Independent Study: Lung Cancer Detection, Segmentation and 3D Reconstruction

Done by: Bardh Rushiti, Mentored by: Dr. Zachary Butler
Rochester Institute of Technology

ABSTRACT

The goal of this project is to develop a three-step pipeline for detecting lung nodules, segmenting and 3d reconstructing them. The first step is detection, in which lung nodules are identified in medical imaging scans. The second step is segmentation, in which the detected nodules are isolated and outlined. For both, the first and second steps, different architectures and parameters were explored. The final step is 3D reconstruction, in which a three-dimensional model of the nodules is created. The final solution employs 3D-CNN and multi-task learning to achieve high accuracy in both classification and segmentation. This model is then used to analyze the nodules for signs of malignancy, such as irregular shape or size. The combination of these three steps allows for the accurate and efficient identification of lung nodules and the determination of their potential risk to the patient. Code is available at [0].

Keywords: pulmonary nodule, detection, segmentation, 3d reconstruction

1 INTRODUCTION AND BACKGROUND

The detection and diagnosis of lung cancer remain a significant challenge in the medical field. Early detection is crucial for successful treatment, but the identification of lung nodules, small growths on the lung tissue, can be difficult due to their size and location.

Computed tomography (CT) scans are commonly used for detecting lung nodules, but manually identifying these nodules in CT scans can be a lengthy and laborious process for radiologists. Accurate and precise segmentation of nodules can provide more detailed information about their shape, size, and rate of change, which can be useful for diagnosis and treatment planning. When a nodule is identified, it is often necessary to perform follow-up scans over a period of 3-12 months to assess its growth rate [1].

There are several deep learning techniques that have been used in the literature for the detection and segmentation of malignant lung nodules. These include convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [3]. In recent years, there has been a shift towards the use of 3D CNNs and multi-task learning approaches for the detection and segmentation of lung nodules. These techniques have shown improved performance over traditional methods, particularly in handling complex 3D data and predicting multiple labels simultaneously.

In addition, techniques such as multi-task learning, transfer learning, and ensemble learning have also been used to improve the accuracy and robustness of these models. Other approaches, such as generative adversarial networks (GANs) [4] and autoencoders [5], have been used to augment the training data or to perform unsupervised learning on large datasets.

Here, we explore various network architectures and parameters to investigate lung cancer detection, and segmentation and utilize 3D reconstruction of lung nodules in medical imaging scans to visualize the results and establish if there's any irregular shape or size in the nodule. The final results developed in our project are an extension of the work developed by the NoduleNet team at University of California Irvine [6] and [7], specifically in the areas of prediction post-processing and visualization. To achieve this, we made minor adjustments and de-

veloped python scripts to be integrated into the NoduleNet codebase, which allowed the model to work seamlessly on our machine. These modifications were necessary to ensure that the model was optimized for our specific dataset and computing resources.

EXPERIMENTS 2

Data and experiment configurations 2.1

The data used for these experiments is LUNA16 (LUng Nodule Analysis) [8]. It is a public dataset of CT scans used for the evaluation of computer-aided detection and diagnosis of lung cancer. It was created by the National Cancer Institute (NCI) and the LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) as part of a series of challenges organized by the NCI to stimulate research in the field of lung cancer detection. The dataset includes both two-dimensional (2D) and three-dimensional (3D) images, as well as annotations from up to four radiologists on the presence, size, and location of lung nodules. For these experiments, only CT scans that met the selection criteria of LUNA16 were used. Specifically, if two masks had intersection over union (IoU) greater than 0.4, it was considered that both masks were referring to the same nodule. It's important to note that along different phases of these experiments, different size of data was used, depending on the performance of the model.

Both the 2D-CNN and 2.5D-CNN approaches utilized the TensorFlow framework for data preparation, model building [9], model training, and model testing [10]. The processes of tflite, tfzoo, and tf2.0 were shared between the two approaches, with only minor differences in the specific implementations. Overall, the majority of the preprocessing work was shared between the two approaches, allowing for a more efficient and streamlined workflow.

Given the nature of the problem, it is also important to note that there was a very big gap in the numbers between positive and negative cases. To better regulate this gap, we sampled 3 (instead of 130) negative cases for every 1 positive case.

2.2 *Evaluation Metrics*

Given that this project has multiple subsystems that utilize machine learning methods, each subsystem has a metric with which we can evaluate progress. Classification: Object classification: We use precision and recall as evaluation metrics. Precision measures the proportion of correct positive predictions, while recall measures the proportion of actual positive cases correctly predicted by the model. Both metrics provide valuable insights into a model's strengths and weaknesses, with precision being important for minimizing false positives and recall for minimizing false negatives.

Detection: In object detection, Intersection over Union (IoU) measures the overlap between the predicted bounding box and the ground truth bounding box. It allows us to evaluate the precision of the prediction and is important for accurately localizing objects in the image. We are using IoU in this project to compare the performance of different models and determine the best fit for our task.

