Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Bonn-Aachen
International Center for
Information Technology

Fraunhofer
IAIS

Master Thesis Proposal

Object Detection in Dense Volume Data

Ramit Sharma

Supervised by

Prof. Dr. Paul G. Plöger
Second Supervisor
Laura Anger

December 2020

1 Introduction

Object detection has attracted a lot of attention in recent years because it has a wide range of applications and technological breakthroughs. The task of object detection has been explored extensively in academia as well as real-world applications like monitoring security, transportation surveillance, autonomous driving, drone scene analysis, robot perception etc. [16] In this project, our work is focused on object detection in dense volume data. One of the types of dense volume data is computerized tomography (CT) scan.

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 [15]. 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 [33], COVID 19 CT scans [17] where they provide labeled data to solve the problem of automatic diagnosis of CT scans. The datasets like DeepLesion dataset [2], covid-19-chest-xray-lung-bounding-boxes-dataset [5] 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 both the accuracy as well as the speed of the object detection models is the resolution of the image.[28] Hence in this project, we intend to look deeper into the impact of resolution on the overall performance of the model.

1.1 Problem statement

The problem statement section is divided into two sections. In section 1.1.1 we discuss about the problem, and in section 1.1.2 we discuss about the relevance of the problem.

1.1.1 The problem

The approaches to object detection could be classified into one stage approach and two-stage approach. The one-stage approaches are fast but are less accurate. The two-stage approaches are computationally expensive and cannot perform in real-time.[51] One of the approaches to address this issue of complexity is to use low-resolution image.[46]

But the papers [4, 10, 13, 21–23, 25, 27, 34, 35, 37, 38, 40–42, 44, 50, 52, 53, 56] don't provide details about how the model would perform when the resolution is varied. Hence in this project, we intend to look deeper into the impact of resolution in the accuracy as well as the speed of the model. We intend to find out which model has better speed as well as accuracy when the resolution is varied. To do so, we would do the literature survey of the various object detection models. We would select two state-of-the-art models for object detection and compare their performance at various resolution.

1.1.2 Relevance of the problem

Our project focuses on dense volume data. One of the types of dense volume data is CT scan. CT scans are used in airport baggage screening, assembly lines, and the object detection systems in these places should be able to detect objects in real-time. Low-resolution CT scanning is fast. Using low-resolution image is one of the ways to address the issue of computational complexity and making the object detection systems fast. Hence using low-resolution image is helpful in making the entire process of scanning as well as detection faster. Even in the medical domain, to reduce the radiation dose, the exposure time of the patient should be reduced.[29] The exposure time of patients could be reduced by allowing low-resolution CT scans. So, finding out which object detection model has better accuracy as well as speed at low-resolution CT scans is quite relevant.

2 Related work

In this section, we first give an overview of object detection 2.1. Then we discuss about the approaches of object detection, namely one stage object detection 2.2 and two-stage object detection 2.3.In subsection 2.4 we discuss about the 3D data used in object detection . In subsection 2.5 we give a brief introduction about resolution and discuss the advantages as well as disadvantage of using low-resolution image. In subsection 2.6 we discuss about the models.

2.1 Object detection overview

The crux of object detection is to locate as well as classify objects. It utilises rectangular bounding boxes in order to locate the objects that are detected and then classify the category of object. Object detection is one of the important areas of computer vision. It has several applications in scientific as well as practical industrial productions like face detection, text detection, pedestrian detection, video detection, logo detection, vehicle detection, medical image detection and so forth. The current state of the art object detection models use deep convolutional neural networks(CNNs) as their backbone as well as detection network for extracting features from input images or videos and solve the task of classification as well as localization respectively. [16]

The approaches to the task of object detection can be divided into two categories:

- One stage approach
- Two-stage approach

2.2 One stage approach

In one stage approach class probabilities as well as bounding boxes of objects are predicted utilising a single-stage network. They don't require region proposal generation as well as post processing. So, single stage approaches are fast. [11]

The one stage approach could be divided into two tasks :

- 2D object detection
- 3D object detection

2.2.1 2D object detection

Unified one stage 2D object detection 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 [37] 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 [38] and YOLOv3 [22] 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 [27] 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 [25] they proposed focal loss in order to handle the imbalance between target and background object bounding boxes.

