International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Quad-YOLOv5: Improved YOLOv5 for Liver Lesion Detection on Bio Medical CT Images

S. Komal Kour[1], Dr. T Adilakshmi[2]

Submitted: 27/01/2024 **Revised**: 05/03/2024 **Accepted**: 13/03/2024

Abstract: The liver is an important organ in the human body, performing numerous vital functions that are essential for overall health and well-being. Accurately identifying lesions in medical CT scans has long been one of the most difficult problems in the field of medical image analysis. Appropriate treatment and management strategies can be implemented to address the underlying liver condition and optimize patient outcomes. The proposed method for Liver Lesion Detection, utilizes the DeepLeison dataset which contains many biomedical CT scan images with a variety of liver pathologies. The method proposes Quad-YOLOv5 which is based on popular driven object detection deep learning model called YOLOV5 (You Only Look Once) model. We build up the medical image dataset of Liver Lesion by collecting 6335 CT images by augmentation. To enhance the performance of the Quad-YOLOv5 model, we have implemented data augmentation techniques and conducted extensive experiments using the DeepLeison dataset. Our findings demonstrate that the model exhibits strong performance coupled with remarkable interpretability. Through meticulous experimentation, we have refined the model's capabilities, ensuring that it delivers superior results in lesion detection tasks while maintaining a high level of interpretability. Our model is trained and evaluated based on its performance. The precision and recall for the model are 96% and 93%. It is obtained that Quad-YOLOv5 model is with Mean Average Precision (mAP50) of 97%. The model increases the efficiency and accuracy of diagnosing and treating liver lesions. It can be incorporated into existing clinical workflows to aid radiologists in the interpretation of CT scans.

Keywords: *- Quad-YOLOV5, Deep Learning, DeepLesion, Lesion Detection, CT Images, Localization*

Introduction

The liver is one of the most crucial organs in the human body, performing a variety of critical functions. Accurate detection of the Liver Lesion from medical imaging scans is crucial for diagnosis and for treatment. However manual detection of the liver lesion is a laborious and time taken job and the task is to develop an automated liver detection algorithm using deep learning techniques, which can accurately and robustly classify and detect the liver lesion from medical imaging scans such as CT scan. The algorithm should be able to handle variations in image quality, patient anatomy, and imaging modality, and should be computed on a huge collection of datasets of medical imaging scans. The main goal is to execute high accuracy, precision, recall, and Probability similarity scores for liver lesion detection. The algorithm trained is scalable and efficient and can process large volumes of medical imaging data in a reasonable amount of time. Adaptive threshold method was used to isolate the liver from the rest of the body, and spatial fuzzy clustering was used to segment the cancerous lesions in the liver. The informative characteristics were extracted from the segmented cancerous region and classified into two categories of liver cancers using two different classification algorithms i.e., multilayer perceptron and C4.5 classifiers and had a comparative study [4]. The Fully Convolutional Network model (FCN) was to enhance segmentation by incorporating a self-supervised contour-guiding mechanism. This pioneering approach amalgamated shape and contour characteristics to achieve precise delineation of the target object. Notably, the network adeptly learned contour features to demarcate the complementary contour region through a self-supervising framework [8]. Leveraging domain knowledge in medical imaging data can greatly enhance the development of robust lesion detection networks. DKMA-ULD, a novel framework, aims to detect lesions across multiple organs with heightened sensitivity compared to existing methods. By integrating insights from medical expertise into its design, DKMA-ULD surpasses current state-of-the-art techniques. This approach allows DKMA-ULD to adapt to diverse imaging modalities and clinical scenarios, enabling comprehensive lesion detection with improved accuracy and reliability [10].

The use of deep learning algorithms for medical image localization tasks has gained significant attention in recent years. Previous studies have utilized various deep learning algorithms such as Convolutional Neural Networks (CNNs) for liver lesion classification. However, there are still challenges in achieving accurate liver segmentation and lesion detection due to factors such as variations in lesion size, shape, and appearance.