It is worth noting some more terminology used in the images that describe the model performance, specifically area, and maxDets. When making predictions on testset, Tensorflow classifies the bounding boxes of the image into 3 categories based on size: small (anything less than 32x32 pixels), medium (anything between 32x32 and 96x96 pixels), and large (anything greater than 96x96 pixels). MaxDets refer to the performance of the model given 1, 10, and 100 detections, with more detections the model has more chances to get the target correctly, and vice versa. They serve, as conditionals for model performance. For example, [Average Pression (AP) @ [IoU = 0.50:0.95 | area = all | maxDets = 100] means the precision is calculated with IoU ranging from 0.5 to 0.95 (with 0.05 as step size, all detections with IoU in this range are considered positive detections), and area ranging from small, medium and large, and maximum number of detections of 100.

Segmentation: In image segmentation, the Dice coefficient (DC) measures the overlap between the predicted segmentation mask and the ground truth mask. It allows us to evaluate the precision of the prediction and is important for accurately isolating the object of interest from the background. We are using DC in this project to present the performance of the NoduleNet neural network. Detection Model:

In this project, we are using the EfficientDet object detection model as our deep learning model for detecting potential pulmonary nodules. EfficientDet has demonstrated strong performance on lung detection tasks and is known for its efficiency, making it a good the choice for our limited resources. Its lightweight and fast design allow for efficient computation, making it an ideal solution for this problem. Overall, the combination of EfficientDet's performance and computational efficiency makes it a promising candidate for detecting lung nodules.

2.3 *2D-CNN Approach*

In this approach, the problem was divided into three independent steps: detection, segmentation, and 3D reconstruction. The initial idea was to use object detection models to classify and localize potential lung nodules and then extract contour information for the nodule shape using classical computer vision techniques, such as contrast enhancement, binarization, and morphological operations. This contour information was then used to create a 3D polygon mesh of the nodule using computer graphics tools.

This approach used single slices from CT scans to train the detector for the classification and localization of lung nodules. Since the detector required 3 channels, the slice was replicated 3 times to match the required dimensionality of the model. Due to limited processing and time, only 30% of the full data from the LUNA16 dataset was used.

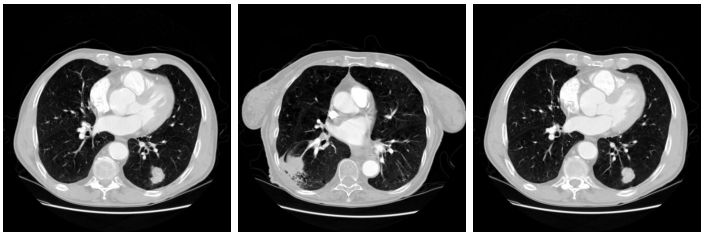
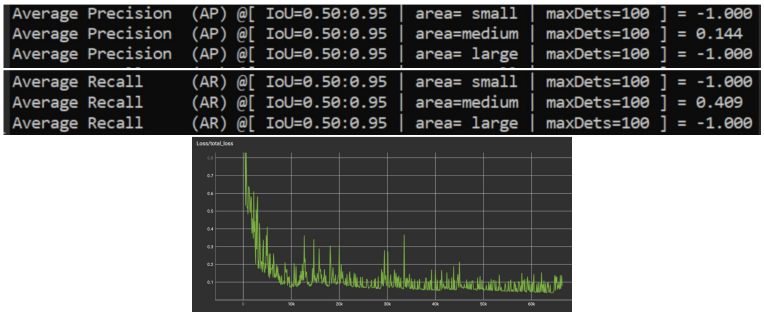


Figure 1: Images of pulmonary nodules

It is worth noting that this machine learning model, like many others, relies on the assumptions of Identical Independent Distribution (IID). This means that the data samples (in this case, CT slices) are independent of one another and that all data samples come from the same static distribution when separated into train, validation, and

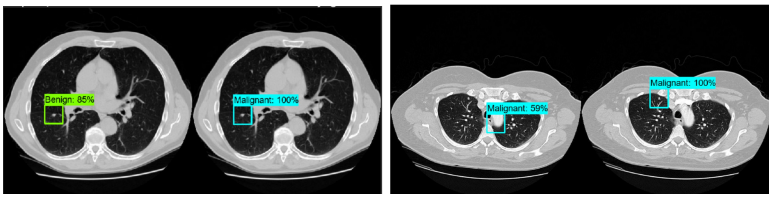
test sets. While this approach violates the "independent" aspect of IID, it was still deemed worthy of experimentation as a baseline performance.

