20029527

Supervisor: Xin Chen

Module Code: COMP3003

2021/05

Automatic Organs Localisation in 3D Medical Images

Submitted May 2021, in partial fulfilment of

the conditions for the award of the degree **BSc Hons Computer Science with Artificial Intelligence**

20029527

School of Computer Science

University of Nottingham

I hereby declare that this dissertation is all my own work, except as indicated in the text:

Signature _____YJ X_____

Date 06/05/2021

I hereby declare that I have all necessary rights and consents to publicly distribute this dissertation via the University of Nottingham's e-dissertation archive.*

Abstract

Computed Tomography (CT) and Magnetic Resonance (MR) images are medical imaging techniques used in many clinical practices. Organ localization is an essential preprocessing step for many medical image processing tasks such as organ classification and organ segmentation. Currently, most organ localization methods can only locate organs in a single image modality, for example, either in CT or in MR images. Users may need to retrain the model and re-tune its parameters if they want to apply it to other imaging modalities, which is time-consuming. In this project, a 3D region proposal network is used as the fundamental method for building a multi-modality model, which can detect organs in both CT and MR images without retraining the model. To achieve higher localization accuracy of the model, the Sobel operator and a novel preprocessing technique, called multi-resolution local image normalization, are implemented to preprocess the raw medical images. All models are evaluated on two clinical abdominal CT datasets and one abdominal MR dataset, where at most 11 body organs are included. As the results have shown, the multi-modality model, trained on the abdominal CT dataset, can detect organs in both MR and CT images with high detection precision and localization accuracy. Compared with the original model, one updated model, using Sobel edge enhanced image as input, achieves higher localization accuracy in CT images but lower in MR images. Another updated model, using multi-resolution edge enhanced image as input, increases the detected accuracy in CT images and some organs accuracy in MR images.

Acknowledgements

I want to express my gratitude to my supervisor, Dr. Xin Chen, for his consistent guidance and encouragement throughout the project and my friend for having the patience to listen to me talk about my difficulty countering in this project. I would also like to thank one of the authors of the 3D region proposal network model, Xuanang Xu, for his instant and kind help with any questions regarding their model input and output and framework setup. Finally, I would like to thank my family for their continued support. Without these people, I would not complete the project as high a standard.

Contents

1 Introduction 1
1.1 Backgrounds and Motivations 1
1.2 Aims and Objectives 1
1.3 Descriptions of the work 2
1.4 Main Contributions 2
2 Related Work 3
2.1 Classical Machine-Learning Methods 3
2.2 Deep Learning-based methods 4
2.2.1 2D ConvNet-based models 4
2.2.2 3D ConvNet-based models 4
2.3 Literature Review Summary 5
3 Methodology 6
3.1 3D Region Proposal Network (3D RPN) 6
3.1.1 Model Overview 6
3.1.2 Backbone Network for Feature Map Extraction 7
3.1.3 RPN for Bounding Box Prediction 7
3.1.4 Multiple Bounding Box Fusion Strategy 8
3.1.5 Training Label Assignment 8
3.1.6 Loss function 8
3.2 Evaluation criteria 9
3.3 Data Preprocessing and Augmentation 10
3.3.1 Sobel Edge Detection (Sobel Operator) 10
3.3.2 Multi-Resolution Local Image Normalization 11
4 Design 12
4.1 Datasets 13
4.1.1 Abdominal CT Dataset 13
4.1.2 Independent Abdominal CT Dataset 13
4.1.3 Abdominal MR Dataset 14
4.1.4 Bounding Box Annotation 15
4.2 Experiments Design 16
5 Implementation 16
5.1 Datasets Setup 16
5.1.1 Image File Conversion 16
5.1.2 Image Orientation Setup 17

5.2 Annotations Extraction 18
5.3 Data Preprocessing 19
5.3.1 Data Preprocessing in Model 19
5.3.2 Image Preprocessing with Sobel Operator 20
5.3.3 Image Preprocessing with Multi-Resolution Local Image Normalization 20
5.4 Model Training 21
5.5 Model Output 21
5.6 Model Evaluation 21
6 Results & Discussions 22
6.1 Qualitative Results 22
6.2 Quantitative Results 24
6.2.1 Results for Abdominal CT Dataset 25
6.2.2 Results for Independent Abdominal CT Dataset 26
6.2.3 Results for Abdominal MR Dataset 27
6.3 Analysis & Discussion 28
6.4 Future work 29
7 Summary and Reflections 29
7.1 Project Management 29
7.2 Contributions 31
7.3 Reflections 32
8 Conclusion 32
9 Bibliography 33
10 Appendix 34
10.1 Abbreviations 34

1 Introduction

1.1 Backgrounds and Motivations

Computed Tomography (CT) and Magnetic Resonance (MR) images are medical imaging techniques used in many clinical practices to assist medicine surgery, illness diagnosis, and patient's therapy. They help produce detailed pictures of organs, bones, and other tissues and detect abnormalities such as organ failure and cancer. Although both techniques depict similar anatomical structures, their output images vary in many aspects, such as pixel intensity range, slice thickness, in-plane spacing, and noise, which is because the CT scan uses X-rays to generate the picture, but MRI uses radio waves and a powerful magnet.

Efficient and accurate automated organ localization in both image protocols is an essential prerequisite for organ classification, organ segmentation, and lesion detection. Accurately estimating the target organ's position can help users quickly focus on Regions-of-Interests (RoI), contributing to subsequent medical image processing tasks. In organ segmentation, for instance, RoIs is commonly extracted from the targeting images by organ localization as the first step to discard most of the non-relevant information. Effective organ localization can not only save time and memory but also improve the accuracy of the subsequent segmentation. It can also help develop intelligent software tools in medical images to boost diagnosis efficiency and accuracy, and the tedium and oversights of some medical images can be reduced.

However, due to various causes, precise organ localization in medical images is hard to achieve. Low-contrast soft tissues and noise surrounding the target organs can sometimes blur the organ boundaries. Each patient usually has a unique appearance of the target organs, which requires the algorithm to detect correctly even in some variations. Considerable time and memory are consumed due to the high dimensionality of CT scans and MRI, which is a troublesome issue. The classifier's vagueness could increase if the target organs were truncated too much through data processing.

In addition, the imaging modalities used in biology and medicine are based on a variety of energy sources, such as ultrasound, lasers, and x-ray. Different modalities could depict similar anatomical structure but output very different images. However, most of the state-of-the-art organ localization methods can only locate organs in a single image modality, for example, either in CT or in MR images. This may require users to retrain the model using different datasets and optimize each model's parameters if they want to apply it to other imaging modalities, which is time-consuming.

In this project, the research focuses on building a multi-modality model to automatically locate multiple organs in both MR and CT images without retraining the model. This model can save the time used in training different models and collecting training data. Also it is more possible to be applied in the real-life application, as one application can handle various imaging techniques. As the experiments are carried on 3D medical images, the 3D bounding box is used to locate different organs in a medical image.

1.2 Aims and Objectives

The project's main aim is to borrow the structure of a state-of-the-art model to

automatically locate different organs in 3D CT images and extend it to detect organs in MR images to achieve multi-modality organ detection with one single model. The model is evaluated on two separate abdominal CT datasets and one abdominal MR dataset. The final model should perform multi-organs localization accurately and efficiently in both 3D CT and MR images.

The key objectives of this project are:

1. Identify one state-of-the-art method that can efficiently locate multiple organs in CT images.
2. Retrain the selected model using CT images and test its localization accuracy on CT images from same dataset.
3. Evaluate the trained model using another independent abdominal CT datasets to check the robustness of the selected model.
4. Extend the model to be able to locate organs in MR images.
5. Evaluate the multi-modality model using abdominal MR datasets and tune the model parameter to optimize the results.
6. Preprocess CT and MR images with Sobel Operator, retrain and evaluate the multi-modality model using processed images.
7. Preprocess CT and MR images with a novel multi-resolution local image normalization method, retrain and evaluate the multi-modality model using processed images.
8. Compare the performance of three different models and summarize the results.

1.3 Descriptions of the work

The project is designed to give results of the localization accuracy of the multi-modality model in CT and MR images to show the potential of the success of this model. To further explore the methods to improve the organ localization accuracy, two techniques are implemented to preprocess the input images before feeding them into the network. The first method uses the Sobel operator to enhance the edge information in the medical images. Another method is multi-resolution local image normalization, which can emphasize some weak edges that could be ignored by standard gradient filters and mitigate the intensity discrepancy between MR and CT images to make them more similar. In total, three multi-modality models are developed and compared in this project.

1.4 Main Contributions

The major contributions of this work are in three folds:

1) A multi-modality model based on 3D Region Proposal Network is built, with the ability to detect organs in both MR and CT images with high detection precision and localization accuracy. It shows the potential of the multi-modality model and brightens the future that using a single model to detect organs in many imaging protocols.
2) A novel image normalization method that could remove some noise in the original images and minimize the intensity discrepancy between CT and MR images is proposed. Benefiting from this method, the updated model remains high detection accuracy in CT images and hugely increases some organs' accuracy in MR images.
3) The Sobel operator is found to improve the localization accuracy of the model in CT images, which gives other researchers a possible option to preprocess the raw image to increase the organ detection and localization accuracy in CT images.

2 Related Work

In this section, a literature review related to the proposed approaches to locate the organs in 3D medical images is presented. ***Table 1*** summarizes the methods that will be discussed, including classical machine learning methods and deep learning-based methods.

Table 1 - Literature Reviews Summary

Literature Review				
Author	Image Type	Achievements	Method	Evaluation method
Zhou et al	3D CT image	Detect **heart, liver, spleen, two kidneys** locations in 3329 3D scans with 2770 used for training	Collaborative majority voting decision based on emsemble learning	Comparing the detected 3D rectangle with the ground-truth in 3D MBR (Use Jaccard similarity and the Euclidean distance
Zheng et al	2D MRI Images 3D CT image	**Left ventricle** Detection using total 795 MRI images with 400 for training. Liver detection in 226 3D CT images	Marginal Space Learning(MSL) Constrained MSL Constrained+Nonrigid MSL	Center-center distance and vertex-vertex distance (the mean Euclidean distance between vertices of the box)
Criminisi et al	3D CT image	Detect 26 **organs** based on 400 CT volumns, with 318 CT volumes used for training	Multivariate regression forest	Whether the centroid of the predicted organ bounding box is contained by the ground truth bounding box
Gauriau et al	3D CT image	Detect **live, two kidneys, spleen, gallbladder and stomach** based on 130 CT volumes with 50 for training	global-to-local cascade of regression rabdin forest	The distribution of mean distances to ground truth for each organs
Humpire-Mamani et al	3D CT image	Detect **8 organs** based on 1884 thorax-abdomen CT scans with 1130 used for taining and 377 for validation	Multi-label Convolutional Neural Network	Compared to the ground truth using Dice score, Jaccard Coefficient and the wall distance to the reference bounding box
de Vos et al	3D CT image	Localising **heart, aortas, aortic arch** in total 400 3D CT scan with 200 used for training	ConvNet using 2D image slices as input	Use distances between automatic bounding box and the reference bounding box and compute their centroids.
Xu et al	3D CT image	Detect **11 organs** in 201 abdominal CT dataset (131 training) and **12 head organs** in 119 head CT dataset (80 training)	3D Region Proposal Network	Average Presion: for detection precision IoU: for localisation accuracy Average Processing Time: for processing efficiency
Criminisi et al	3D MRI image/3D CT image	Detect: **26 organs** in 400 3D CT with 318 for training / **5 organs** in 33 2-channel MR Dixon Sequences with 20 for training	Multivariate regression forest	1. Centroid of bounding box 2. Compare with atalas-based registration approach using mean error
Keraudren et al	3D MRI image	Detect **fetal brain** in Fetal MRI based on 59 healthy fetuses with 39 of them used for training	SIFT features aggregated with Maximally Stable Rxtremal Regions to detect brain location and use RANSAC to find the cube	The distance of the ground truth bounding box and the detected bounding box
Xu et al	3D CT image	Detect **11 organs** in 201 abdominal CT dataset (131 training)	3D Fully Convolution Network (triple-branch FCN)	Intersection-over-union between the predicted bounding box and the ground-truth bounding box

2.1 Classical Machine-Learning Methods

Many classical machine learning methods were proposed for multiple organ localization in different medical images in the early days. Zhou *et al*. [1] proposed to use a majority voting algorithm based on ensemble learning with Adaboost to automatically locate organs in 3D CT images. They trained some Adaboost classifiers using 2D Haar-like features and implemented majority voting to detect a target organ in 3D volumes based on the previous classifier's training result. To handle high dimensionality, they considered 3D organ localization as detecting several independent 2D objects, which significantly reduced the feature dimension and expanded the training samples. Keraudren *et al*. [2] located the fetus brain in fetal MRI with Bundled SIFT features. They first extracted the brain area from 2D slices using maximally stable extremal regions (MSER) combined with SIFT features (called Bundled SIFT features). Then they used a random sample consensus procedure (RANSAC) to find a best-fitting 3D cube whose dimensions are inferred from prior knowledge. Marginal Space Learning (MSL), a popular approach for localization, was proposed by Zheng *et al*. [3]. They used MSL to detect the left ventricle in 2D MR images and introduced a novel constrained MSL combined with non-rigid MSL to perform liver detection in 3D CT volumes. By considering 3D objects to be nine dimensions (three for the position, three for orientation, and three for anisotropic scaling) and searching low-dimensional marginal spaces only, they discounted the number of testing hypotheses and successfully exploited MSL to perform the localization problem efficiently.

