q\approx I(Q)}\right] \tag{2}$$

Contrast of an image defines the degree of pixels variation among the image internal objects and it helps in achieving the high accuracy in object detection and segmentation operations. Perumal et al. [11] stated that the adequate contrast enhancement process will improvise the image quality, visualization, boundaries and pixel differences. To improvise the kidney tumor segmentation accuracy, the 3D-cGAN model adopted the contrast enhancement as one of the prominent factor in synthetic images generation as part of the data augmentation process. Contrast enhancement or adjustment helps in differentiate the tumor boundaries from its neighbor tissue boundaries.

In general, the medical image datasets with limited images causes to create the over-fitting problems in learning the process models. In case of over-fitting due to less input medical data availability, the model fit the available limited features, noise and variations from training data in memory, which may fail or less accurate in handling the unseen real life test data. The generalization inabilities (over fitting) of a model will record the high variance in performance over varied datasets. K-fold cross-validation method is employed in testing phase to determine our model is suffering from the over-fitting issues or not. To overcome the over fitting issues in our model, we selected the 3D-cGAN and K-fold cross-validation techniques. D-cGAN generates the high-quality synthetic data images to increase the dataset size and K-fold cross-validation technique for dataset partitions and hyper parameter tuning.

Shorten et al. [12] specified that, adding the Gaussian Noise (GU) or noise injection to the input images under augmentation process will enhances, the trained model ability to perform the efficient segmentation process on low contrast and noisy images. Rotations are the best way of augmentation, which creates the reliable synthetic images by just rotating the image objects without disturbing them. According to Kalra et al. [13], the augmentation with rotation not only increases the dataset size, but also improves the trained model efficiency in classification and segmentation.

Apart from the augmentation, the preprocessing phase employed the normalization, noise removal, clipping and resampling techniques for comprehensive preprocessing. Data normalization makes the calculations feasible at model generation. The image input vector pixel values are normalized before training, by rescaling the values in range of 0-1using the Z-Score method. Noise reduction is another preprocessing technique, which eliminates the unnecessary information from the input images. Fan et al. [14] discussed a set of smoothing and filtering (Gaussian and Fuzzy based) techniques, which are employed in this 3d-CU-Net model for noise reduction from input images. In order to keep the MRI image pixel vector values in a consistent range and to eliminate the uninterested background from the images, the image clipping technique is applied. As part of clipping, the minimum and maximum thresholds are evaluated and the pixel ranges are clipped. Image resampling process is applied on input MRI image vectors to overcome the imaging geometry issues. Gurjar et al. [15] explored the efficient resampling methods, which works by transforming the original image orientation, resolution and size values. By using the aforementioned image synthesis and preprocessing techniques (shown in Figure 2), the proposed 3D-CU-Net model completes the preprocessing phase and provides the error free and reliable input data for training and testing process.

3D-CU-Net Model Configuration Phase: Soon after data preprocessing, the next phase of our model is 3D-CU-Net model design and configuration with respective parameters. Compared to the other medical object segmentation models, the kidney tumor segmentation (tumor pixels detection) and boundaries lineation is a complex and sensitive operation. Parvathi and Jonnadula [16] specified that, designing the kidney tumor segmentation model become a challenging research topic due to the objects overlapping, object detection ambiguity and morphological diversity issues. As on, some former researches concentrated on medical image segmentation process and implemented the U-Net architecture for medical image segmentation. Although the U-Net performed better segmentation than its former segmentation models, it’s still suffering from some intrinsic limitations at encoder-decoder networks are discussed in section-2.

To address the U-Net limitations in kidney tumor segmentation process, in this paper the U-Net architecture is customized as the 3D-CU-Net (is shown in Figure 3), to increase the prediction accuracy in 3D MRI kidney tumor segmentation. The proposed 3D-CU-Net model features are customized at encoders and decoders level, to support the 3D analysis and to achieve the high accuracy in segmentation process.