[1]*Research Scholar, CSE Department, University College of Engineering (UCE), OU*
[1, 2] *Department of Computer Science and Engineering, Vasavi College of Engineering, Hyderabad, India*
komalkour@staff.vce.ac.in[1], t_adilakshmi@staff.vce.ac.in[2]

LITERATURE SURVEY

Few related works done on Lesion Detection and Segmentation using deep learning approaches are briefly discussed. A weakly supervised segmentation method has been devised to efficiently transform extensive collections of RECIST-based lesion diameter measurements, archived within hospitals' digital repositories, into comprehensive 3D lesion volume segmentations and measurements. This approach, while simple in its implementation, yields remarkable efficacy [3]. Liver lesion segmentation in CT scans serves various critical purposes such as quantifying tumor burden, devising treatment strategies, forecasting clinical responses, and monitoring progression. To tackle this challenge, a Hybridized Fully Convolutional Neural Network (HFCNN) has been introduced specifically for liver tumor segmentation and detection. This model offers a promising approach to address the pressing issue of liver cancer. [7]. To obtain accurate lesion structure, the segmentation was performed on high contrast CT scan images and Liver extraction using adaptive thresholding [5]. An innovative semi-automatic RECIST labeling technique employs a cascaded Convolutional Neural Network (CNN) architecture, which consists of an improved Spatial Transformer Network (STN) and Spatial Hierarchical Network (SHN). Enhancements to the STN include multi-task learning and the integration of a supplementary coarse-to-fine pathway to enhance the accuracy of transformation parameter prediction [2]. The proposed method for lesion segmentation got 75.2% of Dice using ensemble method [9]. Liver segmentation by U net and multi-scale candidate generation method to obtain the blocks. Active contour model (ACM) is used to refine the tumour segmentation [6]. The 3D Context Enhanced Region-Based CNN (3DCE) is designed to harness the 3D context for lesion detection in volumetric data. Its implementation consistently enhances detection accuracy on the Deep Lesion dataset [1]. The proposed method RCNN got 97.4% of Dice for SLIVER07 dataset and 96.55% of Dice for 3Dircadb dataset [11]. The proposed paper used DefED-Net for Dice Coefficient of liver and liver segmentation with 96.30 and 87.52 respectively [12].

Previous studies have demonstrated the potential of deep learning algorithms for liver lesion segmentation in medical images. However, deep learning-based detection methods still have room for improvement, and more research is needed to overcome their limitations and challenges. One major challenge is the requirement of a large amount of annotated data for model training, which can be time-consuming and expensive. Another challenge is the need for careful selection and tuning of model hyper parameters to achieve optimal performance.

PROPOSED METHOD

Deep learning techniques have demonstrated considerable utility in clinical settings and have achieved notable successes. Despite this, there has been a scarcity of deep learning methods specifically tailored for the identification and classification of liver lesions on CT images. This gap arises from the absence of openly accessible CT image datasets focused on liver lesions for training and validating these models. Consequently, there is a growing interest in leveraging deep learning methodologies to address the challenges associated with identifying and classifying liver lesions on CT images.

3.1 DeepLesion

The DeepLesion Dataset by NIH is a large-scale, multi-institutional dataset of radiology studies, consisting of over 32,000 CT studies from more than 10,000 unique patients. The dataset is primarily focused on the detection and classification of lesions within the body, and contains a wide range of lesion types, including lung nodules, liver lesions, bone tumours, and more. One of the key features of DeepLesion is its large size and diversity. This dataset is specifically focused on lesions within the thoracic and abdominal regions. The radiologist manually verifies the CT images and its fatigue for them to predict the lesion and its location. The dataset was created from different institutions, to determine location of lesions.

