International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Liver Lesion Detect SSD: Liver Lesion Detection with Single Shot Detector Utilizing ResNet-50 and ResNet-34 Backbones

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

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: Liver Lesion detection is a critical task in medical imaging, essential for early diagnosis and treatment planning, requiring high accuracy and efficiency to ensure effective patient care. This paper presents a robust approach to liver lesion detection using a Single Shot Detector (SSD) framework with ResNet-50 and ResNet-34 as a backbone networks. The SSD model, known for its object detection capabilities, is enhanced by the deep feature extraction power of ResNet architectures, enabling precise and fast identification of liver lesions. We comprehensively evaluate the performance of the proposed method on the DeepLesion dataset from the National Institutes of Health, demonstrating significant improvements in detection accuracy compared to conventional approaches. The ResNet-34 backbone achieves an impressive Mean Average Precision (mAP) of 91%, with ResNet-50 achieving 85%, demonstrating their effectiveness in accurately localizing liver lesions. The ResNet-50 and ResNet-34 backbones provide a balanced trade-off between computational complexity and detection accuracy, making the model suitable for practical medical applications. This research showcases the potential of SSD with ResNet architectures for accurate and efficient liver lesion detection, offering promising advancements for clinical applications.

Keywords: *LiverLesionDetect SSD, Single Shot Detection, Liver Lesion Detection, CT Images, mean Average Precision, ResNet backbone Model*

1. 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. It is a difficult task and time consuming to identify the cancer tissue manually in the present scenario. Viral hepatitis alone causes 1.34 million deaths every year. Problems with liver patients are not easily discovered in an early stage as it will be functioning normally even when it is partially damaged. An early diagnosis of liver problems will increase patient's survival rate. 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 detect the liver lesion from medical imaging scans such as CT scan. The detection of liver lesions in CT images can be used to assess the lesion, plan treatment based on the prediction. The goal is to achieve high accuracy, precision, recall, and mAP (mean average precision) for liver detection, while also minimizing false positives and false negatives. The algorithm trained is scalable and efficient and be able to process large volumes of medical imaging data in a reasonable amount of time. The use of deep learning algorithms for medical image detection tasks has gained significant attention in recent years. A Deep Learning approach is to develop a detection and localization of liver lesion.

[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]

2. Literature Survey

There are some advance research going on this bio medical imaging field for segmentation, Detection and localization of lesion in CT Images. SSD is used for detecting and segmenting lesions in CT images. They propose a modified SSD architecture that incorporates multi-scale feature fusion to improve detection accuracy [1]. SSD applied to detect pancreatic tumours in CT images. It highlights the use of SSD in conjunction with data augmentation techniques to improve detection accuracy for small and irregularly shaped tumours [3]. The paper presents a hybrid detection method that combines SSD with 3D Convolutional Neural Networks (CNNs) for detecting small lung nodules in CT images. The SSD framework is used for initial detection, followed by refinement using 3D CNNs [4]. Methodology for detecting liver lesions in CT images using SSD. It discusses the challenges in liver lesion detection and proposes a modified SSD architecture to improve detection performance [6]. SSDL-CT is a specialized SSD variant for liver tumour detection in CT images. The research addresses the issue of detecting small tumours and proposes techniques to improve detection accuracy [7].

3. Proposed Method

Accurate, automated lesion detection in Computed Tomography (CT) is an important yet challenging task due to the large variation of lesion types, sizes, locations and appearances. Deep learning is a fast and evolving field that has a lot of implications on medical imaging field. Currently medical images are interpreted by radiologists, physicians etc., but this interpretation gets very subjective. Radiologists often have to look through large volumes of these images that can cause fatigue and lead to mistakes. So, there is a need for automating the prediction of lesion through CT Images. One challenge in implementing the deep learning algorithms is the scarcity of labeled medical image data.

