Optimizing CT scan slice count through Lesion detection using YOLO

Abhijith Tammanagari
Duke University
Durham, NC 27708
at396@duke.edu

Satvik Kishore
Duke University
Durham, NC, 27708
sk741@duke.edu

Abstract

Computer Tomography(CT) scans are often used to identify lesions of different types in bone, liver, lung, kidney, and soft tissues. CT scanning machines have various specifications with one distinguishing factor being the number of slices in the receiver. This slice count is a driving factor in CT machine prices. We want to find out what is the minimum number of slices in a CT scanner that retains the ability to identify lesions. Reducing the number of slices brings down cost of operation while also reducing the radiation exposure to the patients. We used sinogram reconstruction to simulate the various slice counts of 16,64, and 128. We then trained an object detection model on that data-set and determined that a 64 slice CT scanner, in terms of mean-average precision, has similar performance to that of a 128 slice CT scanner.

1 Introduction

CT scans are the primary source of identifying lesions within tissues and organs. CT scanning machines have various number of slices in the receiver and the number of these slices range anywhere from 16 to 300 with the common range being within 16 and 128 slices[1]. The more slices there are in the receiver the more clear the constructed image will be. These slice counts have two main complications, one of which is that increasing the number of slices raises the cost of the CT scanner. 16 slice CT scanner are typically priced between $80k - $100k whereas the 128 slice CT scanners are priced at $200k+[2]. Also, it is known that the more slices that are in the CT scanner the more radiation is required which is an inherent risk to the patient [3]. If we simulate the number of slices in the CT scanner via radon and inverse radon transformations, we can find a lower slice count that results in similar performance in terms of identifying lesions as the 128 slice CT scanner. This is important because if the 64 slice or 16 slice CT scanner performs as well as the 128 slice CT scanner then we would be saving the hospital money as well as reducing the radiation load on a patient during a CT scan.

2 Related Work

Heggie et. al. talk about why optimzation of CT scanning technology is important to reduce the radiation dose to the patients. They optimize the physical parameters of the machine in order to improve the amount of information captured within the geometry of the tissue at lower radiation dose levels [3]. Deep learning is opening up new avenues in image recognition tasks and is seeing application optimizing low dose CT scanners as well. It has been used in improving image quality from by denoising the images [4].

We decided to go with a you only look once (YOLO) model primarily because it has been proven to be effective in identifying melanoma which is a type of skin lesion. The paper by Yali Nie

demonstrates that even with a small sample size of 300 images the YOLO architecture accurately identifies melonoma with a mean average precision of .82. [5] This gives us the faith that using a YOLO architecture should be able to identify lesions in CT images.

Yan et. al. mined sources of CT image data collected and annotated by radiologists to create the DeepLesion dataset. They trained neural network models like the VGG-16 to build a universal lesion detector and they were able to do so with an 81% precision.[6]

We know that using a simultaneous algebraic reconstruction technique (SART) is a valid method of reconstructing a CT but we need to validate how effective this method is for a limited number of projections. A paper by Hengyong Yu, suggests that with the SART method including a fewer number of projection that contain a sparsity constraint allows us to effectively construct the CT with increased clarity and reduce the need for more projections.[7] Therefore using an approach that evenly spaces out the CT slices will allow us to maintain a similar amount of information within the image even if fewer slices in total are utilized.