To overcome the variance in network depths issue, our 3D-CU-Net model is designed with a unified encoder-decoder network, which consists of various network depths (L1, L2, L3 and L4) in the same model is shown in Figure 3. The encoder-decoder nodes of this architecture with various network depths (L) are: L1={X(0,0), X(0,1)}, L2={X(0,0), X(0,2)}, L3={X(0,0), X(0,3)} and L4={X(0,0), X(0,4)}. In this model, all networks are sharing a common encoder (X(0,0)) to receive the training inputs but having their own decoders {(X(0,1)), (X(0,2)), (X(0,3)), (X(0,4))} to generate the output. According to the requirement in deep analysis, the appropriate network depth is selected while processing. In any deep network level from L1 to Ln, the connectivity node is (X(i,j)), in which the ‘i’ stands for the encoder index and ‘j’ stands for dense skip convolutions, are used to calculate the target feature maps. When the ‘j’ value is 0 means, convolution receives only single input from its former encoder path layers and feature maps are calculated as:

$$x^{i,j} = \lfloor \Delta(\omega(x^{i-1,j}) \rfloor \tag{3}$$

Similarly when the ‘*j*’ value is a non-zero means, it receives the two inputs from two immediate levels of the encoder path layers and the feature maps are calculated as:

$$x^{i,j} = \left\{ \Delta \left[x^{i,p} \right]_{p=0}^{j=1} \oplus \alpha \left(x^{i++,j--} \right) \right\} \tag{4}$$

where, the convolution (Δ) process is implemented by the respective activation functions in up-sampling layer ω and down sampling layer ∝ with concatenation (⨁) operator. The inputs from the former encoder paths and dense skip connections will use the appropriate activation functions with ‘P’ kernels for deep supervision through deep feature mapping process. Before to the output layers, the feature (background/foreground) relative probabilities of the input image voxels are determined using the soft-max layer.

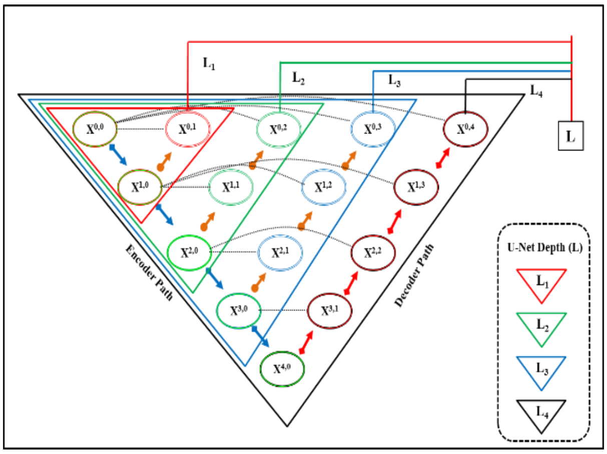

Figure 3. Unified 3D-CU-Net model with encoder-decoder networks to support the depth variance

In addition to the depth invariance problem, our model addresses the U-Net skip connections related enforced feature fusion issue (explained in section-2) also. To control the forced feature fusions, our 3D-CU-Net model is designed with a set of completely connected dense skip connections at each layer of the encoders and decoders is shown in Figure 4. In our 3D-CU-Net model, the dense skip connections are designed for complete connectivity among the deep layers of the architectures. Apart from the same level connections and completely connected dense networks, the descendent layers also connected via dense skip connections. Unlike the other U-Net models, 3D-CU-Net is completely connected, means each node in the network is having the direct connectivity with all others of same layer and with their symmetric descendants of the below layers too is shown in fugure-5. Due to the complete dense connectivity, the semantic gap is reduced between the layers. Our skip connection allows selection of any best feature node from either the same layer or the descendant layer. This facility enables the in-depth analysis of input data at network layers and performs the deep feature fusion at decoders, to extract the crucial features for foreground prediction. To overcome the complete dense connectivity model created additional burden in training, the feature rich nodes are tagged at each layer and the tagged nodes only considered for propagation in training. In this way, the 3D-CU-Net model addresses the slow pace of learning rate and controls the delays in model training too. Loss functions are designed specifically to find the training loss at 3D-CU-Net layers.