3.1 Dataset

The DeepLesion dataset from the National Institutes of Health's Clinical Center is indeed a significant resource for the medical imaging community. It is a large-scale dataset of CT images publicly available to improve detection accuracy of lesions. While most publicly available medical image datasets have less than a thousand lesions, this dataset, has over 32,000 annotated lesions (220GB) identified on CT images. DeepLesion, a dataset with 32,735 lesions in 32,120 CT images of 4,427 unique patients. There are a variety of lesion types in this dataset, such as lung nodules, liver tumours, enlarged lymph nodes, and so on. It has the potential to be used in various medical image applications.

3.2 Steps to Process and Analyze the Data

3.2.1 Load and Explore the Data

Each CT Image has dimensions of 512 x 512, has extracted the 1267 liver CT images from the Deeplesion dataset along with 18 features in the CSV file. Load the CSV file and images into analysis environment. Explore the data to understand the distribution and relationships between features. Data set Features listed below:

- File_name
- Patient_index
- Study_index
- Series_ID
- Key_slice_index
- Measurement_coordinates
- Bounding_boxes
- Lesion_diameters_Pixel_
- Normalized_lesion_location
- Coarse_lesion_type
- Possibly_noisy
- Slice_range Spacing_mm_px_
- Image_size
- DICOM_windows
- Patient_gender
- Patient_age
- Train_Val_Test

3.2.2 Augment the Data

The images are in raw files, have converted into human readable form. Extracted the entire liver images and saved it in the folder along with their bounding boxes. To enhance Dataset, applied data augmentation techniques such as Brightness, Contrast, Noise, and Translation on to Liver CT Images. In order to increase the size and variety of the dataset, the labelled images are augmented. The bounding box coordinates were adjusted according to the applied augmentations. After augmentation, the images were tested to accurately segment the lesion coordinates in the CT images. Figure1 represents the sample augmented images along with Bounding Box coordinates.

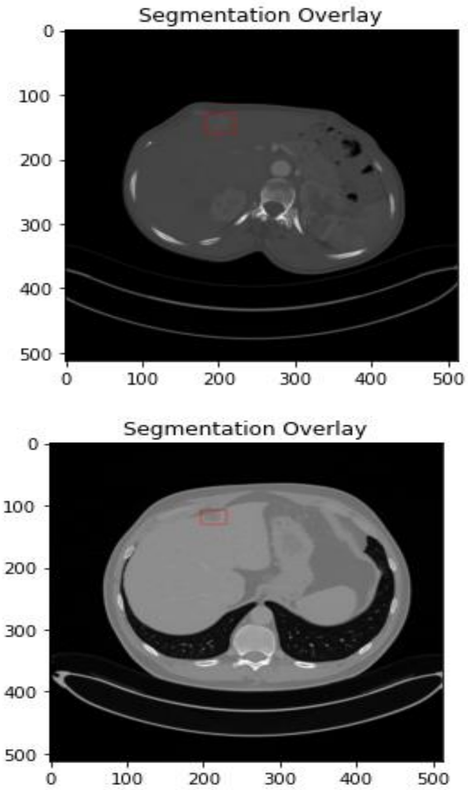


Fig 1: Sample Augmented Images of Contrast and Brightness Images

Sample snippet for contrast augmentation:

```
import imgaug.augmenters as iaa
contrast_seq = iaa.Sequential([
        iaa.GammaContrast((0.5, 2.0))
    ])
```

Similarly, the brightness, noise and translation augmentation were applied and created the dataset. After the augmentation technique the total number of images trained on the model is 6335 CT Images.

3.3. Methodology

3.3.1 Overview of SSD (Single Shot Detector)

A SSD is a type of deep learning-based object detection framework designed to detect objects in images quickly and efficiently. Unlike traditional object detection methods, which often involve a two-stage process (region proposal followed by classification), SSD performs both detection and classification in a single forward pass through the network, which is why it's called "single shot." SSD has two components: a backbone model and SSD head. Backbone model usually is a pre-trained image classification network as a feature extractor. This is typically a network like ResNet trained on ImageNet, thus left with a deep neural network that is able to extract semantic meaning from the input image while preserving the spatial structure of the image albeit at a lower resolution.

