Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

Fraunhofer IAIS

Master Thesis Proposal

Object Detection in Volume Data

Ramit Sharma

Supervised by

First Supervisor
Second Supervisor
Laura Anger

December 2020

1 Introduction

CT scans are not limited to medical domain; they are also used in industries. In industries, they are used for detection of flaws like cracks and voids as well as particle analysis of materials. They are used in metrology for the measurement of internal and external geometry of complex parts. According to the iData Research's medical imaging procedures analysis, over 75 million CT scans are performed each year in the United States alone. This number is forecasted to reach 84 million procedures by 2022 [11]. Analysis of CT scans for the diagnosis of the disease is a tedious task and requires a lot of human effort and working hours, and a small human error in the diagnosis could put the patient's life to risk. So to minimize this risk, a lot of research is being done to perform automatic as well as semi-automatic diagnosis of CT scans. In platforms like Kaggle, we can find competitions like RSNA pneumonia detection challenge [20], COVID 19 CT scans [12] where they provide labeled data to solve the problem of automatic diagnosis of CT scans. The datasets like DeepLesion dataset [3], covid-19-chest-xray-lung-bounding-boxes-dataset [4] have been provided by the medical institutes to openly involve people to develop systems to perform object detection in CT scans.

One of the factors that impact the performance of object detection models is the resolution of images [17]. Low-resolution CT scans have several advantages which are discussed in the related work section 2.1. Hence in this project, we intend to find out which model performs better even at lower resolution. In order to find this out, we would survey the various object detection models. 3D object detection models are computationally expensive [26]. Hence we intend to look into the 2D object detection models. The 2D object detection models could be broadly classified into two categories, namely, the one-stage approach and the two-stage approach.

So we would implement two models, the first implementation would belong to the category of one-stage approach, and the second implementation would belong to the category of two-stage approach. We would evaluate the performance of both the models at various resolution. We would also look into the frames per second attribute of the models at different resolution and analyse how the frames per second the model can predict changes when the resolution is varied.

2 Related Work

In this section we would discuss about the advantages of using low-resolution images 2.1. Then we would discuss about the 3D object detection models 2.2 and 2D object detection models 2.3. In the end of this section we would discuss about the major findings from this section.

2.1 Advantages of Low resolution images

The advantages of using Low-resolution images are as follows:

Low memory requirement: The memory required to store the images reduces when we use the image of lower resolution; this is well illustrated in figure 1.

- **Memory requirement**

Inch size (changed)	Resolution (changed)	Pixel dimensions (you set)	File size
2 x 2 in	100 ppi	200 x 200 px	117.2 KB
3 x 3 in	100 ppi	300 x 300 px	263.7 KB
6 x 6 in	100 ppi	600 x 600 px	1.03 MB

Figure 1: memory requirements [1]

- **Frame Per Second(Fps) increases**: As the resolution of the image decreases, the fps of the object detection model increases [25].

The problem of object detection could be classified into two categories:

- **3D object detection**
- **2D object detection**

2.2 3D object detection

A significant improvement has been obtained in the field of 3D object detection because of Convolutional Neural Networks (CNNs). Most of the previous works used to convert the point cloud to volumetric representations and CNNs were generalized to 3D CNNs for the task of object detection. 3D FCN [14] uses 3D CNNs to predict 3D bounding boxes as well as class labels. VoxelNet [27] uses 3D CNNs to encode 3D input volumes to 2D feature maps, and theses features are fed to subsequent detection network. In Vote3Deep [7], the sparsity of 3D volume is utilized to accelerate the 3D convolution. A major drawback of these 3D algorithms is that they are computationally expensive [26].

2.3 2D object detection

The problem of 2D object detection could be divided into two categories:

- **One stage approach**: Unified one stage approach refers to architectures which directly predict class probabilities as well as bounding box offsets from images with single feed-forward Convolutional Neural Network(CNN) in a monolithic setting which does not involve generation of proposal region or post classification that encapsulates all computation using a single network.

  YOLO [21] divides the input image into M x M gird cells and utilizes CNNs to get the bounding box regression, confidence scores as well as class probabilities of each grid cell. YOLO0000 [22] and YOLOv3 [13] further improve the performance. Even though YOLO is fast, it misses small objects because of the coarse segmentation of input images. These drawbacks were addressed by SSD [16] by utilizing feature pyramids for single stage object detection. In SSD for every feature map locations anchor boxes of various aspect ratios and scales are generated. In RetinaNet [15] they proposed focal loss in order to handle the imbalance between target and background object bounding boxes.