The regression tree-based method was popular as well. Criminisi *et al*. [4], [5] offered a multi-class random regression forest to detect 26 organs, predicting organ bounding boxes based

on a set of the ground-truth organ bounding box and gradient features. However, the accuracy was better for larger organs. Those with smaller size or greater positional variability were difficult to detect. They also detected five organs in 33 2-channel MR Dixon Sequences and proved that their methods could significantly minimize the MR bounding box localization errors compared with an atlas-based registration approach [5]. Gauriau *et al*. [6] further improved the multi-class random regression forest by generating a global-to-local cascade of the regression random forest tree, which cascaded the initial regression forest with dedicated per-target forests. It reduced localization errors but increased localization time.

2.2 Deep Learning-based methods

In recent years, deep learning-based methods have gained more and more attention in medical image analysis. Some methods processed 3D medical images slice by slice to build the model, while others were fully implemented in a 3D manner.

2.2.1 2D ConvNet-based models

Humpire-Mamani *et al*. [7] proposed a multi-label convolutional neural network to locate eight organs and three bony structures in 3D thorax-abdomen CT scans. They represented the organs in continuous 2D slices. Three ConvNet were used—one for each orthogonal view—to predict the location. By combining the predicted results from three ConvNets, the 3D bounding boxes around each organ were obtained. De Vos *et al*. [8] introduced a single ConvNet to detect one or more anatomical structures in 3D CT images using 2D image slices. By combining the result calculated through the ConvNet in all slices, the 3D bounding box was then created. Their network was trained to detect the presence of the anatomical structure of interest in axial, coronal, and sagittal slices extracted from 3D CT volumes. Spatial pyramid pooling was applied to allow the network to analyze slices of various sizes.

However, methods [7], [8] predicted the organ's location using 2D image slices in CT images and combined the predicted results after separate calculations. It could be time-consuming and redundant, as the calculation of slices of CT images cannot run simultaneously, and the adjacent slices may present similar contents. Also, implementing 3D spatial contextual information in a 2D way could not fully utilize the benefit brought by the 3D context, which could result in poor accurate organ localization.

2.2.2 3D ConvNet-based models

A deep learning network that fully exploited the spatial context information in medical images using a 3D neural network is more promising. Xu *et al*. [9] presented a 3D Region Proposal Network (RPN) model for organ localization in CT images. As each organ in CT images contains only one instance, they directly exploited their model to generate several bounding boxes and fused them to make only one prediction. 3D RPN located organs more accurately than those previously mentioned networks. It, however, had a low Intersection over Union overlap ratio in the bladder and the pancreas detection, which might be because of the low-contrast boundaries and varied appearance of these two organs.

They compared their results with some previous methods [7], [8] using the wall distance and average processing time. To make the comparison more convincing, they re-implemented the methods [7], [8] on the same dataset and fine-tuned the model's hyper-parameters and training configurations. The result in ***Table 2*** proved that their 3D RPN model had higher organ localization accuracy with less processing time, which indicated the strength of the model implementing in a 3D manner.

Table 2 – The Wall Distance (The Mean and The Standard Deviation of The Absolute Wall Distance Between The Predicted Bounding Box and The Ground-Truth Bounding Box) and Average Processing Time of Different Body Organ Localization Methods Evaluated on Same Dataset. Source: Adapted from [9].

Organs	Wall dist. [mm]: Mean (Standard Deviation)		
	Methods		
	De Vos et al. [8]	Humpire et al.[7]	Xu et al. [9]
L-lung	6.8 (13.2)	5.5 (5.4)	**5.1 (3.8)**
R-lung	6.3 (14.2)	**4.6 (5.1)**	4.9 (4.9)
Heart	9.2 (14.1)	6.1 (6.3)	**4.1 (4.6)**
Liver	11.7 (15.8)	10.7 (13.6)	**8.5 (9.4)**
Spleen	11.8 (13.5)	10.1 (12.3)	**6.3 (6.7)**
Pancreas	13.1 (12.0)	11.9 (10.4)	**9.2 (7.9)**
L-kidney	10.1 (12.4)	8.5 (9.9)	**4.3 (4.2)**
R-kidney	8.7 (10.5)	8.2 (9.2)	**3.9 (3.5)**
Bladder	11.1 (9.7)	10.4 (8.7)	**7.3 (6.5)**
L-femur	4.9 (4.8)	5.8 (4.2)	**2.1 (1.9)**
R-femur	4.9 (4.8)	6.0 (4.6)	**1.9 (1.6)**
Times [s]	1.2	5.4	**0.3**

To simplify the network structure and make the network easy to deploy in other applications, Xu *et al*. [10] proposed a triple-branch fully convolutional network (triple-branch FCN) to locate 11 organs in CT scans. The implementation was more straightforward than the 3D RPN, as only fundamental components in ConvNet were included, like the convolution layer, pooling layer, and SoftMax function. A density enhance filter was used to enhance structures in the input CT images. Before feeding the images into the network, these enhanced images needed to integrate with the raw images to form a three-channel image. This process, however, consumed more time and memory. Although the predicted result of all organs' accuracy increased, the trade-off between the time and memory consumption and accuracy must be considered.

2.3 Literature Review Summary

Table 3 - The Wall Distance (The Mean and The Standard Deviation of The Absolute Wall Distance Between The Predicted Bounding Box and The Ground-Truth Bounding Box) and Average Processing Time of Different Body Organ Localization Methods Based on Their Paper Results. Source: Adapted from [9].

Methods	Wall dist.[mm]: Mean (Standard Deviation)											
	L-lung	R-lung	Heart	Liver	Spleen	Pancreas	L-kidney	R-kidney	Bladder	L-femur	R-femur	Time
Criminisi et al. [4], [5]	12.9(12.0)	10.1(10.0)	13.4(10.5)	15.7(14.5)	15.5(14.7)	-	13.6(12.5)	16.1(15.5)	-	10.6(14.4)	11.0(15.7)	4.0
Gauriau et al. [6]	-	-	-	10.7(4.0)	7.9(4.0)	-	5.5(4.0)	5.6(3.0)	-	-	-	3.2
De Vos et al. [8]	-	-	**3.2(4.0)**	8.9(15.0)	-	-	-	-	-	-	-	6.4
Humpire et al. [7]	**2.3(3.1)**	**2.0(2.6)**	-	**5.8(12.7)**	**3.4(8.4)**	-	**2.7(7.2)**	**3.0(9.3)**	**4.7(7.9)**	**1.0(2.3)**	**1.0(2.5)**	4.0

As can be seen from the results summarized by Xu *et al*. [9] in ***Table 3***, deep learning-based models [7], [8] detected more accurately than the classical machine learning methods [4], [5], [6]. Also, ***Table 2*** demonstrated that the methods that were fully implemented in the 3D manner [9] were proved to have better performance than the one generating the final 3D bounding box by assembling the detected target organs in all slices.

It can be seen from the data in ***Table 1*** that most models only focused on organ localization in either CT or MR images. Although the method [5] proposed by Criminisi *et al*. applied their multivariate regression forest to both image modalities, the model retrain process was

required to detect organs in different image modalities. Therefore, in this project, a 3D deep learning network is borrowed to achieve accurate organ localization in CT images and extended to be a multi-modality model to simplify the organ localization process across different image modalities.

3 Methodology

This section describes data preprocessing and augmentation strategy, the selected model, and the evaluation criteria used to analyze the model. Also, two image preprocessing techniques preprocessing the input images and improving the model localization accuracy are introduced.

3.1 3D Region Proposal Network (3D RPN)

Although countless algorithms were proposed to tackle this challenging localization task, after comparing with various machine learning models, the 3D RPN model proposed by Xu *et al*. [9] is selected as the base model in this project, as it outperforms most of the state-of-art models with less time and memory consumption and higher organ localization accuracy. In addition, the input size of the images is not constrained in this network, as there is no fully connected layer. Benefiting from this design, the model is unlimited to the input images size, which is helpful, as medical images produced by different machines and techniques have various slice sizes and thicknesses.

Originally, this proposed model only detects organs in CT images. In this project, the original model is extended to a multi-modality model, which can detect organs in both MR and CT images. The fundamental structures of the 3D RPN model are remained, but the model input is extended to ensure it can receive MR images as well. Input images are preprocessed depends on its image modality. The model's data preprocessing parameters for each image modality are optimized individually.

3.1.1 Model Overview

As shown in ***Figure 1***, the feature maps of an input CT image are first extracted using ConvNet. Reference bounding boxes of various sizes are then allocated to each cell in the feature map. For each reference bounding box, the region proposal network predicts multiple classes' scores and bounding box adjustment parameters of the feature map. Those adjustment parameters are then applied to the reference bounding box, which creates multiple class-specific candidate bounding boxes. To eliminate the redundancy of these boxes, the model employs a multiple bounding box fusion strategy on those boxes and finally produces the target organ's final bounding box.

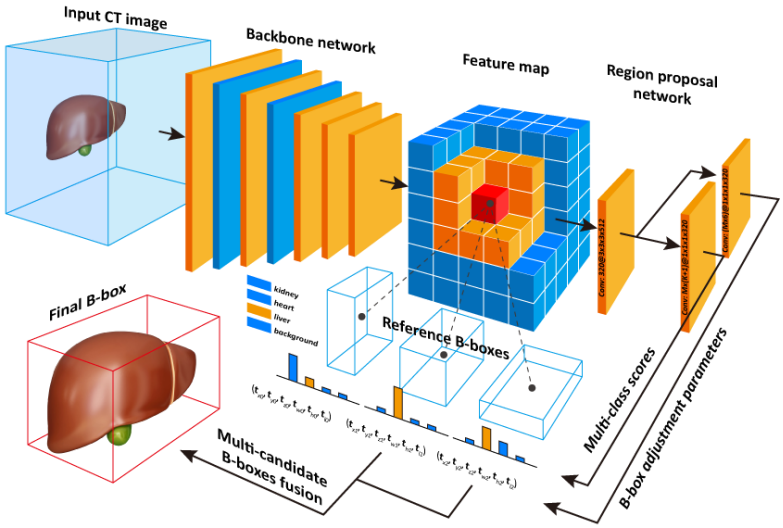

Figure 1 - Schematic Representation of 3D Region Proposal Network. Source: Adapted from [9].

3.1.2 Backbone Network for Feature Map Extraction

A novel backbone network, extended from the AlexNet structure, is built to generate a high-resolution feature map for feature extraction. As shown by the architecture in ***Figure 2***, hierarchical feature maps of the input CT images with various resolutions are produced through the convolutional layers and pooling layers. Typically, feature maps in low-level have rich spatial information, but the high-level ones contain strong semantic knowledge. Therefore, the feature maps in various layers are aggregated to hallucinate high-resolution feature maps. A deconvolutional layer is used for upsampling the high-level feature map in this process. Only the top feature map is merged with the second-top one because of the restricted memory size. The model applies batch normalization [11] after each convolutional layer to accelerate the network's learning convergence.

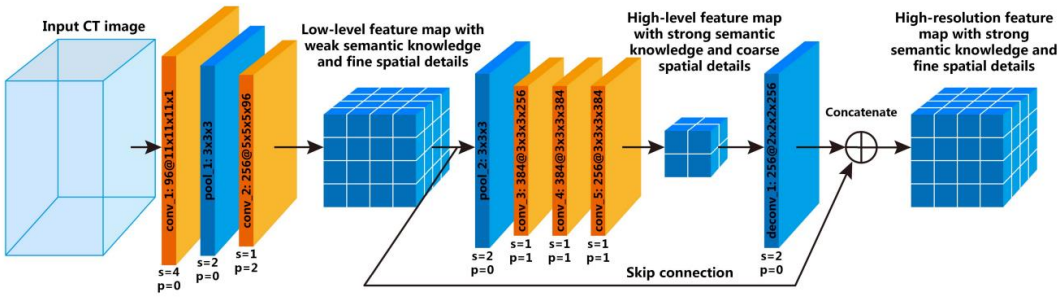

Figure 2 - Structure of Novel Backbone Network. Source: Adapted from [9].

3.1.3 RPN for Bounding Box Prediction

As can be seen from ***Figure 1***, a $3 * 3 * 3$ convolutional layer with ReLU activation function, together with two $1 * 1 * 1$ convolution layers, are used in this 3D RPN. In each feature map cell, a sliding window in size of $3 * 3 * 3$ perform the searching, and each window is mapped to a 320-d feature vector. For two $1 * 1 * 1$ convolutional filters, one layer calculates the scores of multiple classes, while the other predicts the bounding box adjustment parameters of the feature map. In each spatial dimension, four base sizes, $30mm, 60mm, 120mm, 240mm$, generate 64 reference bounding boxes at each cell in the feature map.