3.2 Data Visualization

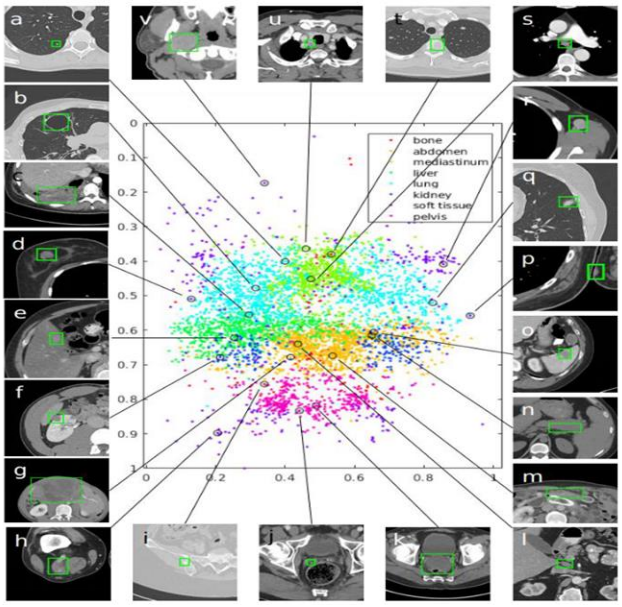

Figure 1. - Data Visualization of DeepLesion

A portion comprising 15% of the DeepLesion dataset is depicted in visual form. The scatter map illustrates lesion locations, with the x- and y-axes representing the x- and z-coordinates of each lesion's relative position within the body, as indicated in Figure 1. DeepLesion encompasses a vast dataset containing 32,735 lesions distributed across 32,120 CT slices derived from 10,594 studies involving 4,427 individual patients. Each CT image typically

contains 1 to 3 lesions, each accompanied by bounding boxes and size measurements, totaling 32,735 lesions overall. The dataset encompasses a diverse array of lesion types, including those related to bone, abdomen, mediastinum, liver, lung, kidney, soft tissue, and pelvis, among others. From 32,120 CT slices, we've identified and extracted 1,284 CT images featuring the liver and liver lesion. Subsequently, we augmented these images using a variety of techniques, including noise injection, translation, brightness adjustment, and contrast enhancement. Furthermore, adjustments were made to the lesion coordinates respective to their augmentation during this process. Figure 2 represents the Liver CT Images with bounding boxes shown the lesion position.

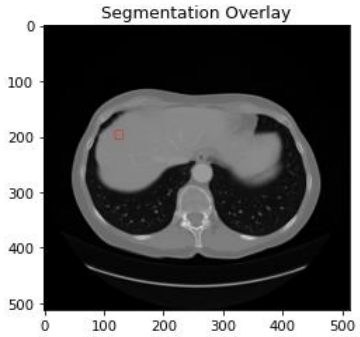

Figure 2: CT image represents the Liver Lesion Segmentation

Dataset. The experimental data set used here is obtained as follows:

a. Collecting 1284 CT images with livers lesion from 32,120 CT slices.

b. Among the 1284 CT images, 17 images are with same image name and with two lesion bounding boxes, it's difficult for training. Hence, removed those images and considered 1267 CT images for Augmenting and training a model.

c. With the Liver CT images, augmented the images with Noise, Contrast, Translate and Brightness. Total merged Dataset have 6335 CT images.

d. 6335 CT images were labelled with bounding box of the lesion position in csv file, which consists of all the augmented data with image names in a column File_name and along with their lesion coordinates in a column Bounding_boxes.

e. Dataset and labelled coordinates in a CSV file were used. Among them 80% of data used to train the Quad-YOLOv5 model and computed the precision, recall and mAP50 (Mean Average Precision).