For ResNet34, the backbone results in a 256 7x7 feature maps for an input image. The SSD head are convolutional layers added to this backbone and the outputs are interpreted as the bounding boxes and classes of objects in the spatial location of the final layers activations. Figure 2 and 4 represents the SSD architecture with ResNet34 backbone model and ResNet50 Backbone Model separately and it provides a framework for liver lesion detection, combining speed and accuracy to offer a powerful tool for medical imaging. Its ability to handle multiscale detection, combined with deep learning techniques, makes it well-suited for the complex task of identifying and classifying liver lesions.

Components of SSD

1. **Feature Extraction**: SSD uses a backbone network (e.g., Resnet34, Resnet50) to extract feature maps at different layers. These feature maps capture details at multiple scales, which is crucial for detecting objects of varying sizes.
2. **Pyramid Structure**: By employing a pyramid-like structure with multiple feature maps, SSD can handle small and large objects more effectively.
3. **Direct Prediction**: SSD predicts class and bounding boxes directly from feature maps. Non-Maximum Suppression is used to filter out overlapping bounding boxes that are predicted for the same lesion. The box with highest score will be considered for detection.

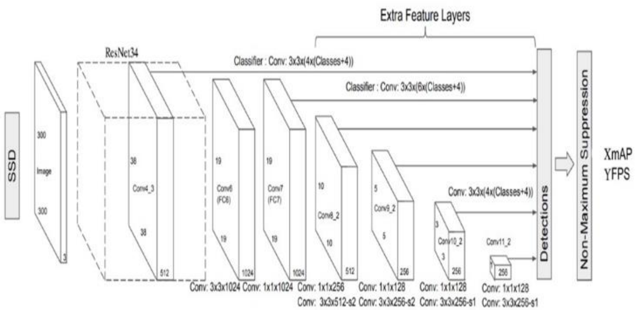


Fig 2: LiverLesionDectect SSD Model Architecture with Resnet34 as Backbone

3.3.2 ResNet Model:

ResNet models such as ResNet34 and ResNet50 are used to extract feature maps from input images. The feature maps provide rich representations that can be used to detect objects of various sizes and shapes. The initial layers of the SSD network are replaced with a pre-trained ResNet model, excluding its final fully connected layers. The remaining layers of SSD are then added to further process these feature maps for object detection. The Figure3 represents the ResNet Block with residual connections. ResNet introduces residual blocks, which include shortcut connections (skip connections) that bypass one or more layers. These connections allow the network to learn residual functions, effectively simplifying the learning process and mitigating the vanishing gradient problem. Equation for calculating the ResNet Block i.e., Y= F(X,W)+X.