3.1.4 Multiple Bounding Box Fusion Strategy

Through the RPN, for each organ, several candidate bounding boxes are generated. Therefore, to remove the redundancy of the candidate bounding boxes and further increase the localization accuracy, a multiple-candidate fusion strategy is applied. Two conditions are checked to see whether the associated candidate bounding boxes B_i should be kept or not.

- The class score p_i > an absolute threshold T_1 (0.9)
- The class score p_i included in top T_2 (10%) percent of the total number of related candidate bounding boxes

In each CT image, there is only one instance of each organ. Therefore, as shown in the ***Equation 1***, the final bounding box B_{final} are produced using the weighted mean of these candidate bounding boxes.

$$B_{final} = \frac{\sum_i p_i B_i}{\sum_i p_i} \quad (1)$$

3.1.5 Training Label Assignment

In this model, the reference bounding boxes are assigned with class-specific labels. Intersection-over-Union (IoU) overlap ratio between the reference bounding box and every ground-truth bounding box in the input CT image are computed separately. The highest IoU value then compares with different thresholds to assign the label. If it is higher than $T_f(0.35)$, a foreground threshold, the corresponding ground-truth label (μ > 0) is assigned to the reference bounding box. If it is lower than $T_b(0.25)$, a background threshold, a background label (μ = 0) is assigned to the reference bounding box. If neither cases are satisfied, an ignored label (μ = -1) is assigned, and the model skips those boxes during the training phase.

For each ground-truth bounding box, the model correlates the reference bounding box whose IoU overlapping with it is the highest and the target adjustment parameters t^* relative to it are computed.

3.1.6 Loss function

Together with the label assignment, the model optimized itself by minimizing the loss function shown in ***Equation 2*** below.

$$Loss = L_{cls} + L_{reg} \quad (2)$$

$where$

$$L_{cls} = classification\ loss\ function$$
$$L_{reg} = bounding\ regression\ loss\ function$$

Considering the inter-class imbalance in the training set and the fact that many small organs are easily treated as background or ignored labels, the classification loss function (***Equation 3***) is used, which can rebalance the different classes loss and focus training on hard-classified reference bounding boxes.

$$L_{cls} = -\sum_{i=0}^{n} [\mu_i \geq 0] \frac{1}{N_{\mu_i}} (1 - p_i)^{\gamma} log p_i \quad (3)$$

$where$

$$i = the\ index\ of\ a\ reference\ bounding\ box$$

$\mu_i = ground - truth\ label$
$N_{u_i} = total\ number\ of\ the\ reference\ bounding\ box\ labelled\ with\ \mu_i$
$p_i =$ predicted probability of reference bounding box i belonging to class μ_i
$\gamma = hyper\ parameter\ (determined\ by\ Cross\ Validation - set\ to\ 2\ in\ this\ case)$

$NOTE$: $(1 - p_i)^{\gamma}$ controls the weights of the hard − classified examples.

The bounding regression loss function L_{reg} (***Equation 4***) is an extension of the L_1 loss function used in faster R-CNN [12].

$$L_{reg} = \frac{1}{N_{total}} \sum_i [u_i > 0]\ smooth_{L_1}(t_i - t_i^*) \quad (4)$$

$in\ which$

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2,\ if\ |x| < 1 \\ |x| - 0.5,\ othereise \end{cases} \quad (5)$$

$where$

$t_i = predicted\ adjustment\ parameters$
$t_i^* = adjustment\ parameters\ relative\ to\ the\ ground - truth\ bounding\ box$
$N_{total} = number\ of\ the\ foreground\ reference\ bounding\ box$
$\mu_i = Assigned\ label\ of\ the\ reference\ bounding\ box$

3.2 Evaluation criteria

Intersection-over-Union (IoU) overlap ratio between the predicted and ground-truth bounding box is computed as an evaluation metric to measure the localization accuracy of the current method. It is defined as follow:

$$IoU = \frac{V_{predict} \cap V_{groundtruth}}{V_{predict} \cup V_{groundtruth}} \quad (6)$$

$where$

$$V_{predict} = volume\ of\ the\ predicted\ bounding\ box$$
$$V_{groundtruth} = volume\ of\ the\ ground - truth\ bounding\ box$$

The higher the overlap ratio between those two bounding boxes, the higher the scores, as it is helpful to detect whether the prediction match as closely as possible to the ground truth.

Detection rate is used to check how many percent of a specific organ can successfully detected by the model. Formula is shown in ***Equation 7***.

$$organ\ detection\ rate = \frac{\sum_{i=0}^{N} organ\ exist\ in\ p_i \cap organ\ exist\ in\ gt_i}{\sum_{i=0}^{N} organ\ exist\ in\ gt_i} \quad (7)$$

$where$

$$p_i = prediction\ results\ for\ file\ i$$
$$gt_i = ground\ truth\ results\ for\ file\ i$$
$$N = number\ of\ images\ in\ testing\ set$$

3.3 Data Preprocessing and Augmentation

The inputs images are resampled to a uniform spatial resolution using bilinear interpolation, and their voxel intensities are rescaled to [0, 1] in the data preprocessing process. In the training stage, data augmentation is also applied to boost the model's robustness and generalization ability and avoid overfitting. The input images are randomly selected to be cropped to a sub-scan. For those sub-scan images, the organs bounding boxes truncated during this process are treated as background. Finally, the input images are randomly shifted by a maximum of 10 mm along the x and y axes.

To achieve higher localization accuracy of the 3D RPN model, two edge detection techniques are implemented to emphasize the boundary of key organs in both CT and MRI. The model then uses the generated gradient magnitude of the raw images as the input. Edge detection is worked by detecting discontinuities in brightness for inputting images. Typically, it is used to find the approximate absolute gradient magnitude at each point in an input grayscale image. In this work, the Sobel edge detections and a novel method, called multi-resolution local image normalization, are used to pre-process the input CT and MR image to remove some background noise and emphases the organ's contour information.

3.3.1 Sobel Edge Detection (Sobel Operator)

Sobel operator, invented by Sobel Irwin Edward in 1970, is one of the most popular edge detection techniques in image processing and computer vision. The main advantage of this operator is its simplicity because of the approximate gradient calculation. In this project, a 3D Sobel operator consisting of a $3 * 3 * 3$ convolution kernels is used to extract the edge information in 3D images. The implementation of this operator is based on the idea published by Irwin Sobel [13]. ***Figure 3*** shows an example of x kernels in this $3 * 3 * 3$ kernel. The x kernel is just rotated as required to obtain the kernel in the y and z directions.

-1	-3	-1
-3	-6	-3
-1	-3	-1

$Plane\ x-1$

0	0	0
0	0	0
0	0	0

$Plane\ x$

1	3	1
3	6	3
1	3	1

$Plane\ x+1$

*Figure 3 - X Direction Kernels of 3*3*3 Kernel*

Convolution is done by moving the kernel across one pixel from the original image at a time. The gradient of a pixel is a weighted sum of pixels in the 3-by-3-by-3 neighborhood. The final gradient magnitude for a particular pixel is calculated by:

$$G = \sqrt{G_x^2 + G_y^2 + G_z^2} \quad (8)$$

$where$

$$G_x = gradient\ in\ x\ direction$$

$$G_y = gradient\ in\ y\ direction$$

$$G_z = gradient\ in\ z\ direction$$

In this work, this Sobel filter is applied to enhance the edge in the original CT and MR images to improve the localization accuracy. The modified RPN network takes gradient magnitude of images as input, followed by the same data preprocessing and augmentation as the original model. All images use the Sobel operator to generate their corresponding gradient magnitude. By removing some noise pixels, the final gradient image is generated. The examples of gradient magnitudes of CT and MR images are shown in ***Figure 4***.

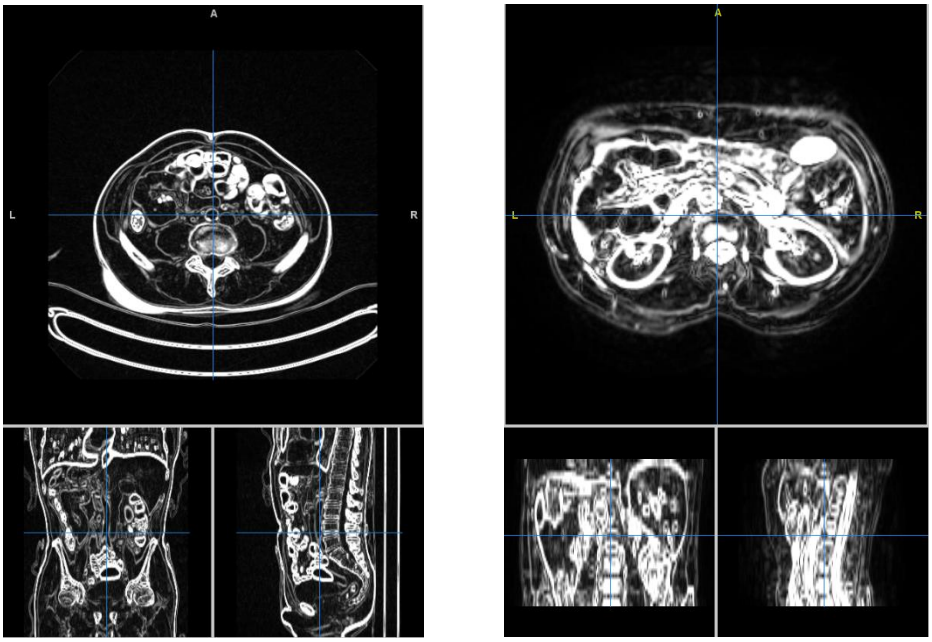

Figure 4 - Sample of Abdominal CT (Left) and MR (Right) Images in Gradient Magnitude

3.3.2 Multi-Resolution Local Image Normalization

As the Sobel operator can be affected by the noise ratio in the images, the gradient magnitudes between CT and MR images are very different. Therefore, a novel preprocessing technique, called multi-resolution local image normalization, is proposed. It aims to emphasize some weak edges that could be ignored by standard gradient filters and mitigate the intensity discrepancy between MR and CT images to make them more similar. As this method can remove some homogenous regions during calculation, different threshold value selection can affect the noise ratio in the generated image. By combining gradient maps in different resolutions, some weak edges can be shown through this method. It only applies to 2D images and has not yet been expanded to implement in a 3D manner fully. Therefore, by considering 3D images as continuous 2D slices, images are normalized slicewise. Individual gradient maps in 2D are finally combined to a new 3D gradient image.

For each individual slice, it is downscaled to several sizes, followed by calculating the gradient magnitudes of all resized images. Then these gradient maps are processed separately. A sliding window is defined to determine the image region normalized by the algorithm, and its size is image dependent. The normalization happens window by window. Each window calculates the standard deviation of the region and compares it with a pre-defined threshold to decide whether this region is homogeneous, the region where the pixel value is quite similar. Some homogeneous regions are removed to reduce the noise ratio in the normalized map. Finally, the algorithm resizes the normalized map back to the original image size and combines all the normalized maps calculated using different-scales images by adding the pixel value.

The final image intensity is normalized between 0 and 1. The pseudocode is shown below:

Algorithm 1 Multi-Resolution Local Image Normalization

```
img = ReadImgData(file)
for curSlice ← 1 to totalSlice do
    I = img[:, :, curSlice]
    nLevel =User predefined value
    imgScale = 1./2.^[0 : nLevel − 1]
    hW =User predefined sliding window size
    mulGM = Empty matrix has same size as I
    for level ← 1 to nLevel do
        tmpImg = resize(I, imgScale[level])
        gradientImg = gradient(I)
        stdI = std(gradientImg)                ▷ std means standard deviation
        normGM = Empty matrix has same size as gradientImg
        for r ← hW + 1 to gradientImgRow − hW do
            for c ← hW + 1 to gradientImgColumn − hW do
                windowImg = gradientImg[r−hW : r+hW, c−hW : c+hW]
                if std(windowImg) < stdI/threshold then
                    normGM[r, c] = 0
                else
                    windowImg = normalize(windowImg)
                    normGM[r, c] = windowImg[hW + 1, hW + 1] ▷ Use the
center value from the window as the value for pixel in [r, c]
                end if
            end for
        end for
        mulGM = mulGM + resize(normGM, IRow, ICol)
    end for
    mulGM = normalize(mulGM)
    img[:, :, curSlice] = mulGM
end for
writeImgData(img, file)
```

In this work, the intensity is scaled from 0-1 to 0-255 and the pixel type is reset to be unsigned 8-bit integer. This process simplifies the subsequent manual noise removal and reduces the stored image size. By removing some noise pixels, the final multi-resolution gradient image is generated. The examples of multi-resolution gradient magnitudes of CT and MR images are shown in ***Figure 5***.