Quad-YOLOv5 Model

Overview of YOLOv5 Model

YOLOv5 represents a notable leap forward in object detection technology, building upon the success of its predecessor, the YOLO (You Only Look Once) model, renowned for its real-time performance and accuracy. This latest iteration introduces significant enhancements aimed at further improving real-time object detection capabilities. Key components include CSP-Darknet53 (Cross Stage Partial Network) as the backbone, Spatial Pyramid Pooling (SPP), and Path Aggregation Network (PANet) in the model's neck, along with the head architecture from YOLOv4.The advancements of YOLOv5 are grounded in addressing the challenges associated with information loss in deep neural networks. The Information Bottleneck Principle and the innovative utilization of Reversible Functions are pivotal to its design, ensuring that YOLOv5 maintains high efficiency and accuracy while overcoming limitations inherent in previous architectures. This amalgamation of cutting-edge techniques propels YOLOv5 to the forefront of real-time object detection, setting new benchmarks in performance and capability.

CSP-Darknet53

Quad YOLOv5 uses CSP-Darknet53 as its backbone. Darknet-53 is a convolutional neural network architecture used as a backbone or feature extractor in various computer vision tasks, particularly in object detection models like YOLOv3. It consists of 53 convolutional layers and is known for its ability to capture high-level features from CT images effectively. Developed as part of the Darknet framework, Darknet-53 serves as a robust feature extractor, enabling more accurate and efficient object detection compared to earlier versions of YOLO. YOLO, a deep network, employs residual and dense blocks to facilitate information flow to its deepest layers and mitigate the issue of vanishing gradients. However, the utilization of dense and residual blocks can lead to redundant gradients. CSPNet addresses this challenge by truncating the gradient flow, thereby aiding in gradient optimization.

Spatial Pyramid Pooling (SPP)

The SPP block aggregates input information and outputs a fixed-length representation, as depicted in Figure 3. This design offers the benefit of substantially expanding the receptive field and isolating crucial contextual features, all while maintaining network speed.

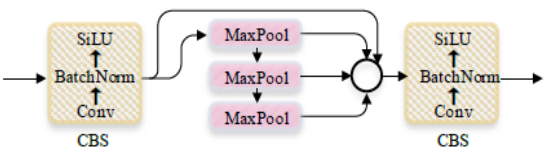

Figure 3: Structure of the SPP block.

Path Aggregation Network (PANet)

PANet serves as a feature pyramid network designed to aid in precise pixel localization for mask prediction tasks.

Quad-YOLOv5

The framework of Quad-YOLOv5 is depicted in Figure 4. We have tailored the original YOLOv5 architecture to specialize in handling the DeepLesion dataset. Through training our model on the DeepLesion dataset [12] with a carefully crafted data augmentation strategy. Our model's key enhancement lies in the addition of an extra prediction head, allowing it to effectively handle lesions of varying sizes. This four-head structure, combined with the existing prediction heads, enhances the model's ability to localize and detect lesion positions accurately. Quad-YOLOv5 is implemented using PyTorch 2.2.1, and model training and testing are conducted on NVIDIA GPUs. During the training phase, we utilize parts of the pre-trained YOLOv5s model, as Quad-YOLOv5 shares most of its backbone (blocks 0 to 8) and some portions of its head (blocks 10 to 13 and blocks 15 to 18) with YOLOv5. Additionally, we incorporate one additional head into our model architecture. In the realm of object detection, YOLO plays a crucial role as a one-stage detector. In this paper, we introduce an enhanced model, Quad-YOLOv5, which builds upon the foundation of YOLOv5 to offer improved performance in lesion detection tasks. The detection pipeline of Quad-YOLOv5 is illustrated in Figure 4. In this architecture, we employ CSPDarknet53 as the backbone and the path aggregation network (PANet) as the neck, following the original design. In the head part, we introduce an additional head specifically for detecting tiny objects. Quad-YOLOv5 comprises four detection heads, each dedicated to lesion detection, thus enhancing its capability to accurately identify lesions across various sizes.