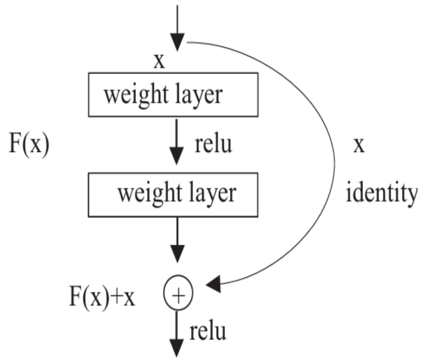


Fig 3: ResNet Block

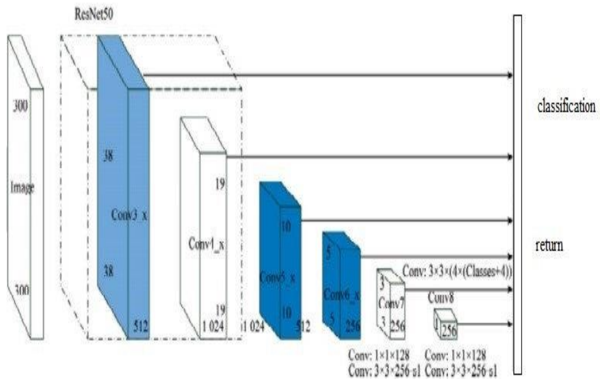


Fig 4: SSD Model Architecture with Resnet50 as Backbone

3.3.3. Algorithm for Proposed LiverLesionDetect SSD Model

1. Initialize the dataset
   1.1. Load the dataset
   1.2. Apply data augmentation techniques to increase dataset variability and robustness.
2. Model Architecture
   2.1. Load and extract the pretrained ResNet-34/50 for feature extraction as the backbone on to the SSD Model for Liver Lesion Detection.
3. Initialize the Head
   3.1. Additional four convolutional layers are added on top of the backbone to create feature maps of different scales.
4. Train & Evaluate the Model
   4.1 Executed the SSD model with batch size as 16 and utilized the SGD optimizer with Nestrov momentum.
   4.2 Evaluated the model's performance on a validation set to assess its accuracy.

The algorithm above represents the proposed LiverLesionDetect SSD model with improved model architecture with variant backbones to assess the performance of the CT Image liver lesion detection. A larger feature map has a smaller receptive field and is good for detecting smaller lesions and a smaller feature map has a larger receptive field and is good for detecting larger lesions.

4. Implementation

The pretrained ResNet34 and ResNet50 were employed as a backbone on SSD Model for training a Detection model. After the augmentation technique the total number of images trained on the model is 6335 CT Images. These 6335 Images were split for training and validation with a split ratio of 80% and 20%. These CT Images were applied on SSD Model. The annotated data, the bounding boxes containing information about the location of lesion in the images, was stored in XML format. XML annotations for images based on the coordinates specified in a CSV file. These annotations are typically used for lesion detection tasks where each image is associated with bounding box coordinates around the objects of interest. The proposed solution for Liver Lesion detection relies on the Single Shot Detector (SSD) model within the PyTorch framework and ResNet as the Backbone architecture. PyTorch is chosen for its flexibility, ease of use, and strong community support, facilitating seamless integration with SSD for both training and deployment. As for the architecture, ResNet34 variants are employed as the backbone due to their effectiveness in feature extraction, essential for accurately detecting Lesions. SSD is selected as the primary model because of its capability to predict bounding boxes at different scales in a single pass through the network, making it suitable for early detection. Additionally, experimenting with variant SSD Backbone models on SSD architecture ensuring the performance. In summary, the solution compares the ResNet34 architecture, and ResNet50 architecture as a backbone on SSD model to propose a robust system for Liver Lesion detection. Initially, it involves collecting and annotating a large dataset of liver images with labeled lesions. Trained separately on the SSD model with ResNet34 architecture and ResNet50 architecture as a backbone learning to recognize the features of liver lesions. Then the model evaluated on a separate test set to ensure it performs well in detecting and localizing lesions accurately. To locate the lesion uses anchor boxes in SSD, which are pre-defined bounding boxes that the model uses to detect lesions in the CT images. These anchor boxes are responsible for a size and shape within a grid cell. Liver Lesion detection with SSD model should include:

- Probability that there is an lesion in the CT Image,
- Height and Width of the bounding box,
- Horizontal coordinate and Vertical Coordinates of the center point of the bounding box

Hyper parameters

The Hyperparameters are a key component in the training and tuning of model, and they control the behavior of the training algorithm and as well as the model architecture. The hyperparameters which set in SSD Model approach are such as a learning rate of 0.0005 was determined to be optimal for efficient model convergence and performance, Batch size of 16 was selected based on available memory and computational resources, balancing training speed and memory usage, Number of Epochs for the model to train was set 50, 75 and 100 epochs to ensure adequate learning, Optimizer used is SGD with learning rate 0.0005, momentum 0.9, and Nesterov momentum. SSD internally uses loss function addressing localization tasks. Localization loss measures the difference between predicted and ground truth bounding boxes. Relu activation function used in the model.