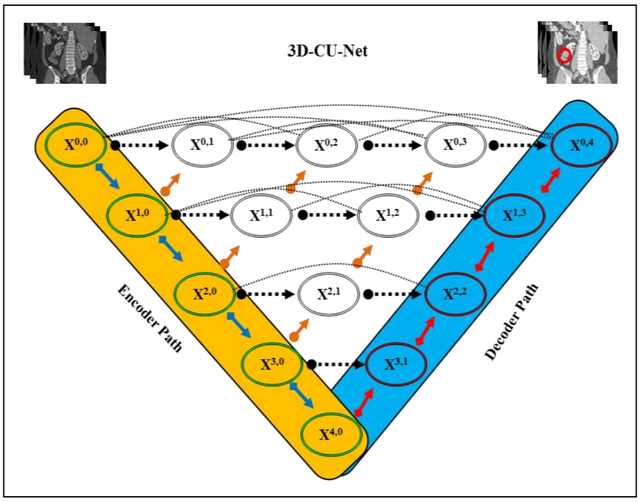

Figure 4. Graphical representation of the 3D-CU-Net process layers

Loss Function: While training and prediction process, the loss function is defined separately for each network (L1 to L4) at their decoders (X(0, n)) with a sigmoid activation function (S). In case of the kidney tumor segmentation process, the kidney tumors are generally very small (granules) in size and scattered over the kidney region. Because of the size invariance in tumor segmentation, finding the local minima in loss function is facing the ambiguity and biased towards the background voxels is specified by Yeung et al. 2022, Hashemi et al 2018. This biased mechanism may fail to detect the foreground (small tumors) part either partially or completely. To overcome the local minima finding issues of the loss function [17], the 3D-CU-Net integrated the Binary Cross Entropy (BCE) loss function and Soft-Dice Coefficient (SDC) loss function features together to find the local minima of the loss in predicted segmentations accurately. This integrated method generalizes the training and prediction loss by redesigning the weights of the foreground features. This integrated model of loss function is designed as:

$$L(y,\tilde{y}) = -\left(y(\log(\tilde{y}) + (1-y) * \log(1-\tilde{y}) \right) \tag{5}$$

In a binary cross entropy loss function $L(y, \tilde{y})$, the y is segmentation mask value and the $\tilde{y}$ is predicted segmentation value, which are joined later with the soft-dice coefficient function $D(y, \tilde{y})$ are:

$$D(y,\tilde{y}) = \left[1 - \frac{2\sum_{s} y * \tilde{y}}{\sum_{s} y^2 + \sum_{s} \tilde{y}^2} \right] \tag{6}$$

In order to balance the precision and recall values in loss function, the foreground voxel weights are redesigned and the integrated loss function $L * D(y, \tilde{y})$ is designed as:

$$L * D(y,\tilde{y}) = 1 - \frac{\sum_{i=1}^{n} p(f_i).p(b_i)}{\sum_{i=1}^{n} p(f_i).p(b_i) + \alpha \sum_{i=1}^{n} p(f_i).p(\tilde{b}_i) + \beta \sum_{i=1}^{n} p(\tilde{f}_i).p(b_i)} \quad (7)$$

where, the $p(f_i)$ is the probability of the predicted voxel foreground value where f_i=1 and $p(b_i)$ is the probability of the predicted voxel background value, where b_i=0. In contrary to this, the voxel actual values are reversed as $\tilde{f_i} = 0$ and $\tilde{b_i} = 1$ to cope up with the hyper parameters α and β of the loss function. In this size invariant tumor segmentation process, to leverage the computational efficiency and to handle the bias in background selection, the min-threshold (δmin) is defined. This threshold value limits the local minima, in which only the voxel contrast value is greater than the min-threshold (δmin) are considered for background separation to optimize the computational depth. Threshold approximation method defined by the Mayala and Haugsøen [18] for nucleus segmentation is adopted in our research to determine the min-threshold (δmin) value. Based on the image pixel intensity and their frequencies approximation from histograms, the min-threshold is evaluated in this research.

4. EXPERIMENTS

4.1 Training and testing

This section explores the proposed 3D-CU-Net model training and testing on kidney MRI tumor segmentation process. Segmentation accuracy is evaluated using the standard metrics and compared against the counterparts of the proposed segmentation model.

Dataset: To conduct the experiments on proposed 3D-CU-Net model, a set of 3D MRI kidney images are selected from TCGA-KIRC dataset [19]. This dataset doesn't contain any private information of data donors and is available in public domain for research analysis. As this dataset is collected from various locations, persons and devices, the dataset information is heterogeneous in image quality, modalities and metadata. The nature of the variance in dataset image properties helps to build the standard processing models, which can later process the data efficiently in real life applications.