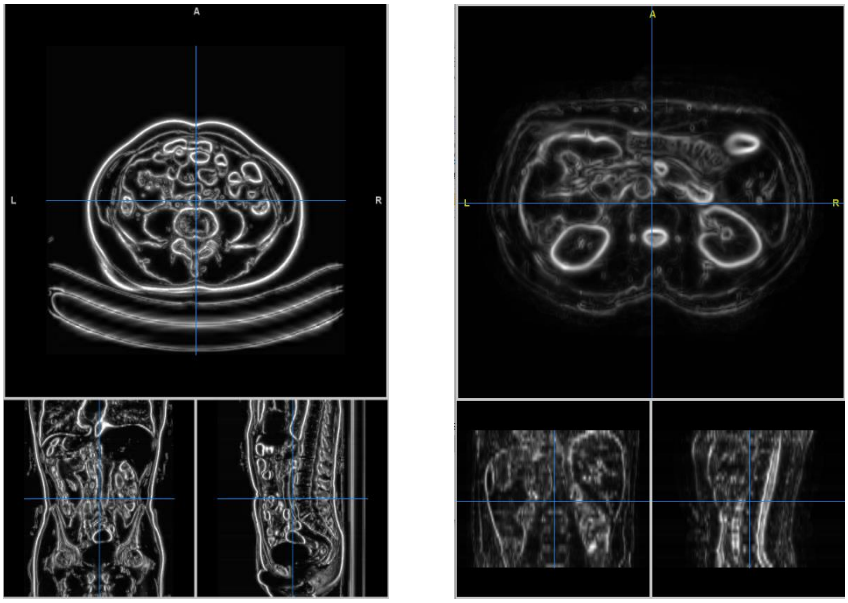

Figure 5 - Sample of abdominal CT (Left) and MR (Right) images in Multi-Resolution Gradient Magnitude

4 Design

The section introduces the design of the experimental datasets, image annotation and the experiments that are carried out to present the final result.

4.1 Datasets

In this project, the experiments are conducted on three datasets, including two abdominal clinical CT datasets and an abdominal MR dataset in T2 weighted. Detailed information about these three datasets is summarized in ***Table 4***.

Table 4 – Parameters of The Datasets Used in This Project

Dataset	Subset	Image Number	Slice Size	Slice Number	In-plane resolution[mm]	Slice Thickness[mm]
Abdominal CT	Training	131	512*512	74-987	0.56-1.00	0.70-5.00
	Testing	70	512*512	42-1026	0.60-0.98	0.45-5.00
Independent Abdominal CT	Testing	30	192*160	256	2	2
Abdominal MR	Testing	20	256*256 320*320 288*288	26-39	1.36-1.74	7.7-9

4.1.1 Abdominal CT Dataset

The first abdominal clinical CT dataset is built on MICCAI Liver Tumor Segmentation (LiTS) challenge dataset [14], which mainly focus on evaluating the liver and liver segmentation algorithm in contrast-enhanced abdominal CT scans. The image scans were gathered from 6 various medical centers around the world. In this project, the model is rebuilt and trained on CT images in this dataset. Its evaluation results show the model localization accuracy on CT images in this dataset, which test the model's capability on detecting organs in CT images.

The annotation files of CT images in the LiTS dataset are published by Xu *et al*. [9] publicly. They built the annotations of bounding boxes of 11 body organ for all images in this dataset. It consists of 201 CT scans, including 131 images for training and 70 images for testing. In these annotations, 11 body organs bounding boxes are generated, including liver (131/70), left lung (52/21), left kidney (129/70), right lung (52/21), right kidney (131/69), left femoral head (109/66), right femoral head (105/66), bladder (109/67), heart (53/28), spleen (131/70) and pancreas (131/70). In the parentheses, the number illustrates the number of the annotated organs in the training set and testing set of the abdominal CT dataset. The sample image of abdominal CT images from LiTS challenge is shown in ***Figure 6***.

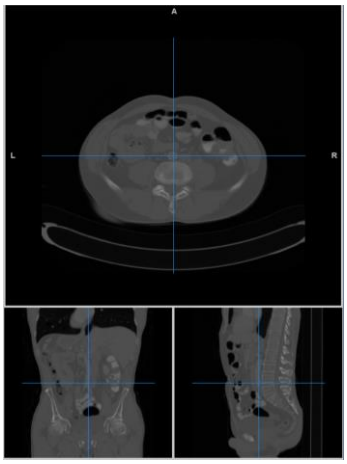

Figure 6 - Sample of Abdominal CT Scans from LiTS Dataset

4.1.2 Independent Abdominal CT Dataset

The independent abdominal clinical CT dataset is built on Multi-Atlas Labelling Beyond the Cranial Vault (BTCV) Segmentation challenge dataset [15], an ongoing benchmark for

evaluating automated segmentation approaches on abdominal organs. The CT scans are randomly selected from a combination of an ongoing colorectal cancer chemotherapy trial and a retrospective ventral hernia study. CT data in the BTCV challenge provides both raw CT image and its corresponding label file with 13 body organs included. Therefore, it is suitable for being a test set.

As the label image of the testing set in this challenge is not provided, only 30 CT images in the training set are chosen as an independent CT testing set to evaluate the model. In these 13 body organs, only the accuracy of six common organs that exist in the abdominal CT dataset is recorded. Six organs include liver (30), right kidney (30), left kidney (30), bladder (28), spleen (30), pancreas (30). In the parentheses, the number illustrates the number of the organs labeled in this dataset. All images slice size is 192 * 160, with in-plane resolution and slice thickness to be 2mm. CT image in this dataset has different image slice sizes and in-plane resolutions from the one in the abdominal CT dataset. Therefore, they can be used as an independent test set to evaluate further the robustness and generalization ability of the model when handling CT images with various configurations. The sample image of abdominal CT images from the BTCV challenge is shown in ***Figure 7***.

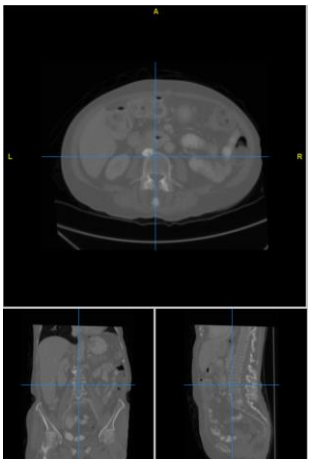

Figure 7 - Sample of Abdominal CT Images from BTCV Dataset

4.1.3 Abdominal MR Dataset

The abdominal MR dataset is built on the Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS) challenge dataset [16], aiming to evaluate algorithms on abdominal organ segmentation in CT and MR images. 20 MR images in the T2-SPIR sequence in the training set of the CHAOS dataset are used as this MR dataset. SPIR (Spectral Pre-Saturation Inversion Recovery) represents a hybrid imaging sequence and uses a T2-weighted contrast mechanism, which is more suitable for abdominal organs study. In addition, by suppressing fat tissue around the body organs, the organ boundary becomes clearer. The adjacent body organs and tissues are more separable from each other because of their high signal value. The purpose of this MR dataset is to analyze the capability for the model performing the multi-modality task. This challenge is selected because it contains MR images in T2-weighted, and their label images are provided. After proper image preprocessing, the difference between CT and MR images can be decreased.

In each labeled image, four abdominal organs are annotated existing in the abdominal CT dataset, including liver (20), right kidney (20), left kidney (20), and spleen (20). In the parentheses, the number points out the number of the organs labeled in this dataset. The

datasets are required by a 1.5T Philips MRI, which produces 12-bit DICOM MR images. The in-plane resolution of MR images is between 1.36 – 1.74mm, while the slice thickness is between 7.7-9mm. The sample image of abdominal MR images from the CHAOS challenge is shown in ***Figure 8***.

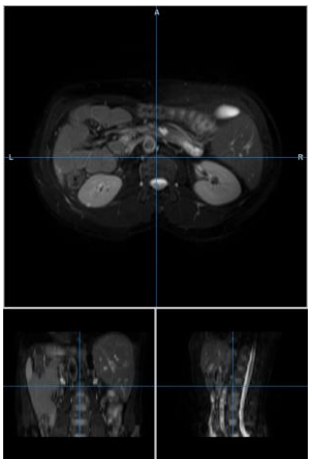

Figure 8 - Sample of Abdominal MR Images in T2-SPIR from CHAOS Dataset

4.1.4 Bounding Box Annotation

The ground-truth bounding box annotation is stored in text files. N lines of records are stored in each text file, representing N organs appearing in the MR and CT image. Each line's data are recorded in the following format:

$$Name\ Label\ x_0\ x_1\ y_0\ y_1\ z_0\ z_1$$

where

$$Name = organ\ name$$
$$Label = organ\ label$$
$$x_0 = starting\ coordinate\ of\ the\ bounding\ box\ on\ x-axis$$
$$x_1 = ending\ coordinate\ of\ the\ bounding\ box\ on\ x-axis$$
$$y_0 = starting\ coordinate\ of\ the\ bounding\ box\ on\ y-axis$$
$$y_1 = ending\ coordinate\ of\ the\ bounding\ box\ on\ y-axis$$
$$z_0 = starting\ coordinate\ of\ the\ bounding\ box\ on\ z-axis$$
$$z_1 = ending\ coordinate\ of\ the\ bounding\ box\ on\ z-axis$$

Figure 9 shows an example of the annotation files. The first line of the file denotes a liver bounding box, which starts at (68, 136, 94) and ends at (312, 377, 141). All the coordinates are measured in voxels. The X-axis points from right hand to left hand; the Y-axis points from face to back; the Z-axis points from feet to head. All the ground-truth annotation files used in either the training or evaluation period are consistent with this format.

```
liver 1 68 312 136 377 94 141
kidney-r 2 130 213 253 348 84 119
kidney-l 3 304 394 269 355 90 127
femur-r 4 104 172 237 305 25 40
femur-l 5 343 410 244 311 24 38
bladder 6 194 315 180 319 24 45
heart 7 212 351 141 272 133 162
spleen 10 331 442 247 358 109 134
pancreas 11 217 346 193 260 100 122
```

Figure 9 - Sample Annotation File

4.2 Experiments Design

In this work, the experiments are conducted based on the three datasets mentioned above. For the abdominal CT dataset, there are 201 abdominal CT scans, with 131 of them for training and the remaining 70 images for testing purposes. The independent abdominal CT dataset consists of 30 abdominal CT images used to test the model's performance on the external CT dataset. The abdominal MR dataset is composed of 20 MR images in T2-weighted, used to test the model's capability on multi-modality tasks. The 3D RPN model is trained based on the training set from the abdominal CT dataset, followed by an evaluation process using 70 images from the abdominal CT dataset, 30 images from the independent abdominal CT dataset. Then the 3D RPN model is extended to a multi-modality model and 20 MR images from the abdominal MR dataset are evaluated.

To further improve the localization of the original model, two extra image preprocessing techniques are applied separately to every image in three datasets. Therefore, in total, three models are built in this project. All models follow the same training and testing process described above, except that their input images are different. For the raw multi-modality model, the input image is the raw CT and MR images in three datasets. For the gradient multi-modality model, applying the Sobel operator as the preprocessing technique, the input image is the gradient magnitude of CT and MR images in three datasets. Finally, for the multi-resolution gradient multi-modality model, which applies multi-resolution local image normalization as preprocessing step, the input image is the multi-resolution gradient magnitude of images in three datasets.

5 Implementation

This section covers the implementation of the dataset setup, annotation extraction, data preprocessing (with model extension), model training configuration, model output and evaluation.

5.1 Datasets Setup

5.1.1 Image File Conversion

This project used third-party libraries, such as Insight ToolKit (ITK) and SimpleITK, to process all images from training and testing set. All model input images were stored using MHD format, an ITK meta image header, and RAW format. Therefore, CT and MR images in three datasets were converted from either Neuroimaging Informatics Technology Initiative (NIfTI) or Digital Imaging and Communications in Medicines (DICOM) to MHD. As both third-party libraries are

cross-platform systems and available, the image conversion script was written in python. Through this process, the processed MR images were in the same format as other CT images, therefore, it can be fed into the network to get the prediction.

DICOM to NIFTI

MR images from the abdominal MR dataset were in DICOM format, which is used worldwide to store, exchange, and transmit medical images. Images in DICOM were converted to NIfTI format using a function called "dicom_series_to_nifti()" in the "dicom2nifti" python library.

NIFTI to MHD

CT images from the abdominal CT dataset and independent abdominal CT dataset downloaded online and processed MR images were in NIfTI format, which is the most commonly used format for multi-dimensional neuroimaging data. In this image conversion stage, functions like " ImageFileReader() " and " ImageFileWriter()" were used to read and write the image data with specific image type. In the meantime, image pixel type was set to be signed integer to downsize the image size, and all image orientations were unified. Detailed image orientation setup is described in 5.1.2.

5.1.2 Image Orientation Setup

All the model input images were treated in "R (Right to Left) A (Anterior to Posterior) I (inferior to Superior) " orientation, whose direction matrix in ITK equals to [1, 0, 0; 0, 1, 0; 0, 0, 1]. Images stored in different orientations may cause the disorder of organs' absolute location (e.g., the liver in one image is on the left, while in another image is on the right). Therefore, images stored in different orientations (e.g., "L, A, I") were converted to "R A I" orientation before fed into the network. This process can be done by flipping the pixel value in the image data and then storing the image in default orientation in "RAI." The standard model input image should have the same image orientation as the one in ***Figure 10*** stored in default direction matrix ([1, 0, 0; 0, 1, 0; 0, 0, 1]).