2.2.2 3D object detection

The single-stage 3D object detection approaches [21, 23, 44, 50, 52, 56] parse the given sparse 3D point cloud to a compact representation like voxel grid or bird eye view image and use CNN to predict the bounding box. This enables single-stage approaches to become simple as well as efficient. A significant drawback of this approach is that these approaches downscale the feature maps progressively. Hence, the spatial resolution of the feature maps get lost and thus, the structural information of point cloud could not be considered explicitly. Therefore single-stage approaches are less accurate when it comes to processing the sparse point clouds.[12]

In [22, 51], they slice the dense 3D data to 2D slices and are fed to 2D object detection models to get the prediction. In [18], Khosravan et al. proposed S4ND, a deep learning method which does lung nodule detection in one step. The main architecture is based on convolution blocks that has dense connections. They also use down-sampling methods in the network as it plays an important role in tiny object detection. In [20] they used single stage network which is implemented as 3D CNN, but it did not perform well as it was not able to converge. Hence they used multistage detector instead. Unfortunately they didn't provide much details about the architecture of the one stage detector used.

2.3 Two-stage approach

In two stage approach several possible regions containing objects are proposed and then region-wise features to predict the category of each region or proposal are extracted. [11]

The two-stage approach could be divided into two tasks :

- 2D object detection
- 3D object detection

2.3.1 2D object detection

Two-stage 2D object detection 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 2D object detection algorithms are best represented by the R-CNN family [10, 13, 40]. 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 [6] position-sensitive feature maps are extracted. These feature maps are fed to RPN to get class scores. Mask R-CNN [14] 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.

2.3.2 3D object detection

Two-stage 3D object detection approach [4, 34, 35, 41, 42, 53] leverage spatial information in the second stage, that focuses on the region of interest (ROI's) which are predicted by the first stage and then predicts bounding box. Compared to one stage approach two stage approach are better as they leverage the regions of interest in first stage and focus only on those regions in second stage. This shows that accurate localization can be achieved when fine-grained spatial information is leveraged. Operating on each point and re-extracting features on each ROI increases the computational cost substantially. Hence it becomes hard for two-stage approaches to reach real-time speed. [12]

In two-stage approaches [19, 57], the dense volume data is fed to the 3D version of R-CNN models to get the prediction. In [47] slice of dense volume data, i.e. slice of 3D CT scans are fed to Mask R-CNN to get the prediction. In [54] Zhang et al. introduce pancreatic tumor detection framework which aims to fully exploit the context information at different scales. The network uses Feature Pyramid network [24] along with Faster R-CNN [39] in the backbone.

2.4 Datasets

Datasets have played a very important role in the history of object detection research. Datasets are not only common ground to measure and compare the performance of algorithms, but also pushed the field towards improving the system so that complex and challenging problems of object detection could be solved. In recent times deep learning techniques have been very successful in several visual recognition problems, and it could not have been possible without the availability of large amounts of annotated data.[26] The datasets for 3D object detection can be broadly classified into two categories :

- Sparse 3D point cloud
- Dense volume data

2.4.1 Sparse 3D point cloud

The sparse 3D point clouds could be divided into two categories namely RGB-D images and LiDAR data. In comparison to RGB-D images , LiDAR data is special. On one hand LiDAR data provides structural and spatial information of relative location and precised depth. On the other hand LiDAR data is sparse, unordered and locality sensitive, and hence it becomes more difficult to process raw LiDAR data.[53] Both data types are discussed in following sections.

2.4.1.1 LiDAR data

Light Detection and Ranging (LiDAR) is a remote sensing technique in which the distance to the target is gauged by illuminating the target utilizing a laser light and using a sensor in order to measure the reflection of light [36]. The difference in wavelengths and laser return times are then used to obtain a 3D representation of the target. LiDAR data has several applications in surveying, geomatics, geography, archaeology, seismology, forestry. The KITTI Vision Benchmark Suite [9] is one the datasets for 3d object detection in the field of autonomous driving. It contains LIDAR data taken using a sensor mounted in the front of the car.

2.4.1.2 RGB-D data

RGB-D data is a combination of RGB data along with its corresponding depth data. One of the RGB-D object detection datasets is SUN RGB-D [45].

2.4.2 Dense volume data

Dense volume data is also referred to as dense 3D data in [7]. Computed Tomography (CT), Magnetic Resonance Imaging (MRI) are some of the examples of dense volume data.