Although this dataset contains the CT and MR image modalities, only the 3D MRI images are selected for training and testing operations. Apart from this some required clinical data forms (i.e. kidney case quality form) are made available with this dataset for researcher reference. Along with the radiology images, the patients MRI volume metadata contains the demographic information and some clinical parameters also. A set of abdominal MRI images of NifTI1 type (with .nii extension) with their associated ground truth (masks) images are selected for research analysis. Health professionals and radiology institution technicians performed the kidney tumor ground truth segmentation for training and testing purpose.

To conduct the experiments with TCGA-KIRC dataset, a prototype application is designed using the Python standard image processing libraries. Initially the NiBabel library functions are used in experiments for all read and write operations. After loading the MRI images and masks from system data paths, they are transformed into the Numpy arrays (3D Vectors) for further processing. A set of preprocess methods are applied on input vectors to prepare the input data for model training. As the selected dataset images are limited in count, a set of augmentation techniques like 3D-cGAN augmentation, Contrast enhancement, Noise injection, smoothing, skews and rotations are applied over the input dataset to increase the input data count 2x more (shown in Figure 5). Scikit Image and Numpy array library functions of python are widely used for input data preprocessing.

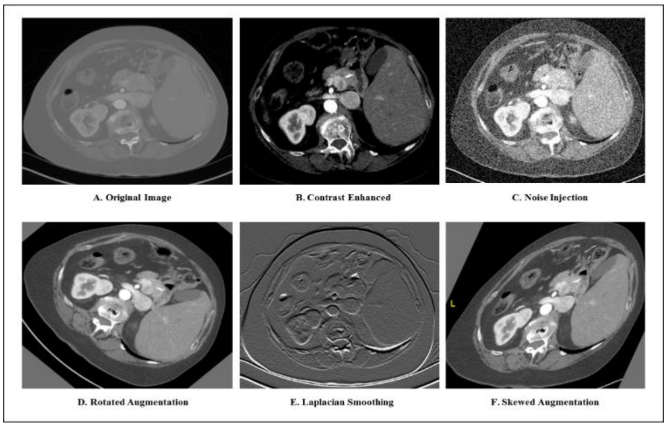

Figure 5. Presentation of various preprocessing techniques applied on kidney MRI images

After completion of the data preprocessing, an abstract deep learning 3D model for custom U-Net is created with Keras library, which is executing on TensorFlow platform. A set of completely connected dense networks are implemented using the concatenations immediate to the convolutions at encoders. The model (depth L4) with contracting layer, middle layer and expanding layers are generated using input the images, neurons and batch normalization functions. As the 3D-CU-Net model is compatible to handle all inputs with depth less than or equal to L4, depth invariance issues are normalized with same model and the completely connected dense skip connections assured the deep future fusions for crucial foreground features extraction.

4.2 Experimental evaluation

The total process of the experimental evaluation is classified into the three different phases are: Model Training, Prediction and Evaluation.

Model Training: The TCGA-KIRC dataset contained MRI kidney image slices and associated masks are shuffled and partitioned as training dataset (70-75%) and test dataset (25-30%). This training dataset images are splits into the batches (with size k) and are arranged into a pipeline for processing. By the time of the model creation, the initiation parameters like batch size, loss function, learning rate, metrics and other GPU settings are configured with respective values. After initializing with basic parameters, the proposed 3D-CU-Net model build function is compiled with weights, metrics, optimizers, loss functions and GPU mirror strategies. The build model is trained using an efficient fit function with the shuffled training data batches, epochs and callbacks. To avoid the class unbalance and over-fitting issues in model training, the weights and bias values are reassigned across the iterations of training process. For efficient memory management in training, a set of custom garbage collection functions are designed to clean up the temporary batch files and unusable data vectors in the middle.

Prediction: Soon after training the model with fitness function, the test dataset is given as input to the predict function for segmentation. Predict function with deep feature fusion capabilities performed the tumor segmentation from the test input data using the trained model knowledge. In prediction process, test dataset images from pipeline are iteratively loaded, segmented, post processed and persisted for future analysis. To ensure the proposed 3D-CU-Net prediction capabilities on data variance, a set of augmented (rotated) synthetic test images also segmented by the same model. Figure 6 is presenting the segmentation results obtained from the proposed 3D-CU-Net model.