Figure 2: Performance of model trained on 65k steps



After training for approximately 65,000 steps, a relatively high average recall and noticeably low average precision was reached. This approach produced a good recall score; however, due to its high false positives, it suffered with low average precision. The results of this performance can also be seen in the predictions down below.

Figure 3: Predictions of the above-mentioned model



After investigation, we concluded that using single slices results in a loss of information, as the model does not have access to contextual information from other slices in the CT scan. Therefore, it leads to a decrease in the model's performance.

2.4 *2.5D-CNN Approach*

Approach 1 - 2.5D

According to [11], an increase of 20% in performance was reported by giving the model context information of nearby slices. Due to varying distances between slices along the z-axis, multiple distances

were explored for a better understanding of the performance based on this hyper-parameter.

For this approach, a training sample was created by stacking 3 adjacent slices of the CT scan, into a 3-channel RGB image. This gave the model more context information on the location of the thoracic spine, other organs, air pathways, and potential nodules. In technical terms, having adjacent slices in a CT scan creates a data sample that resembles 3D, but it's not quite 3D. This is why it's called 2.5D instead.

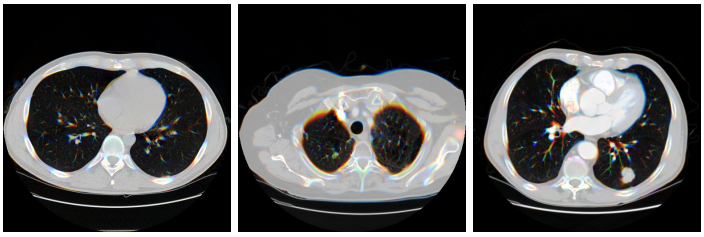


Figure 4: Predictions of the above-mentioned model

After training for approximately 60,000 steps, a slightly higher average recall and noticeably higher average precision was reched.

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.263
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.417
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = -1.000
```

Figure 5: Performance of model trained on 60k steps

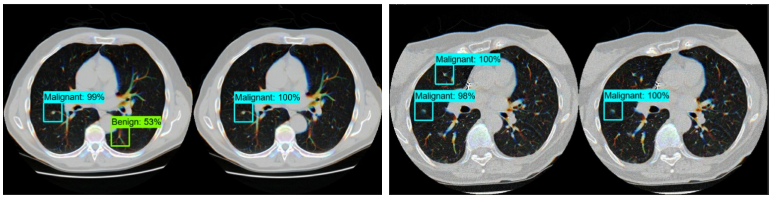


Figure 6: Predictions of the above-mentioned model

This improvement in performance can be seen specifically in average precision. Even though the predictions are slightly better, this performance was not enough to accomplish the end-goal of pulmonary 3D reconstruction.

Approach 2 - Focal Loss

As mentioned previously In the construction of the problem, the data was greatly imbalanced, and we believed that this was one of the reasons for the low performance. In this approach, we explored focal loss [12], as a way address class imbalance by applying different weights to the loss function depending on the class of the sample. For example, if the minority class has a smaller number of samples, the loss function may assign a higher weight to the samples in this class, effectively "focusing" more on the minority class and reducing the bias towards the majority class. This can help the model learn to better distinguish between the two classes and improve its overall performance.

$$\mathscr{FL} = -\frac{1}{N}\sum_{i=1}^{N}\left[-\alpha_t(1-p_{i,y_i})^{\gamma}\log p_{i,y_i}\right] \tag{1}$$

$$p_{i,y_i} = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{if } y = 0 \end{cases}$$

where N is the number of samples, $p_{i,j}$ is the predicted probability of the ith sample for class j, y_i is the true class for the ith sample, and γ is the focusing parameter which is used to down-weight the loss assigned to well-classified examples.

The focal loss at its base is a cross-entropy loss function, with two additional parameters: modulating factor (γ) and weighting factor (α). Modulating factor (γ) reduces the loss contribution from easy examples and extends the range of probabilities that contribute to a low loss value. On the other hand, the weighting parameter (α) is usually the inverse class frequency. α_t is the weighted term whose value is α for positive. See Figure 7 for a more detailed visualization of the influence of γ in the focal loss function.

It is important to note that the optimal value of the modulating factor and weighting factor will depend on the specific characteristics of the dataset and the desired performance of the model. Some set of factors may be more effective in addressing class imbalance in some cases, but it could also lead to overfitting or other issues if set too high.

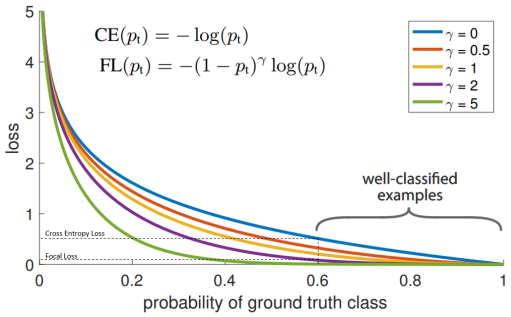


Figure 7: Comparing Focal Loss with Cross Entropy Loss [12]

$\gamma = 2.5$

Increasing the modulating factor in focal loss to 2.5 we expected a likely stronger focus on the harder examples/ classes in the loss function. “Harder examples” refer to the examples that the neural network classifies incorrectly, and does so confidently. Therefore, increasing the modulating factor to 2.5 will result in a greater emphasis on the harder examples class in the loss function, potentially leading to improved performance on this class.

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.231
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.435
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = -1.000
```