2.4.2.1 Computed Tomography (CT)

The term "Computed Tomography" refers to computerized X-ray imaging process in which narrow beam of X-rays are aimed at the patient and rotated quickly around the body which produces signals that are utilized by machine's computer to get cross-sectional images or slices of the body. These slices are referred to as tomographic images and contains more detailed information in comparison to conventional x-rays. [31] The principle remains the same when CT scans are used for other objects as well. 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. Some of the CT scan object detection datasets available are RSNA pneumonia detection challenge [33], COVID 19 CT scans [17], DeepLesion dataset [2], covid-19-chest-xray-lung-bounding-boxes-dataset [5]

2.4.2.2 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging(MRI) is a non-invasive imaging technique that creates three-dimensional anatomical images. It is used for diagnosis and detection of diseases as well as treatment monitoring. It relies on sophisticated technology which excites as well as detects the change in the direction of the rotational axis of protons that are found in water which makes up living tissues. MRI scanners are well suited to get the image of the non-bony parts as well as soft tissues of the body. They are different from CT scans in the way that they don't use damaging ionizing radiation of the x-rays. The spinal cord, brain, nerves, ligaments, muscles, tendons are visible more clearly with MRI in comparison to regular x-rays and CT, this is the reason why MRI is used for imaging the knee as well as shoulder injuries. [32]

In [3, 55], they use MRI data as an input for the task of object detection.

2.5 Resolution of image

Resolution means the number of pixels present in an image. Resolution is identified by the width and height of the image. Resolution also refers to the total pixels an image contains, for example, if an image is 2048 pixels wide and 1536 pixels high. Total pixels = 2048*1536 = 3145728 pixels or 3.1 Megapixels So the resolution image is 3.1 Megapixels.[30] .One of the factors that impact the performance of object detection models is the resolution of images [28].

2.5.1 Advantage of using low resolution image

The advantages of using low resolution images are:

- Low memory requirement
- Frame Per Second(Fps) increases

The advantages are explained in the following sections.

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.

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 requirement at various resolutions [1]

Frame Per Second(Fps) increases As the resolution of the image decreases, the fps of the object detection model increases [48].

2.5.2 Disadvantage of using low-resolution image

Downsampling image leads to information loss and affects the accuracy of the model [43].

2.6 Discussion

3D data could be processed using 3D object detection models, or they could be sliced and fed to 2D object detection techniques for prediction. The 3D object detection technique, whether it is one stage approach or two-stage approach has these limitations:

- A major drawback of these 3D algorithms is that they are computationally expensive [51].
- 3D data are quite difficult to collect, annotate and store and hence there are fewer datasets for deep learning models. [8, 49]

In [46] to address the issue of more memory requirement and higher computational complexity Song et al. used low-resolution image as input for the Region Proposal Network(RPN). So, low-resolution image could be used to solve the problem of computational complexity and high memory requirement of 2D as well as 3D object detection networks. Reduction in resolution would also lead to reduction in accuracy. But the papers discussed in the sections one stage approach 2.2 and two-stage approach 2.3 don't provide the details about the impact of resolution on accuracy and speed of the models. Hence in this project we intend to look into impact of resolution on accuracy and speed of the models.

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

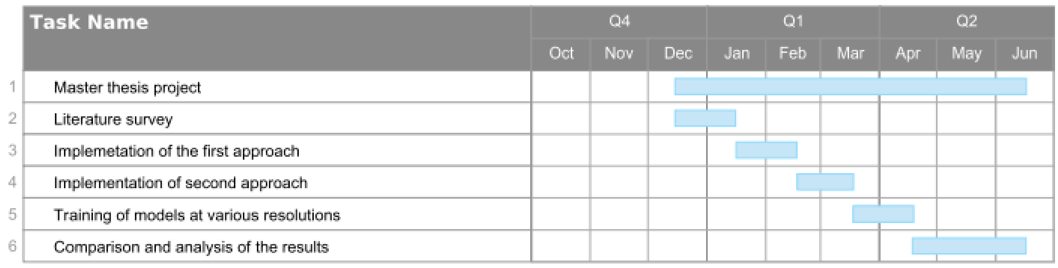
3.1.2 Expected

- 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

4 Project Schedule



References

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

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

[3] Hao Chen, Qi Dou, Lequan Yu, and Pheng-Ann Heng. Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation. In *arXiv:1608.05895v1 [cs.CV]*, 2016.

[4] Yilun Chen, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Fast point r-cnn. 2019.

[5] 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.

[6] 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.

[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] Di Feng, Xiao Wei, Lars Rosenbaum, Atsuto Maki, and Klaus Dietmayer. Deep active learning for efficient training of a lidar 3d object detector. In *arXiv*, 2019.

[9] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

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

[11] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. In *arXiv:1912.12033v2 [cs.CV]*, 2020.