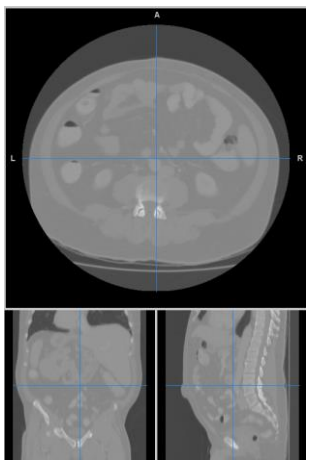

Figure 10 - Sample CT Image

In some cases, although two abdominal images shared the same direction matrix, their initial image orientation may be different. Therefore, the way to correctly flip an image is image-dependent. ***Figure 11*** shows two example images in which both direction matrices are [-1, 0, 0; 0, -1, 0; 0, 0, 1], but their organ locations are completely different. The liver in the left image is on the right, while the right image is on the left. The left image's orientation is consistent

with ***Figure 10***, but the direction matrix is different. Therefore, by flipping the pixel value on the x-axis and y-axis, the image direction matrix can be reset to [1, 0, 0; 0, 1, 0; 0, 0, 1] without seeing visual changes on the output image. However, for the right images, its initial image orientation is different from ***Figure 10***, as the liver is on the left. The algorithm should first change the image orientation to ensure the right image shares the same image orientation as ***Figure 10*** before changing its direction matrix. The processed right image is shown in ***Figure 12***.

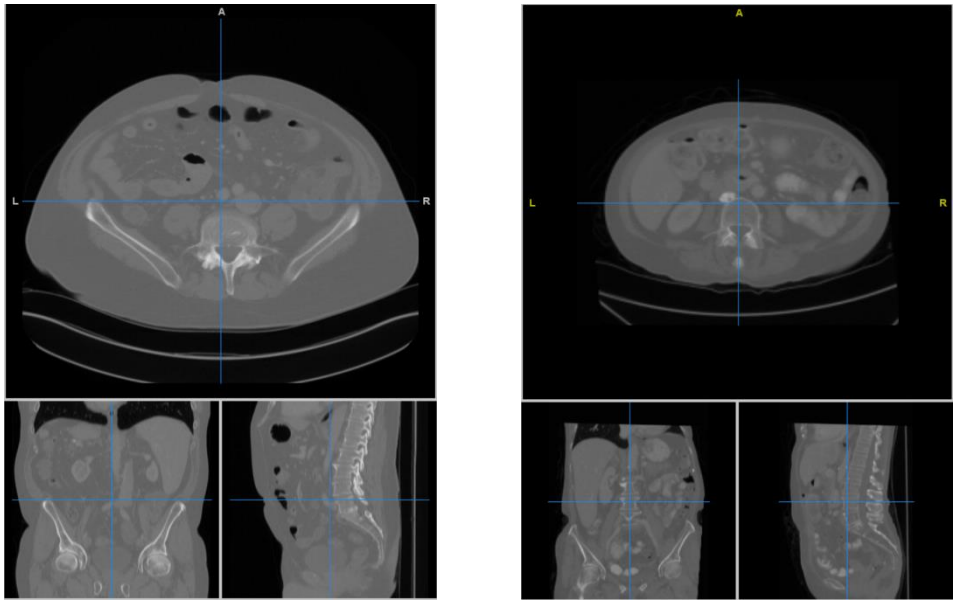

Figure 11 - Two CT images Share Same Direction Matrix. Left Image: liver on right, Right Image: liver on left

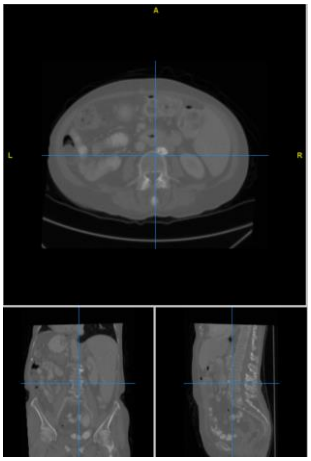

Figure 12- Processed Right Image

5.2 Annotations Extraction

The annotation files for the abdominal CT dataset were provided by Xu *et al.* [9] publicly. Therefore, no more action was needed. For annotations for independent abdominal CT dataset and abdominal MR dataset, the organ's bounding box was computed according to their segmentation masks in the labeled image. For convenience, the format of generated annotations was consistent with one for the abdominal CT dataset. As shown in ***Figure 13***, pixels in the region belonging to a specific organ have identical pixel values. ***Table 5*** and ***Table 6*** summarize the detected organs' corresponding pixel values in different datasets.

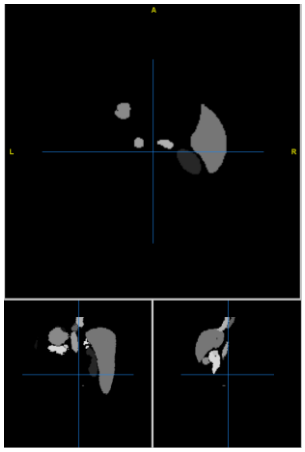

Figure 13 - Example Labelled Image

Table 5 – Organs and Their Associate Pixel Value in Labeled Image in Independent Abdominal CT Dataset

Organ Name	Pixel Value
Liver	6
Right Kidney	2
Left Kidney	3
Bladder	4
Spleen	1
Pancreas	11

Table 6 - Organs and Their Associate Pixel Value in Labeled Image in Abdominal MR Dataset

Organ Name	Pixel Value
Liver	63
Right Kidney	126
Left Kidney	189
Spleen	252

PNG to NIfTI

Labeled images for the abdominal MR dataset are provided in PNG format with continuous 2D slices. Through a function called "ImageSeriesReader()" in the SimpleITK library, separate slices belonging to one image are combined to form individual 3D NIfTI images.

Extract Annotation File Through NIfTI Image

To correctly extract the annotations, the labeled images stored in different orientations were set to "RAI" orientation before extraction. Their image's orientation was paired with their corresponding raw image to ensure the validity of the annotations. The extraction of the bounding box was by counting the maximum and minimum index of a specific pixel value and finally generating the bounding box.

5.3 Data Preprocessing

5.3.1 Data Preprocessing in Model

In the model data preprocessing stage, according to their image modality, the input images

were processed separately. For both CT and MR images, the input images were resampled to a uniform spatial resolution, which is $2.0 * 2.0 * 2.0\ mm^3$. Axial slices were center cropped with a maximum physical size of $300 * 300\ mm^2$. Voxel intensities were rescaled from different ranges Hounsfield Unit (HU) (shown in ***Table 7***) to [0, 1]. The Intensity outside this range was clipped. As the raw CT and MR images had very different pixel intensity ranges, their corresponding parameters in the raw multi-modality model were set differently. The intensity range between CT and MR images was similar in the two updated models because the raw images have already been processed using the external image preprocessing technique.

Table 7 - Intensity Range for Different Models

Model Name	Intensity Range
Raw Multi-Modality Model	[-1000, 1600] for CT, [0, 400] for MR
Gradient Multi-Modality Model	[0, 5000]
MulGM Multi-Modality Model	[0, 255]

The images then went through the data augmentation process illustrated in **Section 3.3**. In this process, to maintain enough context information, the minimum number of slice of sub-scan was set to 50.

5.3.2 Image Preprocessing with Sobel Operator

To improve the localization accuracy of the original model on the multi-modality task, an image preprocessing technique, using the Sobel operator to extract edge information, was applied to all training and testing images. Their corresponding gradient magnitudes were generated and inputted into the network. A function called "SobelEdgeDetection()" in ITK was used to perform the edge detection with the Sobel operator in 3D images. After that, around 10% of the top largest pixels in the gradient image were clipped to a maximum threshold. It was used to remove some noise and make the edge clearer. In the project, the threshold was set to 5000. Any pixel value cannot exceed this threshold. The final gradient image shares the same origin, spacing, and direction matrix as the raw image.

5.3.3 Image Preprocessing with Multi-Resolution Local Image Normalization

To further reduce the impact of pixel noise and narrow the intensity discrepancy between CT and MR images, another novel image preprocessing technique, using multi-resolution local image normalization, was proposed. It was applied to all images in three datasets. Raw images' corresponding multi-resolution gradient magnitudes were generated and inputted into the network.

Several hyper-parameters are required to be set when using this method. For different sizes of 2D image slices, they were downsampled individually. In this project, the larger the slice is, the more it would be downsampled. The downscale level for CT scans in the abdominal CT dataset (slice size: 512*512) was set to be 4, while for the other two datasets, the downscale level was 3, as images in these two datasets had smaller image slice size. For the sliding window size, the value was 3 for all CT images and 13 for MR images, as more noises were included in MR images. The last parameter used to determine the threshold to remove the homogenous regions was set to 7 for three datasets.

After the multi-resolution gradient images were generated, for CT scans, pixels greater than 200 were set to 255 to remove some noise in processed images. No further noise removal was required for MR images, as only a few pixels were greater than 200. The final gradient image

shares the same origin, spacing, and direction matrix as the raw image.

5.4 Model Training

With the aid of ITK, the 3D RPN was mainly built on the Caffe framework [17]. This model used backpropagation [18] and stochastic gradient descent algorithm to train the deep learning network in the training stage. The learning rate was set to 10^4 for 1000 training epochs. The model parameters were saved every 20 epochs. The final model was the one generated in the final iteration. The weight decay was set to $5 * 10^4$ and momentum parameter was 0.99. RPN layers' weights were initialized using a zero-mean Gaussian distribution, whose standard deviation was set to 0.01, and the Xavier algorithm [19] was used for other layer's initialization.

5.5 Model Output

The model outputs were stored in TXT files, one file per testing image. Each result file includes the predicted organs' bounding box location. As shown in ***Figure 14***, each line corresponds to a predicted organ bounding box location printed in the format of $(organ\_label, x_0, y_0, z_0, x_1, y_1, z_1, confidence\_score)$. All the coordinates are measured in voxels. The X-axis points from right hand to left hand; the Y-axis points from face to back; the Z-axis points from feet to head. The printed bounding box coordinate is calculated based on its relative location in each CT image. For example, the first line of the file denotes a liver bounding box starting at (71.182124, 132.060429, 93.690745) and ending at (318.135451, 376.844498, 143.574447) with a 0.990949 confidence score.

```
1 71.182124 132.060429 93.690745 318.135451 376.844498 143.574447 0.990949
4 126.085510 262.071732 83.537687 218.550757 350.948825 120.062632 0.994848
5 301.923174 269.377122 88.677877 395.915674 351.739402 123.325559 0.999740
6 107.293792 237.422204 22.988408 178.244299 304.131586 38.302235 0.995515
7 342.465095 243.415939 23.223446 412.485038 309.216074 39.245315 0.995955
8 191.740610 197.323477 26.056382 309.341623 337.281698 45.463960 0.994350
9 210.598517 140.201671 129.742337 359.943733 269.729788 162.166667 0.987136
10 329.814988 246.207263 111.044865 446.337919 357.466945 135.597362 0.997465
11 224.951126 196.819416 100.380164 378.926318 293.285327 121.893543 0.996283
```

Figure 14 - Sample Model Predicted Output File

The following table summarize the organs name that each organ label corresponding to in the model output.

Table 8 - Organ Name and Its Corresponding Organ Label in Model Output

Organs	Organ Label
Liver	1
R-lung	2
L-lung	3
R-kidney	4
L-kidney	5
R-femur	6
L-femur	7
Bladder	8
Heart	9
Spleen	10
Pancreas	11

5.6 Model Evaluation

IoU accuracy was recorded once the model outputted the predicted result. In this project, only organs that existed in both predicted result and annotation were valid. If an organ

predicted by model cannot be found in annotations, the IoU calculation of the current organ was ignored. When calculating the single organ's IoU accuracy, the coordinates of the starting and ending points in the predicted bounding box and ground-truth bounding box were extracted. The organ's IoU was then calculated using ***Equation (6).*** The mean IoU of an organ was finally computed to represent the organ average accuracy. The mean IoU accuracy for a specific organ is calculated in two different ways: 1. Dividing the sum of the organ IoU by the number of valid organs. 2. Dividing the sum of the organ IoU by the number of organs in annotations. Also, the organ detection rate was recorded to reveal the percentage of organs that the model can successfully detect. The detailed description and formula are covered in **Section 6**.

6 Results & Discussions

In this section, three models trained on different input images were evaluated on abdominal CT dataset, independent abdominal CT dataset, and abdominal MR dataset individually to analyze the model's localization ability in CT and MR images. Some discussion and future works based on these experiments are covered as well.

6.1 Qualitative Results

Visualization provides an intuitive way to observe the performance of the model. It helps to notice whether the predicted bounding box outlines the RoI in a 3D medical image. In this project, cropping the testing images according to the bounding box of organs is used as a visualization technique. The results are partially visualized below. Each example shown below was extracted from the same slice of the ground-truth liver and predicted liver.