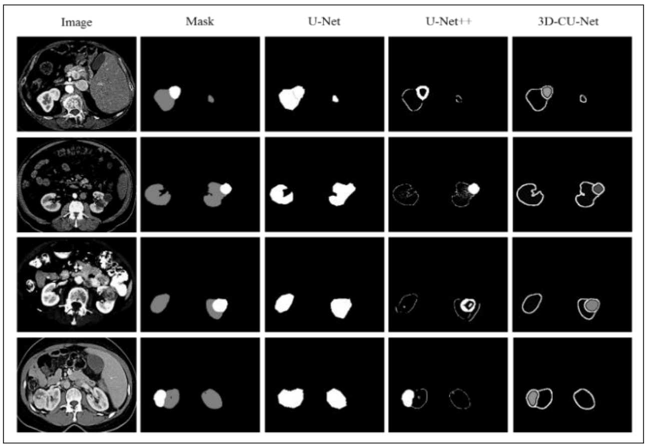

Figure 6. Comparison of 3D-CU-Net tumor segmentation results against the counterparts U-Net and U-Net++

Evaluation: In order to prove the accuracy of segmentation and efficiency of 3D-CU-Net model, experimental results are evaluated with the help of comparative analysis. For this comparison process, the popular deep convolutional networks U-Net and U-Net++ are selected as the counterparts to 3D-CU-Net. After the selected segmentation models training and validation process, the training loss and validation loss are evaluated to showcase the under fitting or over fitting issues. Training and validation functions of these models are set with the gradient descendent minimum learning improvement value as Δ≥0.001, input feature coefficients values as α=0.7 and β=0.3, while training and validation. To perform the hyper parameter tuning process and to find the best fit values for Δ, α and β in training, we adopted the Adam optimization algorithm [20], which is consumes less memory and computational resources for estimation.

Figure 7 is presenting the 3D kidney MRI image training and validation loss values obtained from the loss functions of U-Net with BCE, U-Net++ with SDC and 3D-CU-Net with SDC and CE. The mean value of the validation minimum loss values are U-Net (±0.146), U-Net++ (±0.098) and 3D-CU-Net (±0.091). Due to the efficient training and prediction methods, our 3D-CU-Net model recorded less error rate (loss) than its counterparts will improve the prediction accuracy.

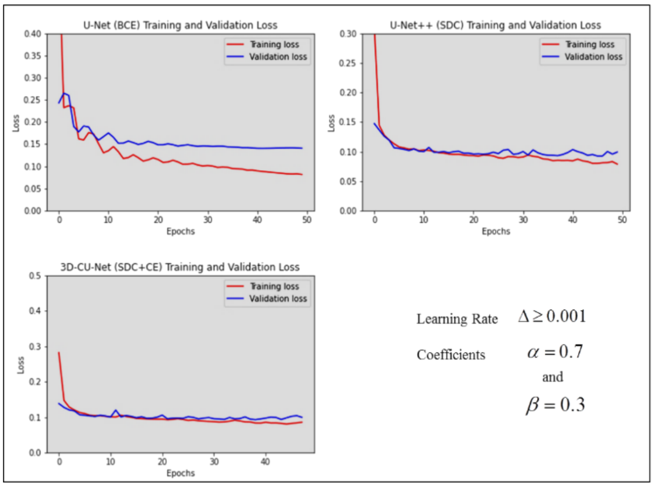

Figure 7. Comparative analysis of the training and validation loss in 3D MRI kidney tumor segmentation

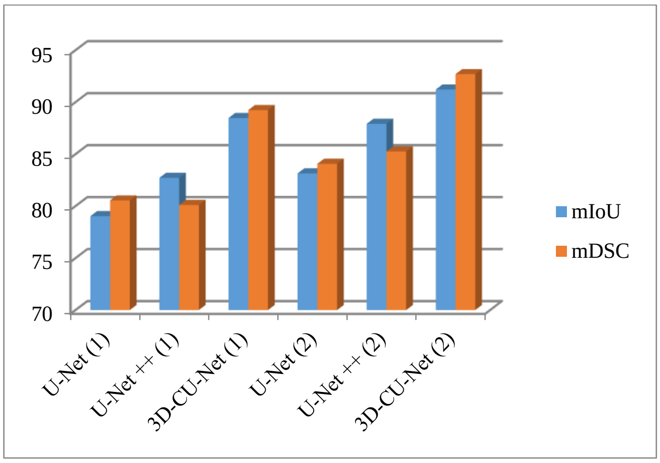

Figure 8. Comparison of the Kidney tumor segmentation accuracy using various deep learning models