3 Methodology

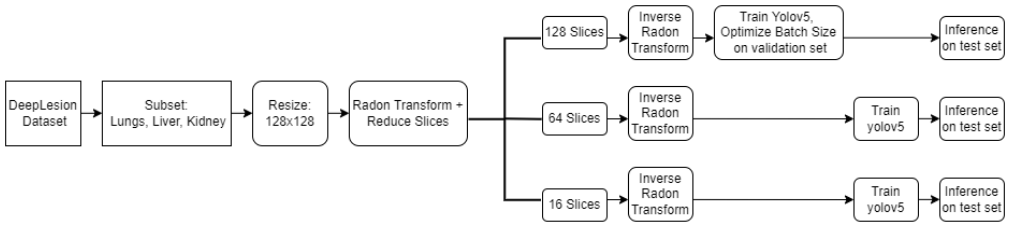


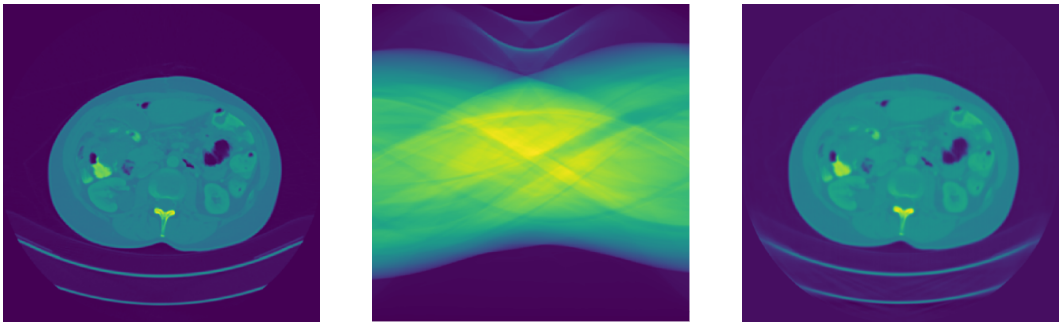
Figure 1: Flowchart representing the experiment design

3.1 Data and Preprocessing

In this study, we used the DeepLesion dataset [6]. This dataset was composed of 32,735 lesions in 32,120 annotated CT slices from 10,594 studies of 4427 unique patients. The annotations defined the bounding boxes of the location and size of the lesions on these selected key slices. The lesions were from multiple types of tissues, and we selected the tissues from the lungs, liver, and kidney. We resized all images to 128x128 pixels and split them into training, validation, and testing subsets.

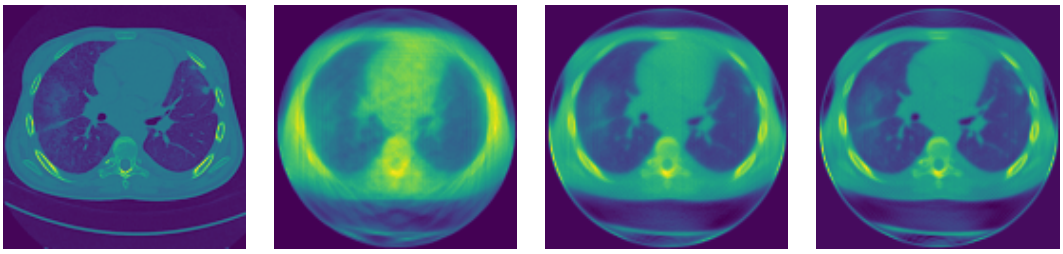
3.2 Physical Layer

We simulated a physical layer by treating the images in the dataset as real subjects. We then applied a radon transformation to create sinogram of the images [8]. The radon transformation involves creating projections of the subject from different angles, each projection representing a shadow that is formed when radiation passes through the subject. These projections are then fourier transformed and placed next to each other to create the sinogram. In our study, we created sinograms from 128 angles spaced evenly around the subject. To simulate the capture of a real image from the subject, we performed an inverse radon transform on the sinogram, using the simulated algebraic reconstruction technique [9]. To simulate inverse radon transformations at differing number of slices, we subsetted the sinogram at the desired number of slices, the slices being evenly spaced each time. Figure 2 illustrates an example image from our data, the corresponding sinogram and the reconstructed image at 128 slices. We observed that there is some blurring and loss of resolution in the reconstructed image and a few reconstruction artefacts being generated in the lower portion of this image. This occurred because the reconstruction is an approximate of the original subject and we expect reconstruction through more slices to generate sharper images. However, the image appeared good enough for us to be able to observe the structures within the tissue without difficulty when compared with the original image. Figure 3 compares the images generated using 16, 64, and 128 slices and we observe that increasing the number of slices improves image quality, however there is not too much difference between the 64 and 128 slice images.