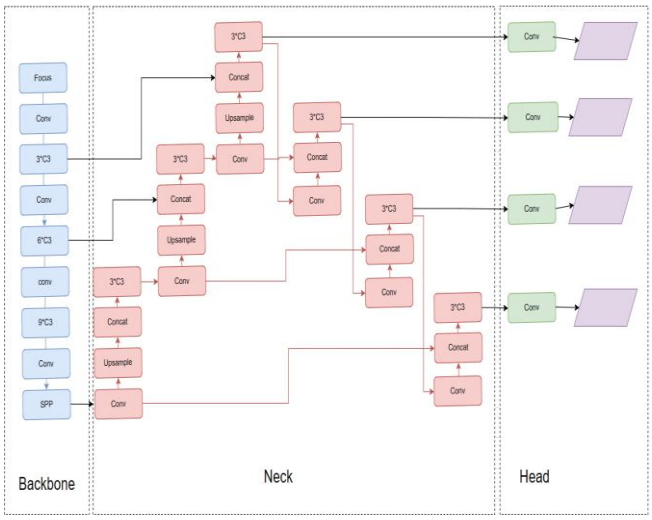

Figure 4: Quad-Yolov5 Architecture

IMPLEMENTATION & RESULTS

While implementing the Quad-yolov5 model the image input size was set to 640x640, with an learning rate of 0.01. The training utilized the SGD optimization algorithm, with a batch size of 16. The merged dataset of 6335 CT images were trained on Quad-YOLOv5 model with labelled data with 80% of training and 20% is for testing. The modelled trained for 100 epoch with precision of 96 % and recall of 93%. Since there are several categories to identify the accuracy, i.e., mean Average precision mAP50 is a well criterion to evaluate the result of lesion recognition is of 97% for our model. The equation to compute the target coordinates b_x, b_y, b_w and b_h for the bounding boxes is shown below.

$$b_x = (2*\sigma(t_x) - 0.5) + c_x$$

$$b_y = (2* \sigma(t_y) - 0.5) + c_y$$

$$b_w = p_w * (2* \sigma(t_w))^2$$

$$b_h = p_h * (2* \sigma(t_h))^2$$

The bounding boxes are detected and computed the Loss function. t_x,t_y are the center point of the bounding box and t_w, t_h are the height and width of the bounding box. c_x, c_y is the grid scaled by grid width and height. p_w, p_h are anchor width and height. Quad-YOLOv5 provides three outputs: the detected object classes, their corresponding bounding boxes, and the objectness scores. To calculate the loss, it employs the Binary Cross Entropy (BCE) method for both the classes and objectness components. Classes Loss measures the error for the classification task. The Objectness Loss evaluates the discrepancy in determining the presence of an object within a specific grid cell. Meanwhile, the CIoU (Complete Intersection over Union) loss is employed to gauge the localization error, assessing how accurately the object is positioned within the grid cell. In object detection models, each bounding box prediction comprises a confidence score indicating the model's certainty regarding the presence of an object of interest within the bounding box. The confidence threshold is a value between 0 and 1, and any bounding box with a confidence score below this threshold is discarded. Setting a higher confidence threshold typically results in fewer but more reliable detections. The final loss is determined by the following equation:

$$\text{Loss} = \boldsymbol{\lambda}_1 L_{cls} + \boldsymbol{\lambda}_2 L_{obj} + \boldsymbol{\lambda}_3 L_{loc}$$

Environments:

Quad-YOLOv5 run in the following verified environments (with all dependencies and python preinstalled packages)

1. Notebooks with GPU –
   a. Kaggle with RAM utilization of 14.8 GB and GPU Memory utilization of 15 GB and Disk space of 5.7 GB.

b. Colab Pro with GPU Memory utilization of 22 GB, Disk Space of 38 GB.

Results:

Few samples of detected lesion position are marked and representing the prediction score of each CT image. The Loss function is shown in below fig. representing the loss and precision, recall, Mean Average Precision (mAP50) is computed at an Intersection over Union (IoU) threshold of 0.5, mAP95 is computed at an IoU threshold of 0.95. If the IoU between a predicted bounding box and a ground truth bounding box exceeds a certain threshold, the predicted bounding box is considered a true positive. Otherwise, it's considered a false positive. Figure 5 showing the performance of Quad-Yolov5 model for 65 epochs and Figure 6 show the detection of lesion and its score.