[12] Chenhang He, Hui Zeng, Jianqiang Huang, Xian-Sheng Hua, and Lei Zhang. Structure aware single-stage 3d object detection from point cloud. In $CVPR_2$020, 2020.

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

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

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

[16] Licheng Jiao, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng, and Rong Qu. A survey of deep learning-based object detection. In *arXiv:1907.09408v2 [cs.CV]*, 2019.

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

[18] Naji Khosravan and Ulas Bagci. S4nd: Single-shot single-scale lung nodule detection. In *arXiv:1805.02279v2 [cs.CV]*, 2018.

[19] Evi Kopelowitz1 and Guy Englehard. Lung nodules detection and segmentation using 3d mask-rcnn. In *arXiv*, 2019.

[20] Kingsley Kuan, Mathieu Ravaut, Gaurav Manek, Huiling Chen, Jie Lin, Babar Nazir, Cen Chen, Tse Chiang Howe, Zeng Zeng†, and Vijay Chandrasekhar†. Deep learning for lung cancer detection: Tackling the kaggle data science bowl 2017 challenge. In *IEEE*, 2017.

[21] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *CVPR*, 2019.

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

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

[24] Tsung-Yi Lin, Piotr Doll´ar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *In Advances in neural information processing systems, pages 91–99*, 2015.

[25] 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.

[26] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep learning for generic object detection: A survey. In *International Journal of Computer Vision (2020) 128:261–318*, 2019.

[27] 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.

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

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

[30] microscope.org. `https://microscope-microscope.org/microscope-info/image-resolution/`. Accessed on 22 October 2020.

[31] NIH. `https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct`, . Accessed on 22 October 2020.

[32] NIH. https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri, . Accessed on 22 October 2020.

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

[34] Charles R Qi, Wei Liu, Chenxia Wu, Hao Su, , and Leonidas J Guibas. Frustum pointnets for 3d object detection from rgbd data. 2018.

[35] Charles R. Qi, Or Litany, Kaiming He, and Leonidas J. Guibas. Deep hough voting for 3d object detection in point clouds. 2019.

[36] AZO quantum. https://www.azoquantum.com/News.aspx?newsID=7018. Accessed on 22 October 2020.

[37] 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.

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

[39] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. 2015.

[40] 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.

[41] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointrcnn: 3d object proposal generation and detection from point cloud. In *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–779*, 2019.

[42] Shaoshuai Shi, Zhe Wang, Xiaogang Wang, and Hongsheng Li. Part-a2net: 3d part-aware and aggregation neural network for object detection from point cloud. 2019.

[43] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. In *Springer*, 2019.

[44] M. Simony, S. Milzy, K. Amendey, and H.-M. Gross. Complex-yolo: an euler-region-proposal for real-time 3d object detection on point clouds. In *In Proceedings of the European Conference on Computer Vision (ECCV), pages 0–0*, 2018.

[45] S. Song, S. Lichtenberg, and J. Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of 28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015)*, 2015.

[46] Shuran Song and Jianxiong Xiao. Deep sliding shapes for amodal 3d object detection in rgb-d images. 2016.

[47] Tang, Y., Yan, K., Tang, Y., Liu, J., Xiao, J., Summers, and R.M. Uldor: A universal lesion detector for ct scans with pseudo masks and hard negative example mining. In *arXiv:1901.06359*, 2019.

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

[49] Qian Wang, Neelanjan Bhowmik, and Toby P. Breckon. Multi-class 3d object detection within volumetric 3d computed tomography baggage security screening imagery. In *arXiv*, 2020.

[50] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. In *MDPI*, 2018.

[51] 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.

[52] Bin Yang, Wenjie Luo, and Raquel Urtasun. Pixor: Realtime 3d object detection from point clouds. In *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 7652–7660*, 2018.

[53] Zetong Yang, Yanan Sun, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Std: sparse-to-dense 3d object detector for point cloud. In *ICCV*, 2019.

[54] Zhengdong Zhang, Shuai Li, Ziyang Wang, and Yun Lu. A novel and efficient tumor detection framework for pancreatic cancer via ct images. In *arXiv*, 2020.

[55] B. Zhao, J. Soraghan, G. Di-caterina, L. Petropoulakis, D. Grose, and T. Doshi. Automatic 3d segmentation of mri data for detection of head and neck cancerous lymph nodes. In *2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pages 298–303, 2018. doi: 10.23919/SPA.2018.8563420.

[56] 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.

[57] Wentao Zhu, Chaochun Liu, Wei Fan, and Xiaohui Xie1. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In *arXiv:1801.09555v1 [cs.CV]*, 2018.