(a) Original image (b) Sinogram (c) Reconstructed Image

Figure 2: Reconstruction of Image using a Sinogram and Radon / Inverse Radon Transformation



(a) Original Image (b) 16 Slices (c) 64 Slices (d) 128 Slices

Figure 3: Comparison of quality of reconstructed images vs original image

3.3 Modeling

Through modeling, we aim to identify the lesions in the reconstructed images and evaluate the performance of the models on the three sets of images in order to determine the best number of slices to use in the CT machines. Our data represents a perfect example of an object detection problem and we used the "You Only Look Once" (YOLO) v5 model [10]. The YOLOv5 is a neural network architecture that has been shown to perform well on object detection problems. It is an evolution of the YOLO [5] architecture with modifications. As the name suggests, the images as passed through the network only once. The model divides the images into a grid, and within each grid it predicts the probability of the existence of bounding boxes with center within the grid, along with their height and width. The network architecture otherwise follows the typical structure of convolutional neural networks, with multiple convolutional layers in sequence that decrease the size of the tensors in height and width while inflating the number of channels, with fully connected dense layers at the end. In our model, we used pre-trained weights from YOLOv5 for the convolutional layers, and trained the dense layers on our data.

For our model training, we used the learning hyperparameters as suggested by the authors of YOLOv5. This included an optimization using a OneCycle Learning rate with inital learning rate 10^{-2} and a final learning rate 10^{-4}, with a momentum of 0.937 and weight decay of $5 * 10^{-4}$. We also performed data augmentation to artificially inflate the number of training images and reduce overfitting. Data augmentation included random horizontal flips, and translation of the image by a maximum of 10% of the image size. We initally trained the model on the 128 slices set on 200 epochs to determine the optimal batch size, which was 64. We then train the model independently on the three datasets with a maximum of 750 epochs set to auto-early-stop training if validation score does not improve in an interval of 100 epochs. The three models auto-early-stopped between 620 and 680 epochs. We then select the model at the epoch with highest validation loss.