4.1 Evaluation Metrics

To evaluate the performance of proposed model, the following metrics were computed, like precision, recall, IoU, mAP50 and mAP50-95:

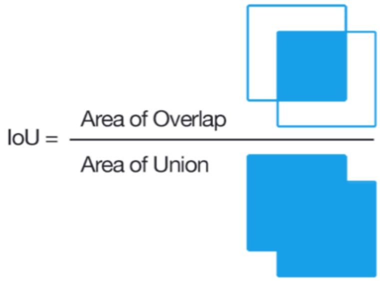
Precision: The proportion of true positive detections (correctly predicted bounding boxes) out of all positive detections (all bounding boxes predicted by the model).

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

Recall: The proportion of true positive detections out of all actual positives (all actual bounding boxes in the ground truth).

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

IoU: The intersection over the union (IoU) is the ratio between the intersected area over the joined area for two regions.



Mean Average Precision (mAP): mAP is the mean of the average precision values across all classes and IoU thresholds. The mAP@50-95 metric evaluates AP at multiple IoU thresholds ranging from 0.50 to 0.95 in increments of 0.05.

4.2 Results

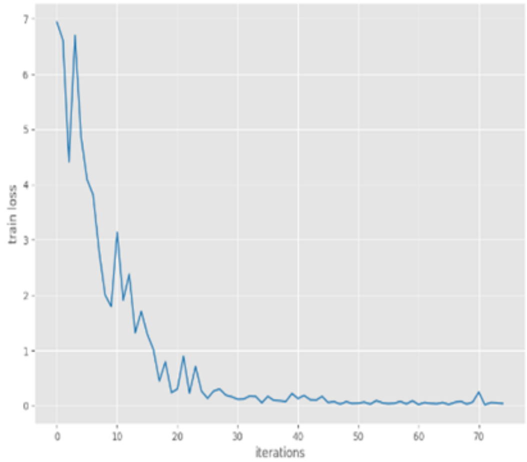
The prediction of liver lesions was carried out using Single Shot Detector (SSD) models, incorporating variants of the ResNet architecture as the backbone for feature extraction. Specifically, ResNet34 and ResNet50 were applied to the SSD framework, utilizing a batch size of 16 and training over 50, 75, and 100 epochs. Stochastic Gradient Descent (SGD) with a momentum of 0.9 was employed as the optimizer. In Table 1-It was observed that the ResNet34-based SSD model outperformed the ResNet50 variant in terms of prediction accuracy. However, training the models for 100 epochs led to over fitting, resulting in a decrease in precision. Each grid cell in SSD can be assigned with multiple anchor/prior boxes. Figure5 - Represents the mean Average Precision (mAP) and train loss for ResNet34 Backbone model on

SSD. Figure 6 – Represents the sample predictions of Liver Lesions. Figure 7 Comparison of Liver Lesion Detection for Variant SSD Models. The Metrics to evaluate the Model are Precision, Recall, mAP50 and mAP50:95 and train loss on variant SSD Models for different epoch rates. SSD predictions are classified as positive matches or negative matches. SSD only uses positive matches in calculating the localization cost. If the corresponding default boundary box has an IoU greater than 0.5 with the ground truth, the match is positive. Otherwise, it is negative.

Table 1: Represents the Metrics for different SSD Models with Resnet as the backbone Model

Metrics / Models	**SSD with Resnet34 Backbone (75 Epochs)**	**SSD with Resnet34 Backbone (100 Epochs)**	**SSD with Resnet50 Backbone (50 Epochs)**	**SSD with Resnet50 Backbone (75 Epochs)**
Precision	**0.915**	0.888	0.76	0.791
Recall	**0.924**	0.881	0.872	0.857
mAP50	**0.917**	0.875	0.854	0.838
mAP:50-95	**0.744**	0.673	0.588	0.611
Train Loss	**0.02**	0.029	0.082	0.056

The below figure 5 and 6 represents the graph. The first graph shows recording of train loss for every iteration of the proposed model i.e., SSD with resNet34 as backbone model. The graph generated for 75 epochs. The model compared with complex structure with ResNet50, but it could not give the impressive results. Though we have increased the epochs the accuracy is decreased.