- **Two stage approach** : Two-stage approaches are region-based frameworks. In the case of two-stage approach region proposals which are category independent are generated from an image. CNN features are then extracted from these regions. After that category specific classifiers are utilized to determine the label of the categories for the proposals.

The two-stage object detection algorithms are best represented by the R-CNN family [8, 9, 23]. Faster R-CNN introduced the Region Proposal Network (RPN). A substantial number of background candidates are filtered out by RPN, and a different network is used to predict bounding box co-ordinates and class labels for each proposal. In R-FCN [5] position-sensitive feature maps are extracted. These feature maps are fed to RPN to get class scores. Mask R-CNN [10] extends Faster R-CNN to instance segmentation, they first find the bounding box coordinates and crop and segment the bounding box region to get the refined mask.

As discussed in section 2.1, reducing the resolution of the image has advantages like reduction in memory requirement, increasing the fps of the model. But the papers discussed in section 2.3 don't provide the details about how these models would perform when the resolution is varied.

2.4 Problem Statement

Today the use of CT scans are not limited to the medical domain, they are used in industries for finding defects in materials, in airport baggage security, and many other applications. The object detection system for these cases should be fast. Suppose we have an object detection system for airport baggage screening, then the object detection system should be fast enough to detect objects in the scans, in other words, the frames per second (fps) of the object detection system should be high. One of the approaches to make the fps high is to use an image of lower resolution [2]. Hence in this paper, we intend to look into object detection systems that are able to detect objects even at a lower resolution and have a higher fps as well as accuracy. In figure 2, we can see that the image resolution of different volume data is different. MRI is of lower resolution, whereas digital radiography, digital mammography and computed radiography have higher resolution. Hence in this paper, we intend to look into the object detection systems that have higher fps and accuracy in both lower as well as higher resolution volume data.

Modality	Image matrix (in pixels)	Dynamic range (bits per pixel)	File size (per image)
MRI	256 × 256	16	131 KB
CT Scan	512 × 512	16	524 KB
Ultrasound	512 × 512	8	262 KB
Color Doppler	768 × 576	8	442 KB
Digital radiography	Up to 3000 × 3000	Up to 16	Up to 18 MB
Digital mammography	Up to 3328 × 4096	14	27 MB
Computed radiography	3520 × 4280	12	30 MB

Table modified from reference 8

Figure 2: Resolution of images of different volume data [6]

The radiation we get from CT, nuclear imaging is ionizing radiation. This radiation could damage the DNA and even could lead to cancer in the long run [24].

In an attempt to reduce the radiation dose, the exposure time of the patient could be reduced, but doing so will increase the noise and decrease the low contrast resolution of image [18]. 3D baggage-CT imagery typically presents with substantial noise, metal-streaking artefacts and poor voxel resolution and is thus generally of poorer quality than medical-CT imagery [19]. Hence if the time permits, we would also add noise signals to the image data and find which object detection model performs better in noisy data.

3 Project Plan

- To carry out the survey of the various object detection models
- To carry out the survey of various CT scan datasets for object detection
- To select one "one-stage object detector" and one "two-stage object detector" for object detection and select the corresponding dataset.
- To prepare various datasets at different resolutions
- To implement both the selected one stage as well as two-stage object detector models
- To compare the performance of the models at different resolutions
- To analyse the impact of resolution in accuracy of the model and frames per second the model can predict
- If the time permits, we would also try to publish a paper in a journal

3.1 Expected Goals

3.1.1 Minimum

- To survey the various CT scan datasets available for object detection
- To survey the various object detection models for CT scans
- To select two approaches, one that belongs to the single-stage object detector category and the other that belongs to two-stage object detector category
- To implement the selected two-stage object detector model