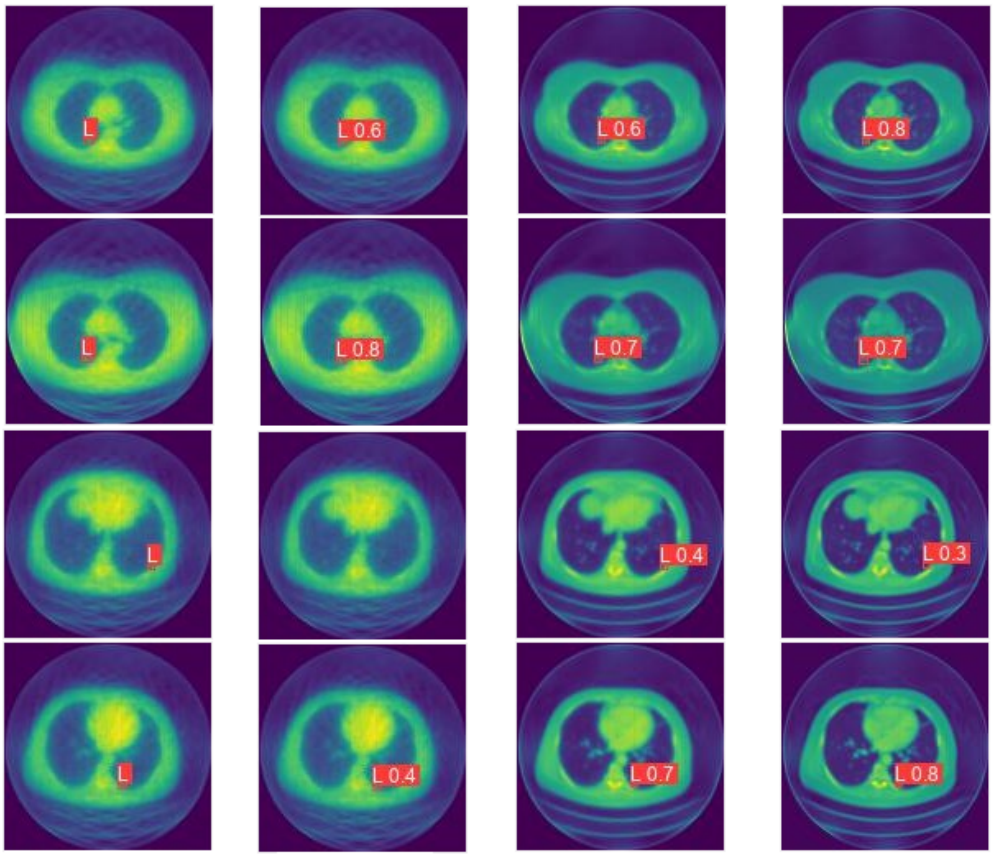


Figure 4: Ground and predictions on a sample of test images

4 Results

4.1 Model Detection Capabilities

Figure 4 illustrates the model results on a sample of test images using prediction and original bounding boxes. We observe that within these four examples, the 64 and 128 slice models are able to locate the lesions with a fair accuracy in terms of size and location. The 16 slice model misses one of the lesions due to the excessive loss in resolution. The ground truth image represents the bounding-box labels from the data. These results suggest that the models pass the minimum baseline of being able to identify lesions in an image.

4.2 Model Performance Comparisons

We decided to use mean Average Precision (mAP) and precision as our two main metrics for evaluating the performance of our models. Mean average precision encompasses the intersection over union (IOU) with a threshold value of .5 for IOU results in an accurate classification. Then precision indicates the instances where lesions were properly identified regardless of the predicted bounding boxes.

Before we analyze the overall performance of our model we need to first establish that it is not over-fitting our training data. Taking a look at the below plots of our data of the mean Average Precision (mAP) score over epochs, we can see that generally the mAP is maximized right around 600 to 630 epochs and does not fall much after that. Thus, we can say that our model does not overfit our training data in any of the scenarios.

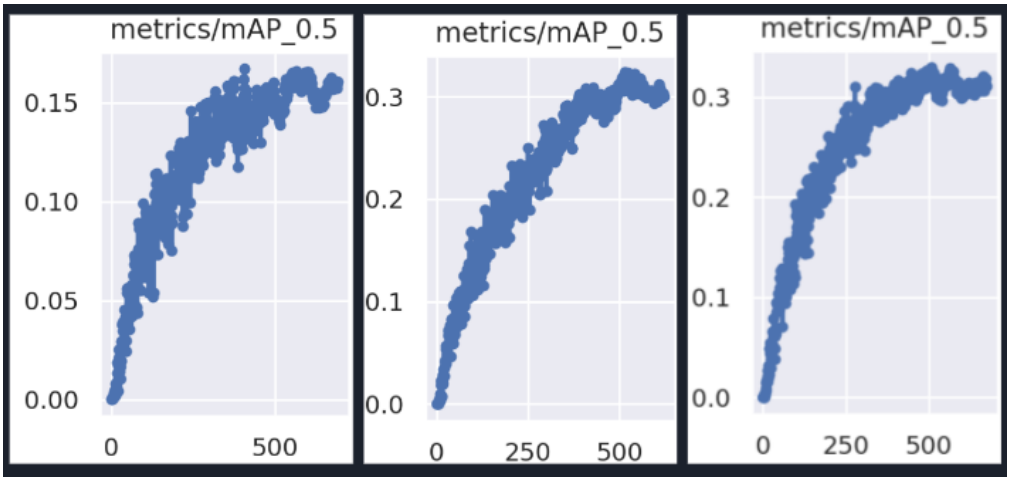


Figure 5: mAP for 16, 64, and 128 slice CT scanner simulation vs Epochs

Table 1: Mean Average Precision and Precision

CT Scanner Slice Count	Mean Average Precision	Precision
16 Slice	0.161	0.468
64 Slice	0.310	0.601
128 Slice	0.326	0.696