Figure 15 compared two livers cropped by a ground-truth bounding box and a predicted bounding box, respectively. Both images were cropped from the same image in the testing set of the abdominal CT dataset. In this example, the IoU between two bounding boxes was around 0.8928. As shown in the figure, their cropped regions were similar.

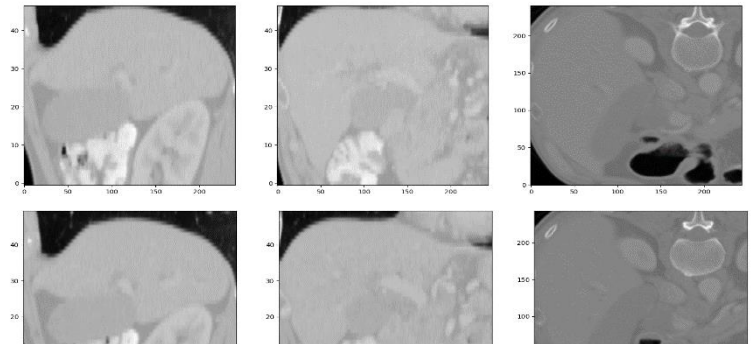

Figure 15 – Sample Livers cropped by Ground-Truth Bounding Box (Top) and Predicted Bounding Box (Bottom) in Abdominal CT Dataset. Liver Localization Accuracy: 89.28%

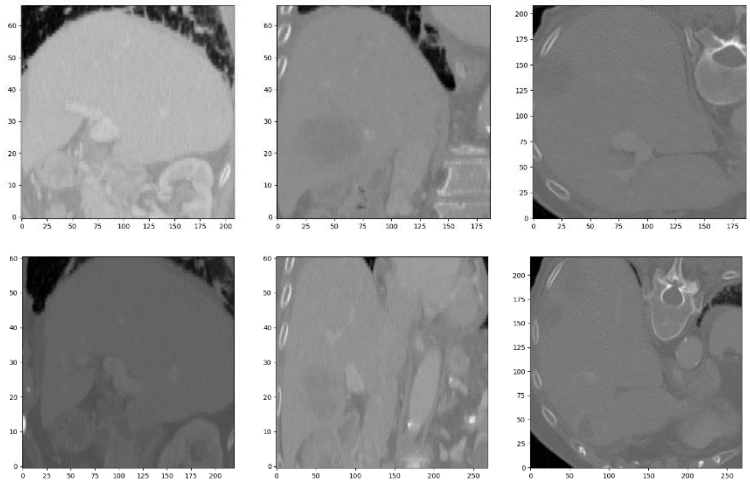

Figure 16 - Sample Livers cropped by Ground-Truth Bounding Box (Top) and Predicted Bounding Box (Bottom) in Abdominal CT Dataset. Liver Localization Accuracy: 59.06%

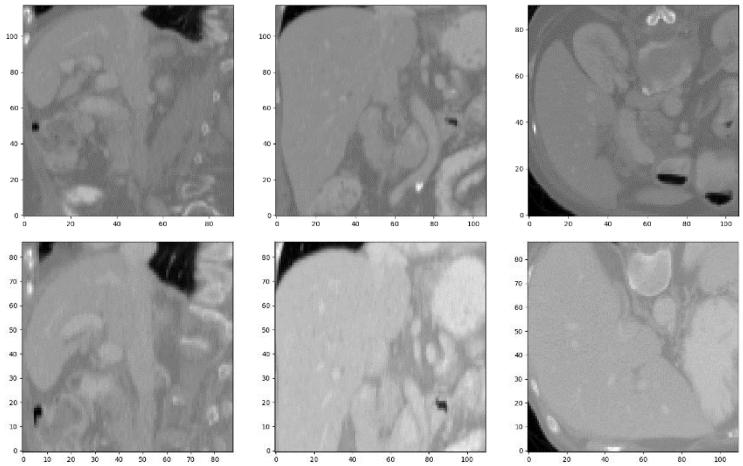

Figure 17 - Sample Livers cropped by Ground-Truth Bounding Box (Top) and Predicted Bounding Box (Bottom) in Independent Abdominal CT Dataset. Liver Localization Accuracy: 69.95%

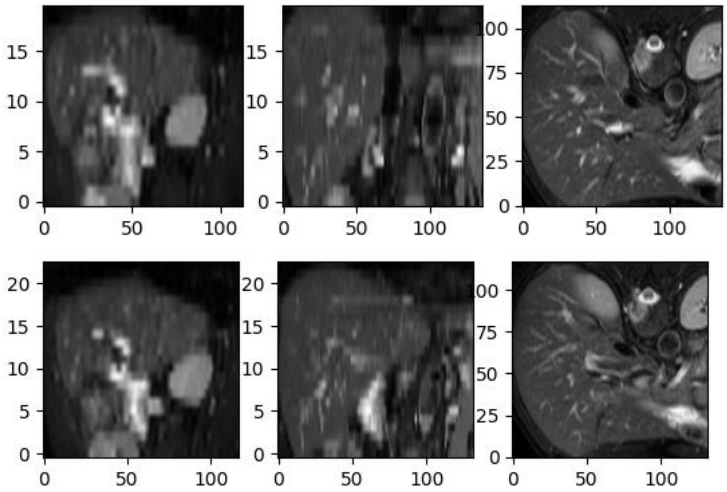

Figure 18 - Sample Livers cropped by Ground-Truth Bounding Box (Top) and Predicted Bounding Box (Bottom) in Abdominal MR Dataset. Liver Localization Accuracy: 77.54%

In the abdominal CT dataset, the lowest IoU in the liver is around 59%, and it is shown in ***Figure 16***. Although the predicted bounding box truncated part of the liver boundary, the model still predicted the rough location of the liver with 59% accuracy. For ***Figure 17*** and ***Figure 18***, they showed the sample livers cropped by a ground-truth bounding box and a predicted bounding box in the independent abdominal CT dataset and abdominal MR dataset, respectively. The localization accuracies were 69.95% and 77.54%.

6.2 Quantitative Results

In this experiment, the organ was valid only if it was predicted by the model and annotated in the testing set at the same time. Only the accuracy and the detection rate for the valid organs were calculated. If a predicted organ could not be found in annotations, the IoU calculation of the current organ was ignored. Three models were: raw model, gradient model, and multi-resolution gradient magnitude model (mulGM Model). The raw model represented the multi-modality model trained on raw CT images. The gradient model denotes the multi-modality model trained on images processed using the Sobel filter. The mulGM model was the multi-modality model whose input images were processed using multi-resolution local image normalization.

In general, the mean IoU accuracy was calculated using the following two formulas:

1. Divided the total IoU by the number of the valid organ.

$$Organ\ Avg\ IoU = \frac{\sum_{i=0}^{N} organ\ IoU_i}{\sum_{i=0}^{N} organ\ exist\ in\ p_i \cap organ\ exist\ in\ gt_i} \quad (9)$$

where

$$p_i = prediction\ results\ for\ file\ i$$

$$gt_i = ground\ truth\ results\ for\ file\ i$$

$$N = number\ of\ images\ in\ test\ set$$

2. Divided the total IoU by the number of the organ annotated in the testing set:

$$Organ\ Avg\ IoU = \frac{\sum_{i=1}^{N} organ\ IoU_i}{\sum_{i=0}^{N} organ\ exist\ in\ gt_i} \quad (10)$$

where

$$gt_i = ground\ truth\ results\ for\ file\ i$$

$$N = number\ of\ images\ in\ test\ set$$

The first method was to estimate the localization accuracy of those valid organs, while the second method was to evaluate the model performance in a more general case, as it included the organs that were not detected by the model but annotated in the testing set. To better distinguish these two methods, the mean IoU calculated through the first method was named "valid IoU," while the one calculated using the second method was named "true IoU."

Table 9 - IoU of Different Methods for The Localization of 11 Body Organs. Source: Adapted from Xu et al. [10]

Organs	Methods				
	Humpire et al.[7]	De Vos et al.[8]	3D Faster R-CNN	Xu et al.[9]	Xu et al. [10]
L-lung	84.87	78.21	79.13	84.84	85.18
R-lung	87.78	81.14	78.76	86.88	87.43
Heart	73.08	65.24	71.51	80.52	81.14
Liver	73.42	70.37	74.58	77.83	86.99
Spleen	58.26	50.66	64.11	70.01	85.75
Pancreas	51.03	46.32	55.33	58.56	58.26
L-kidney	59.15	52.07	71.45	75.29	76.11
R-kidney	59.55	57.35	69.49	76.46	79.83
Bladder	49.42	45.56	54.36	58.23	66.40
L-femur	52.29	56.69	67.75	77.26	74.86
R-femur	51.98	57.81	66.47	79.77	75.35
Global	59.45	56.62	66.63	73.01	76.44

Xu *et al*. [10] compared the mean IoU of five models for 11 body organs localization, and the results of this quantitative comparison are shown in ***Table 9***. The number inside this table denotes the organ's mean IoU accuracy, whose unit is in [%]. The minimum organ IoU recorded was around 45%. Based on the visualization results shown in **Section 6.1**, even the organ's IoU was around 60%, the bounding box still located the organ's rough position without truncating the organ too much. Therefore, in this project, an Intersection over Union score over 0.5 (50%) was considered as a good prediction.

6.2.1 Results for Abdominal CT Dataset

Table 10 - Evaluation Results of 11 Body Organs on The Abdominal CT Dataset

Organs	Raw Model			Gradient Model			MulGM Model		
	True IoU [%]	Detection Rate [%]	Valid IoU [%]	True IoU [%]	Detection Rate [%]	Valid IoU [%]	True IoU [%]	Detection Rate [%]	Valid IoU [%]
Liver	76.99	98.57	78.10	77.79	98.57	78.92	76.55	98.57	77.66
Kidney Right	74.69	98.55	75.79	75.81	98.55	76.92	77.58	97.10	79.89
Kidney Left	74.86	98.57	75.95	75.81	97.14	78.04	78.27	97.14	80.57
Femur Right	76.47	98.48	77.65	77.07	98.48	78.26	75.51	98.48	76.68
Femur Left	74.99	98.48	76.14	76.26	98.48	77.43	73.10	96.97	75.38
Bladder	57.67	98.51	58.54	63.11	98.50	64.07	61.44	97.01	63.33
Heart	76.98	96.43	79.84	75.32	92.86	81.12	77.20	100.00	77.20
Spleen	70.72	98.57	71.75	71.03	100.00	71.03	72.08	98.57	73.12
Pancreas	58.42	98.57	59.26	56.69	98.57	57.51	57.12	95.71	59.67
Lung Right	85.17	100.00	85.17	83.73	100.00	83.73	82.69	100.00	82.69
Lung Left	83.38	100.00	83.38	82.71	100.00	82.71	82.19	100.00	82.19
Global	73.67	98.61	74.69	74.12	98.29	75.43	73.97	98.14	75.31

Table 10 summarized the 11 organs' localization accuracy of three different models on the testing set of abdominal CT dataset. Generally, the overall global detection rates were above 98% in three models, with the lowest organ detection rate at 92.86% when detecting heart using the gradient model. In total, this testing set included 70 CT images, therefore, at most 5 organs were ignored by the models, which was impressive. For the localization accuracy, the global true IoU was 73.67% for raw model, 74.12% for gradient model and 73.97% for mulGM model. When considering only the detected organs' accuracy, the global valid IoU in the raw model was 74.69%, in the gradient model was 75.31%, and in the mulGM model was 75.31%. Almost all organs achieved high IoU except for bladders and pancreas, which had relatively low localization accuracy in these three models.

In this dataset, two updated models, the gradient model, and the mulGM model, provided better results than the raw model, and the gradient model performed the best. Compared

with the raw model, the overall localization accuracy in the gradient model was increased except lungs and pancreas. The IoU overlap ratio in the spleen in the gradient model remained roughly the same, while the detection rate was boosted to 100%. Both proposed models detected bladder more accurately, with the accuracy increasing from 57.67% to 63.11% and 61.44%, respectively. Although the detection rate of left and right kidneys slightly decreased, both models improved the performance on kidney detection, especially the mulGM model. The left kidney accuracy rose around 3%, and the right kidney accuracy rose about 4%. However, two updated models lowered the accuracy of the pancreas, which has obscure boundaries in CT images. Also, the accuracy of left and right lungs was decreased. It may be because both edge enhanced techniques emphasized some tissues' contour inside the lungs, which added more noise to this organ in the gradient magnitude and finally reduced the accuracy. For the mulGM model, although all hearts could be detected, compared with the raw model and gradient model, it sacrificed other organs' detection rate to some extent.

Therefore, as can be seen from this result, it indicates that edge enhancement helps improve to locate the organs having clear contour and low background contrast but has limited enhancement on the organ with blurred boundary. Plus, some non-relevant tissues may also be included in the enhanced map, increasing the difficulty of accurate detection. Currently, the performance of mulGM is unstable. Further experiments are needed to optimize this method.