To conduct the extensive analysis on the selected deep segmentation models, the additional features (i.e. Deep fusion and Dense Skip Connections) are joined in experiments to generate more combinations. Intersection over Union (mIoU) and Dice Similarity Coefficient (mDSC) metrics are selected to calculate the similarity between prediction samples and their associated ground truth labels (masks). Finally, the mean value of these similarity vectors is considered as the prediction result accuracy value. Table 1 is presenting the MRI kidney tumor segmentation accuracy results obtained from various deep learning models with different feature combinations [21-24].

Deep feature fusion with depth invariant unified architecture and completely connected dense skip connections of the 3D-CU-Net helped to record the highest segmentation accuracy with mIoU(92.21) and mDSC(92.69) is presented in Figure 8.

Table 1. U-Net segmentation models prediction accuracy comparison using the metrics mIoU and mDSC

Seg Model	Params	Deep fusion	Dense Skip	Loss Model	mIoU	mDSC
U-Net	7.2M	N	N	BCE	79.03	80.54
U-Net ++	7.7M	N	N	SDC	82.72	80.11
3D-CU-Net	8.2M	N	N	SDC+CE	88.47	89.23
U-Net	7.2M	Y	N	BCE	83.14	84.07
U-Net ++	7.7M	Y	N	SDC	87.91	85.25
3D-CU-Net	8.2M	Y	Y	SDC+CE	91.21	92.69

5. CONCLUSIONS

Deep learning models are playing a vital role in medical image analysis and disease diagnosis. In recent times, U-Net emerged as a popular deep network model for kidney tumor segmentation from 3D MRI images. In this study, the deep learning model U-Net limitations in medical image segmentation are thoroughly discussed. To address the U-Net limitations in medical image segmentation process, 3D-CU-Net model is proposed with the encoder-decoder networks customization. 3D-cGAN and other preprocessing techniques are applied on input dataset for data augmentation and preprocessing. Unified deep network architecture is designed with depth invariant encoders and decoder to process various input images using the same model. A set of completely connected dense connections are designed to avoid the enforced feature fusion in model training. In depth analysis on input data extracts the crucial features, which helps in efficient foreground lineation. BCE and DSC loss function features are integrated to evaluate the pixel level segmentation accuracy between images and masks. TCGA-KIRC dataset is selected for experiments and python libraries are used for model implementation. Comparative analysis of experimental results is proven that, our proposed 3D-CU-Net model recorded the high accuracy in kidney tumor segmentation compared to its counterparts.

REFERENCES

[1] Glaser, C., D'Anastasi, M., Theisen, D., Notohamiprodjo, M., Horger, W., Paul, D., Horng, A. (2015). Understanding 3D TSE sequences: advantages, disadvantages, and application in MSK imaging. In Seminars in Musculoskeletal Radiology, 19(4): 321-327. https://doi.org/10.1055/s-0035-1563732

[2] Parvathi, S.S., Jonnadula, H. (2021). A comprehensive survey on medical image blob detection and classification models. In 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), pp. 1-6. https://doi.org/10.1109/ICAECA52838.2021.9675575

[3] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[4] Ciompi, F., de Hoop, B., van Riel, S.J., Chung, K., Scholten, E.T., Oudkerk, M., de Jong, P.A., Prokop, M., van Ginneken, B. (2015). Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. Medical Image Analysis, 26(1): 195-202. https://doi.org/10.1016/j.media.2015.08.001

[5] Dietterich, T.G. (2000). Ensemble methods in machine learning. In Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings 1, pp. 1-15. https://doi.org/10.1007/3-540-45014-9_1

[6] Zheng, S., Song, Y., Leung, T., Goodfellow, I. (2016). Improving the robustness of deep neural networks via stability training. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4480-4488. https://doi.org/10.1109/CVPR.2016.485

[7] Kumar, P., Nagar, P., Arora, C., Gupta, A. (2018). U-segnet: fully convolutional neural network based automated brain tissue segmentation tool. In 2018 25th IEEE International conference on image processing (ICIP), pp. 3503-3507. https://doi.org/10.1109/ICIP.2018.8451295

[8] Zhang, J., Jin, Y., Xu, J., Xu, X., Zhang, Y. (2018). Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation. arXiv preprint arXiv:1812.00352.