3.1.2 Expected

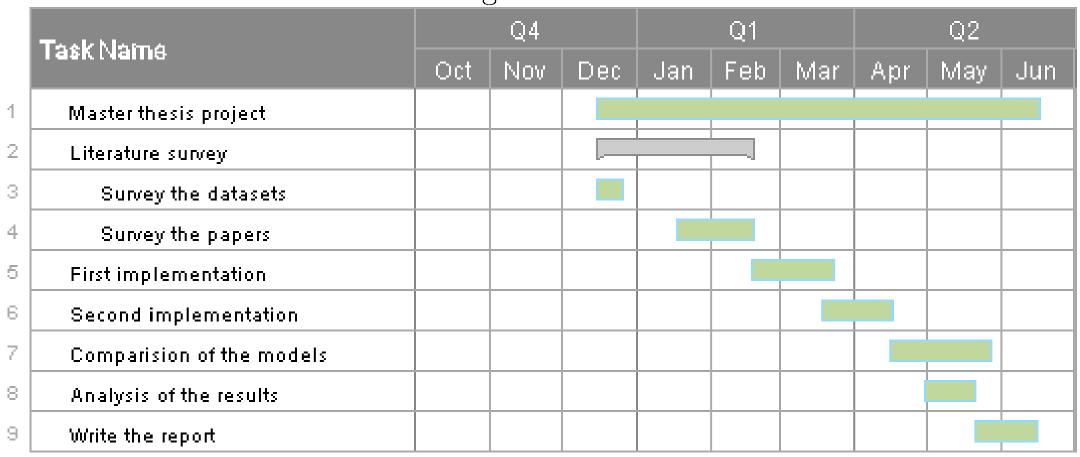
- To implement the selected one stage object detector model
- To compare the performance of both models at different resolution and frames per second the model can predict at different resolution
- To analyse the impact of resolution on the performance of models and frames per second the model can predict
- To select the model which performs the best even at low resolution

3.1.3 Maximum

- To publish a paper in one of the journals

3.2 Project Schedule

Figure 3:



References

[1] Adobe. https://helpx.adobe.com/photoshop/kb/advanced-cropping-resizing-resampling-photoshop.html, . Accessed on 22 October 2020.

[2] Adobe. https://towardsdatascience.com/no-gpu-for-your-production-server-a20616bb04bd, . Accessed on 22 October 2020.

[3] NIH Clinical Center. https://nihcc.app.box.com/v/DeepLesion/folder/50715173939. Accessed on 22 October 2020.

[4] covid-19-chest-xray-lung-bounding-boxes dataset. https://github.com/GeneralBlockchain/covid-19-chest-xray-lung-bounding-boxes-dataset. Accessed on 22 October 2020.

[5] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. In *Advances in neural information processing systems, pages 379–387*, 2016.

[6] Ravi Varma Dandu. Storage media for computers in radiology. In *COMPUTERS IN RADIOLOGY*, 2018.

[7] M. Engelcke, D. Rao, D. Z. Wang, C. H. Tong, and I. Posner. Vote3deep: Fast object detection in 3d point clouds using efficient convolutional neural networks. 2017.

[8] Ross Girshick. Fast r-cnn. In *In Proceedings of the IEEE international conference on computer vision, pages 1440–1448*, 2015.

[9] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In *Facebook AI Research (FAIR)*, 2018.

[10] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In *arxiv.org*, 2018.

[11] iData Research. https://idataresearch.com/over-75-million-ct-scans-are-performed-each-year-and-growing-despite-radiation Accessed on 22 October 2020.

[12] Kaggle. https://www.kaggle.com/andrewmvd/covid19-ct-scans. Accessed on 22 October 2020.

[13] Andréanne Lemay. Kidney recognition in ct using yolov3. In *In Advances in neural information processing systems, pages 91–99*, 2019.

[14] Bo Li. 3d fully convolutional network for vehicle detection in point cloud. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.

[15] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Doll´ar. Focal loss for dense object detection. In *In Proceedings of the IEEE international conference on computer vision, pages 2980–2988*, 2017.

[16] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision, pages 21–37. Springer*, 2016.

[17] Jerubbaal John Luke, Rajkumar Joseph, and Mahesh Balaji. Impact of image size on accuracy and generalization of convolutional neural networks. In *IJRAR*, 2019.

[18] Michael F. McNitt-Gray. Tradeoffs in ct image quality and dose. 2006.

[19] Andre Mouton and Toby P. Breckon. A review of automated image understanding within 3d baggage computed tomography security screening. In *Researchgate*, 2015.

[20] Radiological Society of North America. https://www.kaggle.com/c/rsna-pneumonia-detection-challenge. Accessed on 22 October 2020.

[21] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788*, 2016.

[22] ´J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In *In CVPR*, 2017.

[23] Shaoqing Ren, Ross Girshick Kaiming He, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *In Advances in neural information processing systems, pages 91–99*, 2015.

[24] Harvand Medical School. `https://www.radiologyinfo.org/en/info.cfm?pg=safety-xray`. Accessed on 22 October 2020.

[25] towards data science. `https://towardsdatascience.com/no-gpu-for-your-production-server-a20616bb04bd`. Accessed on 22 October 2020.

[26] Anqi Yang, Aswin Sankaranarayanan, Srinivasa Narasimhan, David Held, and Jen-Hao Chang. 3d object detection from ct scans using a slice-and-fuse approach. In *Carnegie Mellon University*, 2019.

[27] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.