Metrics	YOLOv5 Model (50 Epoch)	Quad-YOLOv5 Model (50 Epoch)	Quad-YOLOv5 Model (65 Epoch)	Quad-YOLOv5 Model (80 Epoch)	Quad-YOLOv5 Model (100 Epoch)
train/box_loss	0.0282	0.0250	0.0250	0.0230	0.0212
train/obj_loss	0.0083	0.0075	0.0073	0.0068	0.0064
val/box_loss	0.0268	0.0218	0.0216	0.0198	0.0180
val/obj_loss	0.0055	0.0046	0.0045	0.0042	0.0039
Precision	0.83	0.90	0.91	0.95	0.96
Recall	0.84	0.87	0.89	0.91	0.93
mAP50	0.89	0.94	0.94	0.95	0.97
mAP95	0.59	0.67	0.72	0.76	0.80

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Mean Average Precision} = 1/N \sum_{i=1}^{N} AP_i$$

$$\text{IOU} = \frac{\text{Intersection}}{\text{Union}}$$

Evaluation Metrics:

The model's accuracy is evaluated based on various metrics, each measuring distinct aspects of its performance. Precision quantifies the proportion of predictions made by the model that are accurate, while recall evaluates the percentage of relevant data points correctly identified by the model. Table 1 represents the comparative study of original YOLOv5 Model and our proposed Quad-Yolov5 Model.