And the second graph represents the mean average precision of the proposed model. The orange line indicates the mAP 50 and red line indicates the mAP50-95

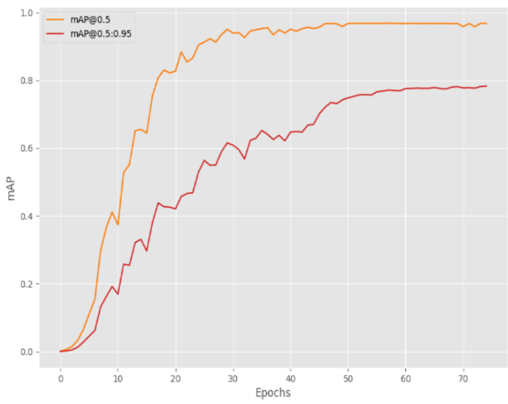
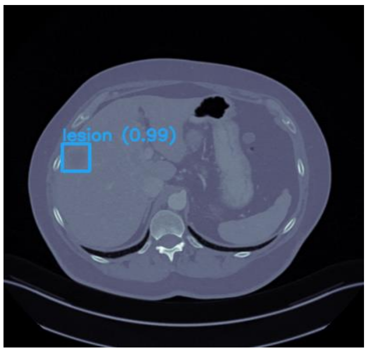


Fig 5: Graphs shows the train loss and mean average precision for SSD

Model with ResNet34 as backbone



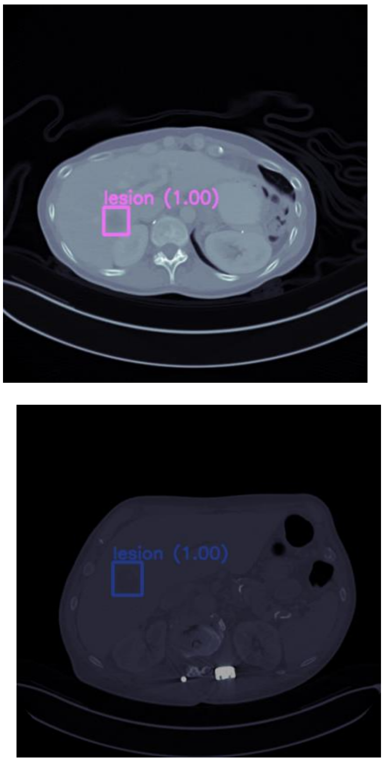


Fig 6: Sample CT Images representing Liver Lesion Detection using SSD with Resnet34 as backbone Model