6.2.2 Results for Independent Abdominal CT Dataset

Table 11 - Evaluation Results of 6 Body Organs on The Independent Abdominal CT Dataset

Organs	Raw Model			Gradient Model			MulGM Model		
	True IoU [%]	Detection Rate [%]	Valid IoU [%]	True IoU [%]	Detection Rate [%]	Valid IoU [%]	True IoU [%]	Detection Rate [%]	Valid IoU [%]
Liver	74.79	100.00	74.79	75.56	100.00	75.56	75.66	100.00	75.66
Kidney Right	70.11	100.00	70.11	73.67	100.00	73.67	69.39	93.33	74.34
Kidney Left	69.12	100.00	69.12	74.56	100.00	74.56	72.93	96.67	75.44
Bladder	0.00	82.14	0.00	0.00	85.71	0.00	0.00	85.71	0.00
Spleen	62.59	100.00	62.59	66.03	100.00	66.03	65.81	100.00	65.81
Pancreas	52.46	100.00	52.46	53.01	100.00	53.01	51.03	96.67	52.79
Global	65.82	100.00	65.82	68.57	100.00	68.57	66.96	97.33	68.81

Table 11 described the evaluation result, including six organs' detection rate and localization accuracy, on the independent CT dataset with three different models. In this dataset, the localization accuracy of the bladders remained zero in three models. As the bladder suffers significant variations of appearance in a different state, the reason for accuracy being zero may be that the appearance of the bladder in this testing set was different from the one in the training set. The bladder's result was excluded when calculating the global detection rate and mean IoU. In this dataset, the global True IoU was 65.82% in the raw model, 68.57% in the gradient model, and 66.96% in the mulGM model. All organs were detected in the raw and gradient models, while the global detection rate in the mulGM model was 97.33%. Therefore, the global valid IoU in the first two models were the same as their true IoU, and the one in the mulGM model was 68.81%. Like the results in the first dataset, the accuracy of the pancreas in this dataset remained the lowest.

The overall localization accuracy was lower than the abdominal CT dataset, which may attribute to two factors: 1. The CT images in the abdominal CT dataset were contrast-enhanced, while the one in the independent CT dataset was raw images. 2. The CT image size in the abdominal CT dataset (slice size: 512 * 512) was different from the independent CT dataset (slice size: 192 * 160). However, the performance on this dataset was still good, which proved the robustness and generalization ability of the original 3D RPN model and its

possibility to be used for various CT images.

Two proposed models performed better than the raw model, and the gradient model outperformed the mulGM model in this dataset. The gradient model raised 3% global true IoU (65.82% to 68.57%) without missing any organs, and it improved all organs' localization accuracy except the bladder. Although mulGM model provided a more accurate prediction for the valid organs, it sacrificed some detection rate, which finally resulted in the lower accuracy in true IoU than the gradient model. Its valid IoUs in left and right kidneys were the highest, but the detection rate decreased, which may because of the blurry boundary generated through multi-resolution local image normalization.

With the apparent improvement in global accuracy, the final results pointed out the proposed models' effectiveness. It also indicated that the edge-enhancement technique could be used as an image preprocessing technique to improve the organ localization accuracy in CT images. However, more experiments should be carried on the multi-resolution local image normalization. As training set and this testing set using CT images in different sizes, their number of down-scaling times are different, which may cause the discrepancy between their generated images. Without proper configuration, using this method could lower the model's localization ability.

6.2.3 Results for Abdominal MR Dataset

Table 12 - Evaluation Results of 4 Body Organs on The Abdominal MR Dataset

Organs	Raw Model			Gradient Model			MulGM Model		
	True IoU [%]	Detection Rate [%]	Valid IoU [%]	True IoU [%]	Detection Rate [%]	Valid IoU [%]	True IoU [%]	Detection Rate [%]	Valid IoU [%]
Liver	68.45	95.00	72.06	0.00	0.00	0.00	12.54	20.00	62.68
Kidney Right	61.45	90.00	68.27	36.67	55.00	66.67	75.98	100.00	75.98
Kidney Left	65.27	100.00	65.27	27.95	45.00	62.10	69.63	100.00	69.63
Spleen	61.82	95.00	65.07	34.52	55.00	62.77	55.75	85.00	65.59
Global	64.25	95.00	67.67	24.78	38.75	47.88	53.47	76.25	65.47

Table 12 summarized the four organs' localization accuracy of three different models on the testing set of the abdominal MR dataset. The detection rate and localization accuracy varied significantly among the three models. The overall detection rate in the raw model remained the highest, which was 95%. There were 20 images in this set; therefore, on average, only one organ was not detected by this model, which was an astonishing result. It proved the possibility of building a multi-modality model with only training on a single CT dataset but detecting organs in MR images. However, the detection rate in the gradient model was the lowest with only 38.75%, and the one in the mulGM model was 76.25%. For the localization accuracy, the global valid IoU accuracy for the raw model was 67.67%, for the gradient model was 47.88%, for the mulGM model was 65.37%. With low detection rates in the gradient model and the mulGM model, although their valid IoUs were acceptable, their true IoUs were lower than the raw model. In the raw model, the global true IoU was 64.25%, in the gradient model was 24.78% and in the mulGM model was 53.47%.

In general, the result in the raw model was far better than the gradient model. Using the Sobel operator to preprocess MR images may not be a proper method to improve the accuracy. Especially for liver detection, the detection rate was 0. The poor results may be attributed to the MR image containing more noise than the CT image. The Sobel operator could emphasize the noise edge, which finally resulted in the gradient magnitude of the MR image is very different from the gradient magnitude of CT scans. Therefore, the gradient model was difficult to identify those organ boundaries effectively.

In this dataset, the mulGM model performed much better than the gradient model but still reduced the organ localization accuracy except for kidneys. With 100% detection rates for left and right kidneys in the mulGM model, their localization accuracy increased from 61.45% to 75.98% and 65.27% to 69.63%, respectively, which was a huge accuracy improvement compared with the raw model. As multi-resolution normalization contains some methods to remove the homogeneous regions and noises in input images, it reduced the discrepancy between CT and MR images so that its result was better than the initial gradient version. It also improved the liver detection rate from 0 % to 20% with 62.68% valid IoU, which means this mulGM model compensates for some disadvantages in the gradient model. Although the liver and spleen performance was not good compared with the raw model, the mulGM model could be a potential future development direction to improve both organ localization accuracy in MR and CT images, as its success was shown on kidney detection in MR.

6.3 Analysis & Discussion

As the final results have shown in the experiments, the multi-modality model can effectively and accurately detect organs in CT and MR images without training two different models. Through the results from the independent abdominal CT dataset, it demonstrates the robustness and generalization ability of the original model. The abdominal MR dataset results prove the success of the multi-modality model. Compared with the original model, one updated model, using Sobel edge enhanced image as input, achieves higher localization accuracy in CT images but lower in MR images. For the other model using multi-resolution local image normalization to process the input image, although its accuracy improvement in CT images does not outperform the gradient model, it boosts the kidney accuracy and performs much better than the gradient model while handling the MR images. It could be a technique that further boosts the organ localization accuracy in MR images while maintaining high accuracy in CT images. Currently, its detection rates are relatively low among the three models, which may because it is only a 2D slice-based method instead of fully implemented in a 3D manner, which could potentially lose some spatial information while calculating the gradient. Also, due to time limitations, more parameter-tunings are needed to generate the optimal multi-resolution gradient magnitude. Therefore, multi-resolution image normalization can be listed as a potential approach to improve the model's localization accuracy on both image modalities.

Both edge enhanced techniques cannot efficiently detect livers in MR images due to tissues and vessels surrounding this organ having similar pixel intensity with the liver. As these noises are not removed beforehand, edge detection methods take these noises when calculating the gradient magnitude, which finally causes the blurry of the boundaries in processed images.

Also, it is notable that the bladder and the pancreas achieve the lowest two IoU overlap ratios among 11 detected organs. For the independent abdominal CT dataset, none of the detected bladders are valid. Different states of bladder and neighboring low-contrast structures make it challenging to have high localization accuracy as other organs, such as liver, lung and kidney. Therefore, more effort is needed to tackle this problem.

The visualization result is currently only achieved by cropping the images, which is hard to present multiple organs bounding boxes in one image. Displaying the bounding box directly on the input 3D images may be more straightforward. However, it demands the design of the GUI system. Hence, it could be listed as a future plan of the project development.

6.4 Future work

As the experiment results are shown above, suitable data preprocessing techniques play a crucial role in improving the model's localization accuracy. The results demonstrate that edge enhancement helps improve the localization accuracy of the proposed network in CT images, but the improvement in the MR dataset may be limited. However, the multi-resolution local image normalization contributes to the improvement of kidney localization accuracy in MR images. It indicates that this method could be a technique that boosts the organ localization accuracy in MR images while maintaining high accuracy in CT images. Currently, this method only normalizes the 3D image slice by slice, which could vanish some edges that have a low gradient in x and y directions but a great change in the z-direction. Also, due to the time limitation, parameters in this technique have not been strictly optimized using a method like cross-validation. Therefore, more future experiments should focus on developing a 3D multi-resolution image normalization method and tuning the parameters to optimize this normalization method. Hopefully, the discrepancy between CT and MR images could be reduced as much as possible with the modified version of multi-resolution image normalization.

Furthermore, in this project, due to time limitation, images in the abdominal MR dataset are in low-resolution with only four body organs. To fully evaluate the capability of this multi-modality model, the model should be tested using the MR images containing more body organs and having higher resolutions. To help the model using edge enhanced technique detect livers more accurately in MR images, some preprocessing techniques that can remove some noises surrounding the key organs could be applied before enhancing the edge information. Plus, more training images with different states of the bladder are needed to classify the bladder in various appearances better.

Finally, in the work of Xu *et al*. [10], they applied a density enhancement technique first to the input images before using the edge enhancement. They then combined all three images to form a new input image to boost the model performance. Their result indicates that with density and edge enhancement techniques, the localization accuracy of the organs having apparent contour and low background contrast is increased. In contrast, their effect on the organs with blurred boundaries is limited. Therefore, the multi-modality model can apply this preprocessing process to remove some noise in MR images with density enhancement technique before extracting the edge information.

7 Summary and Reflections

This section outlines the progress against the initial plan, including the obstacles, risks, and a critical evaluation of how the project has been handled. Contribution and reflection are covered as well.

7.1 Project Management

The project had been run with an agile methodology, with tasks being systematically done moving down the GANTT in ***Figure 20***. Management of tasks included a Kanban-style board, which was done using Trello [20]. As shown in ***Figure 19***, the project's main objectives were listed in the left-most column, to keep track of the current process. Different labels denoted the current state of a specific task, green for a completed task, yellow for an ongoing task, and

blue for a future task. Tasks start off in the "To-Do List" column and move through "In Progress" and "Completed."

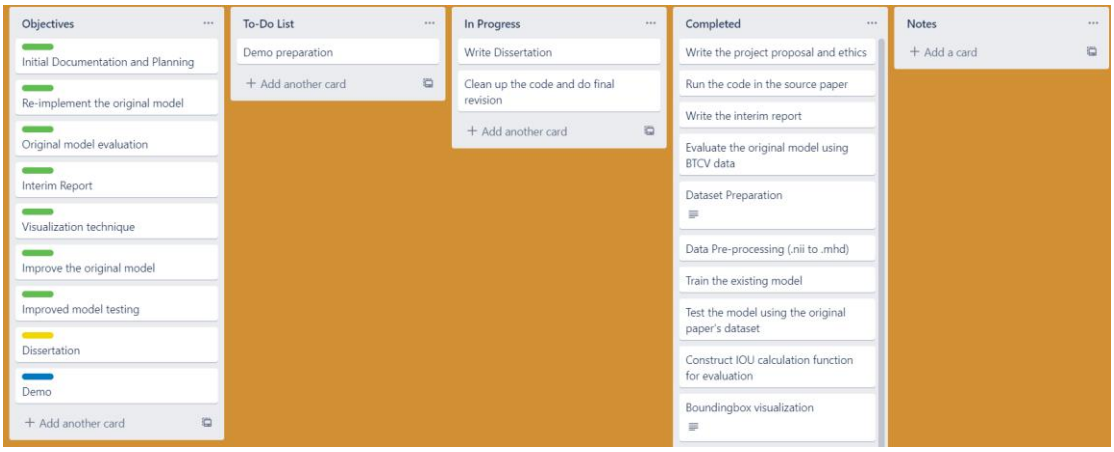

Figure 19 - Trello Board Task Management

The project aims to find a current state-of-art model efficient for automated organ localization in 3D CT images, extend it to a multi-modality model, and improve the model week by week. In the first semester, the focus of the project was to rebuild and evaluate the model's localization accuracy in CT images. In the second semester, the focus was to extend the model to handle MR images and improve the model to increase its localization accuracy in CT and MR images.

Comparing the difference between the two Gantt charts in ***Figure 20***, at first, the plan was to only accomplish the organ localization task in CT images instead of building a multi-modality model. However, after witnessing the robustness shown through the testing on an independent CT dataset, it turned out that building a multi-modality model may be possible. Therefore, more efforts were spent on finding a public MR dataset and evaluating the model's performance in MR images. After acquiring some testing results, further experiments focused on the model improvement, like applying proper image preprocessing technique to process the image first before feeding them into the network.