[9] Mirza, M., Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784. https://arxiv.org/pdf/1411.1784.pdf

[10] Wang, Y., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., ... & Zhou, L. (2018). 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. Neuroimage, 174: 550-562. https://doi.org/10.1016/j.neuroimage.2018.03.045

[11] Perumal, S., Velmurugan, T. (2018). Preprocessing by contrast enhancement techniques for medical images. International Journal of Pure and Applied Mathematics, 118(18): 3681-3688.

[12] Shorten, C., Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1): 1-48. https://doi.org/10.1186/s40537-019-0197-0

[13] Kalra, A., Stoppi, G., Brown, B., Agarwal, R., Kadambi, A. (2021). Towards rotation invariance in object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3530-3540.

[14] Fan, L., Zhang, F., Fan, H., Zhang, C. (2019). Brief review of image denoising techniques. Visual Computing for Industry, Biomedicine, and Art, 2(1): 1-12. https://doi.org/10.1186/s42492-019-0016-7

[15] Gurjar, S.B., Padmanabhan, N. (2005). Study of various resampling techniques for high-resolution remote sensing imagery. Journal of the Indian Society of Remote Sensing, 33: 113-120. https://doi.org/10.1007/BF02989999

[16] Parvathi, S.S., Jonnadula, H. (2022). An efficient and optimal deep learning architecture using custom U-Net and mask R-CNN models for kidney tumor semantic segmentation. International Journal of Advanced Computer Science and Applications, 13(6): 314-320.

[17] Salehi, S.S.M., Erdogmus, D., Gholipour, A. (2017). Tversky loss function for image segmentation using 3D fully convolutional deep networks. In Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, pp. 379-387. https://doi.org/10.1007/978-3-319-67389-9_44

[18] Mayala, S., Haugsøen, J.B. (2022). Threshold estimation based on local minima for nucleus and cytoplasm segmentation. BMC Medical Imaging, 22(1): 1-12. https://doi.org/10.1186/s12880-022-00801-w

[19] Akin, O., Elnajjar, P., Heller, M., Jarosz, R., Erickson, B. J., Kirk, S., Lee, Y., Linehan, M.W., Gautam, R., Vikram, R., Garcia, K.M., Roche, C., Bonaccio, E., Filippini, J. (2016). The cancer genome atlas kidney renal clear cell carcinoma collection (TCGA-KIRC) (Version 3). The Cancer Imaging Archive.

https://doi.org/10.7937/K9/TCIA.2016.V6PBVTDR

[20] Jais, I.K.M., Ismail, A.R., Nisa, S.Q. (2019). Adam optimization algorithm for wide and deep neural network. Knowledge Engineering and Data Science, 2(1): 41-46. https://doi.org/10.17977/um018v2i12019p41-46

[21] Li, S., Song, W., Qin, H., Hao, A. (2018). Deep variance network: An iterative, improved CNN framework for unbalanced training datasets. Pattern Recognition, 81: 294-308. https://doi.org/10.1016/j.patcog.2018.03.035

[22] Yeung, M., Sala, E., Schönlieb, C.B., Rundo, L. (2022). Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Computerized Medical Imaging and Graphics, 95: 102026. https://doi.org/10.1016/j.compmedimag.2021.102026

[23] Hashemi, S.R., Salehi, S.S.M., Erdogmus, D., Prabhu, S.P., Warfield, S.K., Gholipour, A. (2018). Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection. IEEE Access, 7: 1721-1735. https://doi.org/10.1109/ACCESS.2018.2886371

[24] Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., Pal, C. (2016). The importance of skip connections in biomedical image segmentation. In International Workshop on Deep Learning in Medical Image Analysis, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 179-187. https://doi.org/10.1007/978-3-319-46976-8_19