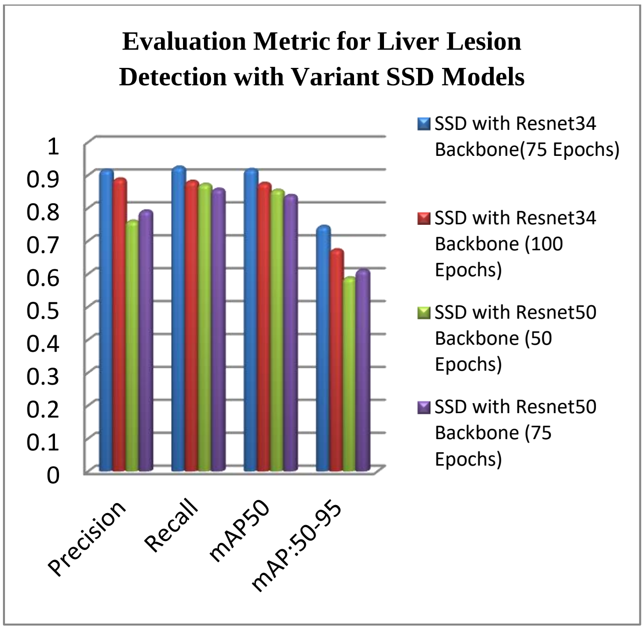


Fig 7: Comparison of Liver Lesion Detection for Variant SSD Models

5. Conclusion and Future Scope

SSD provides a robust framework for liver lesion detection, combining speed and accuracy to offer a powerful tool for medical imaging. Its ability to handle multiscale detection, combined with deep learning techniques, makes it well-suited for the complex task of identifying and classifying liver lesions. Accurate and fast detection of liver lesions can lead to early diagnosis of conditions in liver cancer. Identifying the exact location and size of lesions is crucial for planning treatments. Hence, SSD can be used to monitor the progression or regression of lesions over time, helping in evaluating the effectiveness of treatment. The Future work for the Liver lesion prediction is to predict the multiple lesions on a CT images. And focus on improved accuracy for better detection.

References

[1] Zhang, Z., Chen, H., and Zheng, Y., Deep Learning-based Detection and Segmentation of Lesions in CT Images Using Single Shot Multibox Detector, Medical Image Analysis, 2019

[2] Wang, L., Zhang, Y., Single Shot Multibox Detector for Automated Detection of Colorectal Polyps in CT Colonography, IEEE Transactions on Medical Imaging, 2019.

[3] Chen, J., Wu, Y., Automated Detection of Pancreatic Tumors in CT Images Using SSD, Computers in Biology and Medicine, 2020.

[4] Dou, Q., Chen, H., Yu, L., Detection of Small Lung Nodules Using a Hybrid Approach Based on SSD and 3D CNNs, International Journal of Computer Assisted Radiology and Surgery, 2020.

[5] Li, J., Wang, X., Zhang, S., A Transfer Learning Approach for Lesion Detection in CT Images Using Single Shot Multibox Detector, IEEE Access, 2020.

[6] Zhao, X., Zheng, H., Detection of Liver Lesions in CT Images Using Single Shot Multibox Detector, Journal of Digital Imaging, 2021.

[7] Song, Y., Wei, C., SSDL-CT: Single Shot MultiBox Detector for Liver Tumor Detection in CT Images, Pattern Recognition Letters, 2021.

[8] Wu, X., Li, L., Wang, J., Comparative Study of YOLO and SSD for Detecting Lung Nodules in CT Images, IEEE Access, 2021.

[9] Rushikesh Chopade, Aditya Stanam, Shrikant Pawar "Single shot detector application for image disease localization", biorxiv 2021.

[10] Ranjbarzadeh, R.; Bagherian Kasgari, A.; Jafarzadeh Ghoushchi, S.; Anari, S.; Naseri, M.; Bendechache, M. Brain Tumor Segmentation Based on Deep Learning and an Attention Mechanism Using MRI Multi-Modalities Brain Images. Sci. Rep. 2021.

[11] Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-Captured Scenarios. arXiv 2021.

[12] Zhao, X., Evaluation of YOLO and SSD for Detection of Pancreatic Tumors in CT Images, Computers in Biology and Medicine, 2022.

[13] Shokofeh, A.; Nazanin, T.S.; Negin, M.; Shadi, D.; Amirali, R. Review of Deep Learning Approaches for Thyroid Cancer Diagnosis. Math. Probl. Eng. 2022.

[14] Mansour Alhlalat, Ahmad Sharieh, Mohammed Alzoubi "Lung Disease Detection and Classification Using Single Shot Multi-Box Detector Network: A Comprehensive Study", JMCS Journal,2023.

[15] Jian Ni,Rui Wang and Jing Tang," ADSSD: Improved Single-Shot Detector with Attention Mechanism and Dilated Convolution", March 2023.