In Table 1, we can see that the mAP for 16 slice is far lower than the 128 slice and that along with the sub-par precision tells us that this model did not identify lesions as lesions and it did not provide proper bounding boxes either. On the other hand with 64 slice CT scanner has a mAP for .310 which is similar to that of our baseline of 128 slice which tells us that both models have a precision of .310 with IOU of greater than .5. Therefore, 64 slice CT scanner has comparable performance to the 128 slice scanner. The point in which they differ is in the overall precision metric. The precision metric doesn't mean much in object detection but what it does tell us is that if our model is able to even detect some portion of the lesion even if it was only 10%. In the situation of just identifying even a small portion of the lesion, the baseline model of 128 slice scanner performs slightly better than that of the 64 slice scanner.

5 Discussion

Based on our findings of our experiment, a 64 slice CT scanner has a mAP value of .310 compared to the 128 slice CT scanner which has a mAP of .328. Thus, the 64 slice CT scanner identifies more than 50% of a lesion with similar precision to that of 128 slice CT. Therefore, in order to minimize cost and patient exposure to radiation we suggest that hospitals use 64 slice CT scanners over 128 slice CT scanners. This would save hospitals over $100k for each CT scanner which are used for the purpose of identifying lesions.

One of the key limitations in our analysis is that CT scanners are multipurpose in the sense that they are utilized for more purposes far beyond just lesions. In the case of identifying potential issues with the heart, it is recommended that 128+ slices CT scanner be used to get a better image of the heart. Furthermore, our model was trained for liver, lung and kidneys combined and not each individually. This brings us to potential next steps and future analysis. One key next step that we could take is training each slice count on each type of CT scan and see if the 64 slice performance does well across all types individually. On top of this, we would want to retrain our model on the full 512 by 512 images and see if the model performs better as the images are more clear and sharper. We can also attempt to identify the optimal projection angles to use in a CT scan.

Acknowledgments

We would like to thank Dr. Roarke Horstmeyer for his input and guidance in determining the overall approach to the project.

References

[1] Guide to CT scanners - how to pick a model. LBN Medical. (2022, April 28). Retrieved April 29, 2022, from https://lbnmedical.com/guide-to-ct-scanners/

[2] CT Scanner Price Guide. Medical Imaging Equipment Resources. (n.d.). Retrieved April 29, 2022, from https://info.blockimaging.com/how-much-does-a-ct-scanner-cost

[3] Heggie JC, Kay JK, Lee WK. "Importance in optimization of multi-slice computed tomography scan protocols". Australas Radiol. 2006 Jun;**50**(3):278-85.

[4] Immonen E, Wong J, Nieminen M, et al. "The use of deep learning towards dose optimization in low-dose computed tomography: A scoping review". Radiography **28** (2022) 208-214.

[5] Nie Y, Sommella P, ONils P, Liguori C, and J. Lundgren, "Automatic Detection of Melanoma with Yolo Deep Convolutional Neural Networks," 2019 E-Health and Bioengineering Conference (EHB), 2019, pp. 1-4, doi: 10.1109/EHB47216.2019.8970033.

[6] Yan K, Wang X, Lu L, Summers RM. "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning" Journal of Medical Imaging , **5**(3), 036501 (2018).

[7] Yu, H., Wang, G. (2010). SART-type image reconstruction from a limited number of projections with the sparsity constraint. International Journal of Biomedical Imaging, 2010.

[8] Beyklin G . "Discrete Radon Transform". IEEE Transactions on Acoustics, Speech, and Signal Processing. 35(2). 1987

[9] Andersen AH, Kak AC, Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm, Ultrasonic Imaging 6 pp 8194 (1984)

[10] Jocher G. "YOLO v5". https://github.com/ultralytics/yolov5/releases/tag/v6.1