Thanks to a significant amount of time spent on the first semester on building the model and testing scripts, and setting up the dataset, the progress of MR image evaluation was significantly ahead of schedule. So, a considerable amount of time was spent on model improvement.

Initially, the testing stage was separated from the model improvement process. Nevertheless, throughout the experiments, it is impossible to learn how to improve the model without having any testing results. Therefore, the testing stage was set to be bound with the model improvement.

To sum up, the project's progress was significantly ahead of schedule most of the time, so that more experiments can be carried to further increase the localization accuracy and exploit the model's capability. Overall, the project is well-managed.

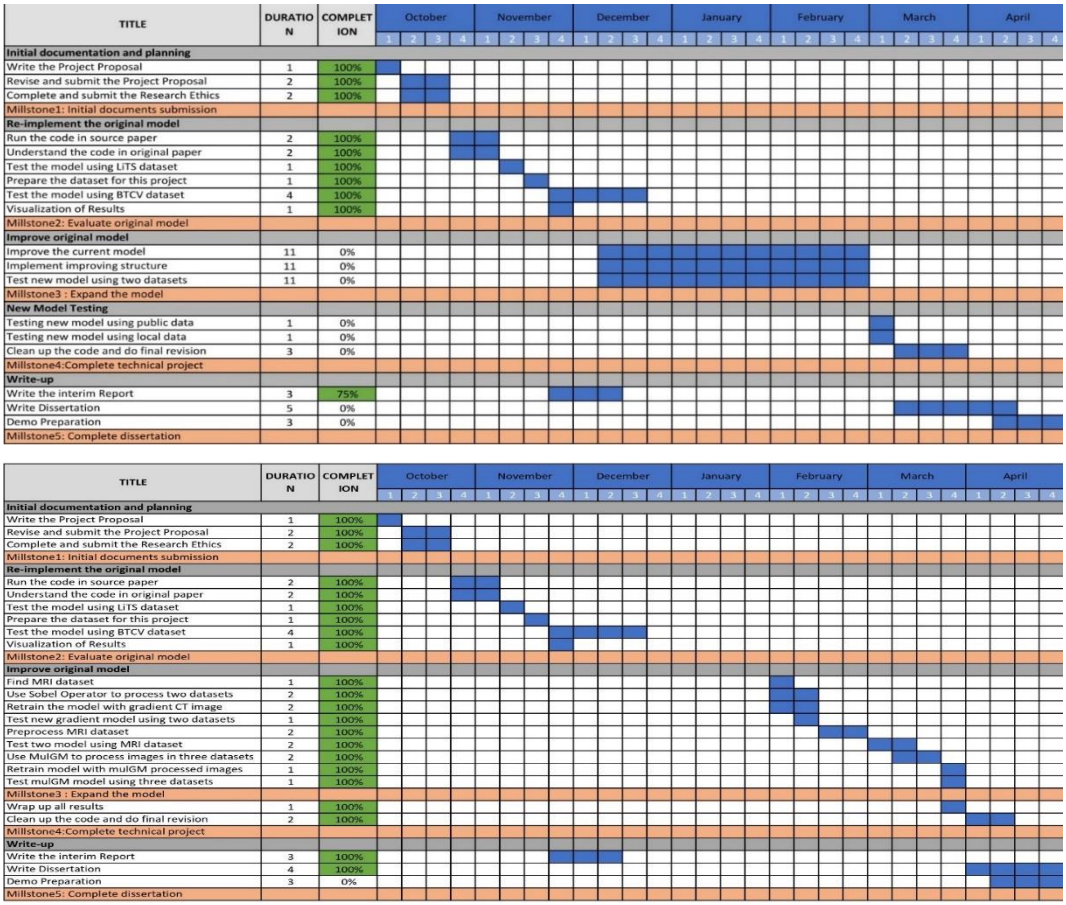

Figure 20 -Original Project Gantt Chart (Top) and Final Project Gantt Chart (Bottom)

7.2 Contributions

In this project, the original 3D RPN model was built and extended to a multi-modality model. This multi-modality model trained on CT datasets can currently locate multiple organs in both MR and CT images with high detection precision and localization accuracy. The result demonstrates the robustness of this model and the hope that this model can be truly applied to some real-world applications, as with one model, organs in two modalities can be accurately detected. It also shows the potential of the multi-modality model and brightens the future that using a single model to detect organs in many imaging protocols.

The second contribution of this project is that by applying the Sobel operator to preprocess the input images, the organ localization accuracy in CT images is improved compared with the raw multi-modality model. With this edge detection technique, some key organ contours can be enhanced and help the model better predict the organs' position.

The third contribution is that a novel multi-resolution local image normalization method is applied to preprocess the input images to enhance the weak edge that could be ignored by the Sobel operator and reduce the intensity discrepancy between CT and MR images. Although its performance on CT images does not outperform the Sobel operator, it still improves the performance in CT image and increases hugely on some organs' accuracy in MR images. Currently, the method only process image in 2D slices, which could eliminate some spatial information through the calculation. Through more thorough experiments, it could be

a method that narrows the difference between CT and MR images and further improves the model prediction.

7.3 Reflections

Overall, the results obtained from this project are satisfactory. It is the most challenging project that I have worked on so far, as it requires knowledge of imaging processing and medical image and the ability to put them into practice. While extracting the annotation files from the labels, the image orientation and image storing orientation are two concepts that truly confused me. As extracting these annotates is the fundamental step for further evaluation, I spend more time fully understanding these concepts and extracting the correct bounding box location. At the beginning of the project, the time taking to run the model was underestimated. As the setup for the Caffe framework was complicated, the environment configuration took me a week to complete, which resulted in the lagging of the re-implementation task. However, frequent discussion with my supervisor is the main reason I can correctly fulfill my tasks on time. Throughout the experiments, I found that continuously testing the model is the only way to verify the effectiveness of the modified version, which is different from an application-based project.

The development of this project was a valuable learning experience. Although stressful at times, it ended up being extremely rewarding. I have learned new technologies and develop a deeper understanding of medical imaging and image processing. Also, skills like critical thinking, model designing, and evaluation, literature reviewing, report writing, and project management skills are fully improved.

8 Conclusion

In this project, an automatic method based on 3D Region Proposal Network for multiple organs localization in 3D medical image are implemented. This method is implemented in a 3D manner, which can perform accurate organ localization by fully extracting the spatial information in input images. Benefiting from the multi-candidate bounding box fusion strategy, the target organs could be located within one prediction. This multi-modality model trained on a CT dataset can locate multiple organs in both MR and CT images with high detection precision and localization accuracy. Through the experiment, it shows the potential of the multi-modality model and brightens the future that using a single model to detect organs in many imaging protocols. The Sobel filter and a novel multi-resolution local image normalization method are applied to this model as a preprocessing techniques to further increase the model's localization accuracy in CT and MR images. As shown from the experiment results, compared with the original model, the model, using Sobel edge enhanced image as input, achieves higher localization accuracy in CT images but lower in MR images, while the model with multi-resolution edge enhancement can bring higher detection and localization accuracy of some organs in both CT and MR images, but have lower global detection rate. Overall, the multi-resolution image normalization method could be a technique that further boosts the organ localization accuracy in MR images while maintaining high accuracy in CT images to better serve the multi-modality model. Further experiments on this novel preprocessing technique are required.

9 Bibliography

[1] X. Zhou *et al.*, 'Automatic localization of solid organs on 3D CT images by a collaborative majority voting decision based on ensemble learning', *Comput. Med. Imaging Graph.*, vol. 36, no. 4, pp. 304–313, Jun. 2012, doi: 10.1016/j.compmedimag.2011.12.004.

[2] K. Keraudren, V. Kyriakopoulou, M. Rutherford, J. V. Hajnal, and D. Rueckert, 'Localisation of the Brain in Fetal MRI Using Bundled SIFT Features', in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, Berlin, Heidelberg, 2013, pp. 582–589, doi: 10.1007/978-3-642-40811-3_73.

[3] Y. Zheng, B. Georgescu, and D. Comaniciu, 'Marginal Space Learning for Efficient Detection of 2D/3D Anatomical Structures in Medical Images', in *Information Processing in Medical Imaging*, Berlin, Heidelberg, 2009, pp. 411–422, doi: 10.1007/978-3-642-02498-6_34.

[4] A. Criminisi *et al.*, 'Regression forests for efficient anatomy detection and localization in computed tomography scans', *Med. Image Anal.*, vol. 17, no. 8, pp. 1293–1303, Dec. 2013, doi: 10.1016/j.media.2013.01.001.

[5] A. Criminisi *et al.*, 'Anatomy Detection and Localization in 3D Medical Images', in *Decision Forests for Computer Vision and Medical Image Analysis*, A. Criminisi and J. Shotton, Eds. London: Springer, 2013, pp. 193–209.

[6] R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch, 'Multi-organ localization with cascaded global-to-local regression and shape prior', *Med. Image Anal.*, vol. 23, no. 1, pp. 70–83, Jul. 2015, doi: 10.1016/j.media.2015.04.007.

[7] G. E. Humpire-Mamani, A. A. A. Setio, B. van Ginneken, and C. Jacobs, 'Efficient organ localization using multi-label convolutional neural networks in thorax-abdomen CT scans', *Phys. Med. Biol.*, vol. 63, no. 8, p. 085003, Apr. 2018, doi: 10.1088/1361-6560/aab4b3.

[8] B. D. de Vos, J. M. Wolterink, P. A. de Jong, T. Leiner, M. A. Viergever, and I. Išgum, 'ConvNet-Based Localization of Anatomical Structures in 3D Medical Images', *IEEE Trans. Med. Imaging*, vol. 36, no. 7, pp. 1470–1481, Jul. 2017, doi: 10.1109/TMI.2017.2673121.

[9] X. Xu, F. Zhou, B. Liu, D. Fu, and X. Bai, 'Efficient Multiple Organ Localization in CT Image Using 3D Region Proposal Network', *IEEE Trans. Med. Imaging*, vol. 38, no. 8, pp. 1885–1898, Aug. 2019, doi: 10.1109/TMI.2019.2894854.

[10] X. Xu, F. Zhou, B. Liu, and X. Bai, 'Multiple Organ Localization in CT Image Using Triple-Branch Fully Convolutional Networks', *IEEE Access*, vol. 7, pp. 98083–98093, 2019, doi: 10.1109/ACCESS.2019.2930417.

[11] S. Ioffe and C. Szegedy, 'Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift', in *International Conference on Machine Learning*, Jun. 2015, pp. 448–456, Accessed: Nov. 05, 2020. [Online]. Available: http://proceedings.mlr.press/v37/ioffe15.html.

[12] S. Ren, K. He, R. Girshick, and J. Sun, 'Faster R-CNN: Towards real-time object detection with region proposal networks,' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[13] I. Sobel. 'An Isotropic 3x3x3 Volume Gradient Operator', in *Technical report* in *Hewlett-Packard Laboratories*, Apr. 1995.

[14] "CodaLab - Competition", 2017. Accessed on: Nov. 20, 2020 [Online]. Available: http://www.lits-challenge.com.

[15] i. Sage Bionetworks, "Synapse | Sage Bionetworks", 2015. Accessed on: Nov. 25, 2020 [Online]. Available: https://www.synapse.org/#!Synapse:syn3193805/wiki/217789.

[16] "CHAOS – Grand Challenge", 2019. Accessed on: Feb. 16, 2021 [Online]. Available at: https://chaos.grand-challenge.org/

[17] "Y. Jia et al., 'Caffe: Convolutional architecture for fast feature embedding', in *Proc. 22nd*

ACM Int. Conf. Multimedia. New York, NY, USA: ACM, 2014, pp.675-678.
[18] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, 'Gradient-based learning applied to document recognition', in *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.
[19] X. Glorot and Y. Bengio, 'Understanding the difficulty of training deep feedforward neural networks', in *Proc. 13th Int. Conf. Artif. Intell. Statist*., 2010, pp.249-256.
[20] "Trello", Accessed on: Oct. 03, 2020 [Online]. Available: https://trello.com.

10 Appendix

10.1 Abbreviations

- CT = Computed Tomography
- MR = Magnetic Resonance
- MRI = Magnetic Resonance Images
- RoI = Regions of Interests
- MSER = Maximally Stable Extremal Regions
- RANSAC = Random Sample Consensus Procedure
- MSL = Marginal Space Learning
- ConvNet = Convolutional Neural Network
- RPN = Region Proposal Network
- Triple-Branch FCN = Triple-Branch Fully Convolutional Network
- ReLU = Rectified Linear Units
- IoU = Intersection-over-Union
- LiTS = Liver Tumor Segmentation Challenge
- BTCV = Multi-Atlas Labelling Beyond the Cranial Vault Segmentation Challenge
- CHAO = Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge
- ITK = Insight ToolKit
- NIfTI = Neuroimaging Informatics Technology Initiative
- DICOM = Digital Imaging and Communications in Medicines