Table 1- Evaluation Metrics for Quad-YOLOv5 Model

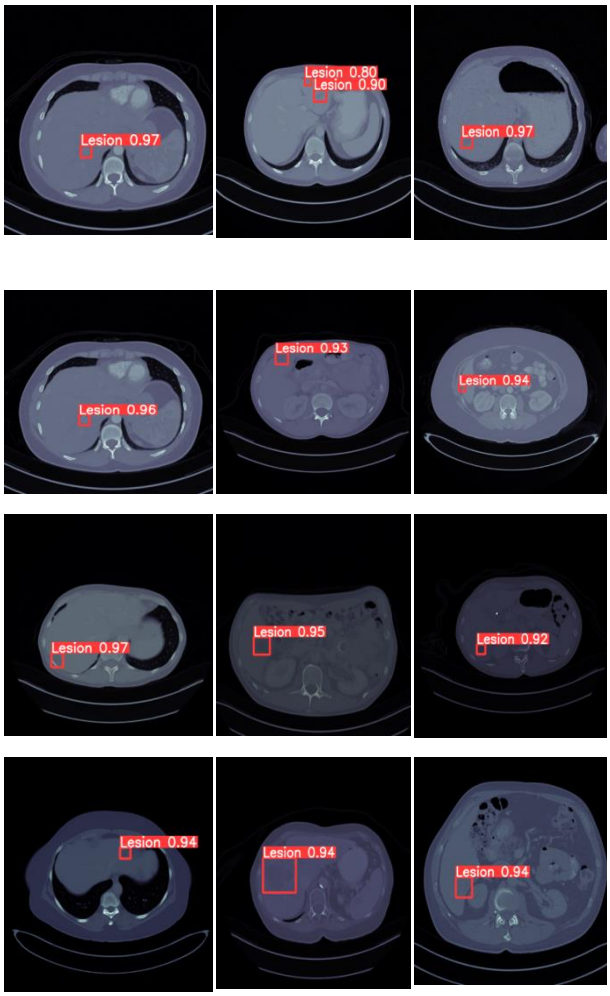

Figure 6: Liver Lesion detection in Quad- YOLOV5 Model

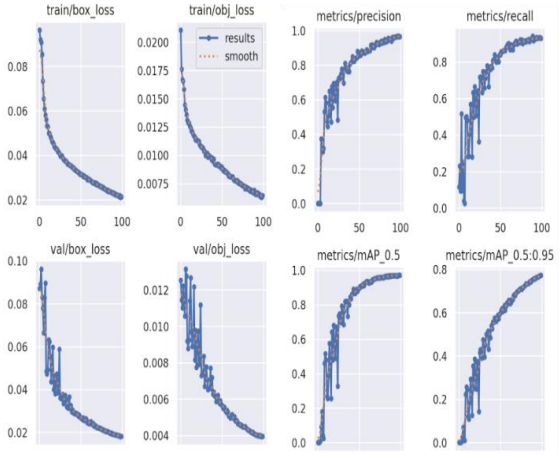

Figure 5: Represents the Loss, Precision, Recall and mAP for Quad-Yolov5 Model

CONCLUSION AND FUTURE SCOPE

Liver Lesion Detection is a crucial step in liver disease diagnosis and treatment planning. In this paper, we have

explored the performance of the Yolov5 architecture on the DeepLesion dataset for liver lesion detection. We have pre-processed the dataset by resizing the images, augmented images and splitting them into training and validation sets. We evaluated the performance of our model using standard evaluation metrics, such as Precision & Recall and visualized the predicted liver lesion regions. Our results suggest that our model performs well on cases with clear liver lesion boundaries with a minimum loss. The model can also be adapted for detection of other organs lesions or other medical imaging modalities. Future work can focus on further improving the accuracy and generalization of our model, as well as exploring its applications in other medical imaging tasks.

References

[1] Ke Yan, Mohammadhadi Bagheri, Ronald M. Summers, "3D Context Enhanced Region-based Convolutional Neural Network for End-to-End Lesion Detection", MICCAI, 2018.

[2] Youbao Tang, Adam P. Harrison, Mohammadhadi Bagheri, Jing Xiao, Ronald M. Summers,"Semi-Automatic RECIST Labeling on CT Scans with Cascaded Convolutional NeuralNetworks", MICCAI, 2018.

[3] Jinzheng Cai*, Youbao Tang*, Le Lu, Adam P. Harrison, Ke Yan, Jing Xiao, Lin Yang, RonaldM. Summers, "Accurate Weakly-Supervised Deep Lesion Segmentation using Large-ScaleClinical Annotations: Slice-Propagated 3D Mask Generation from 2D RECIST", MICCAI,2018.

[4] Amita Das, Priti Das, S. S. Pandaand Sukanta Sabut "Detection of Liver Cancer Using Modified Fuzzy Clustering and Decision Tree Classifier in CT Images" Pattern Recognition and Image Analysis, 2019, Vol. 29, No. 2, pp. 201–211. © Pleiades Publishing, Ltd., 2019.

[5] Ke Yan, Yifan Peng, Veit Sandfort, Mohammadhadi Bagheri, Zhiyong Lu, and Ronald M.Summers, "Holistic and Comprehensive Annotation of Clinically Significant Findings onDiverse CT Images: Learning from Radiology Reports and Label Ontology," CVPR, 2019.

[6] Zhiqi Bai, Huiyan Jiang, Siqi Li, And Yu-Dong Yao, "Liver Tumor Segmentation Based On Multi-Scale Candidate Generation And Fractal Residual Network" ,June, 2019, Volume 7, 2019, Ieee Transaction.

[7] Xin Dong, Yizhao Zhou, Lantian Wang, Jingfeng Peng,Yanbo Lou,And Yiqun Fan "Liver Cancer Detection Using Hybridized Fully Convolutional Neural Network Based On Deep Learning Framework" Ieee Access ,Special Section On Deep Learning Algorithms For Internet Of Medical Things, July 24, 2020, Natural Science Foundation of Zhejiang Province under Grant LY17H160024.

[8] M. Chung, J. Lee, M. Lee, J. Lee, and Y.-G. Shin, ''Deeply self-supervised contour embedded neural network applied to liver segmentation,'' Comput. Methods Programs Biomed., vol. 192, Aug. 2020, Art. no. 105447