Figure 8: Performance of model trained on 60k steps

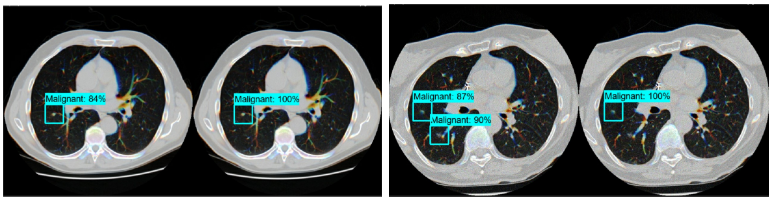


Figure 9: Predictions of the above-mentioned model

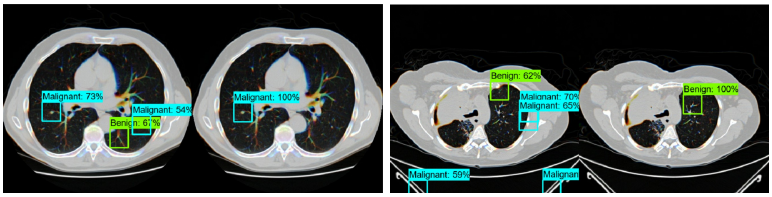
$\gamma = 2.5, \alpha = 0.75$

Additionally, increasing the weighting factor in focal loss to 0.75, we expected a slightly emphasized focus on the minority class, which were the nodules with malignant cases. However, compared to the results above, this change in weighting factor showed a slight increase in average recall, but a massive decrease in average precision. Therefore, this

Figure 10: Performance of model trained on 60k steps

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.055
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.460
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = -1.000
```

Figure 11: Predictions of the above-mentioned model



Although tuning the focal loss was a great educational experience as most datasets in the wild are not perfectly balanced.

After the experimental phase, these two approaches proved to be suboptimal solutions for this problem because of the high number of false positive and negative predictions. These incorrect predictions would need to be handled by a) developing various if-statements to handle these cases better or b) proposing a new neural network that produce more accurate predictions. With regards to the former suggestion, I don't think a classical programming approach would workence making the system less reliable and prone to over-customizing (overfitting) the solution to the current data. Regarding the latter, there is an existing approach that produces better results and hence that's the way we proceed in this project.

2.5 *3D-CNN Approach - NoduleNet*

Handling each task separately proved not to be the right approach because it had some intrinsic limitations. First, training several deep convolutional neural networks is resource-intensive and time-consuming. While detection, and segmentation component of the pipeline serve a different purpose, both extract feature representations with the goal of characterizing lung nodules. Second, the performance of the whole pipeline may not be optimal because focusing and training individual systems in the pipeline does not allow communication and learning feature representation collaboratively.

Here comes in multi-task learning (MTL) and feature sharing as an efficient way to combine different tasks with the same base, con-

volutional neural network. The work of [insert NoduleNet paper] provides an end-to-end framework for solving pulmonary nodule candidate screening, false positive reduction, and segmentation, consisting of three components that share the same underlying feature extraction backbone. For the last and final approach, our work builds on NoduleNet to achieve the final result of pulmonary malignant nodule 3D reconstruction.

NoduleNet is a multi-task learning neural network with three (3) final heads.

Nodule candidate screening: The authors of the paper use a 3D convolutional layer followed by two 1x1x1 convolutional layers to generate classification probabilities and six regression terms for each anchor (a 3D box with specified central coordinates, depth, height, and width) at each voxel on a feature map. They use cube anchors of size 5, 10, 20, 30, and 50 and minimize a multi-task loss function.

Decoupled false positive reduction: They use a different network for false positive reduction, using 3D region of interest pooling from an earlier feature map with a small receptive field and minimizing the same multi-task loss function as the nodule candidate screening network.

Segmentation: For segmentation refinement, they upsample a cropped high-level feature map and concatenate it with low-level features, rather than using a downsampled feature map as in previous approaches. This allows them to only upsample the regions with nodules, saving GPU memory and making whole volume input feasible during training and testing. The segmentation refinement network minimizes the soft dice loss between the predicted and ground truth masks of the input image.

This approach outperformed previous state-of-the-art deep learning based method by 0.95% on DSC, without the need to train a separate and dedicated 3D DCNN for nodule segmentation.

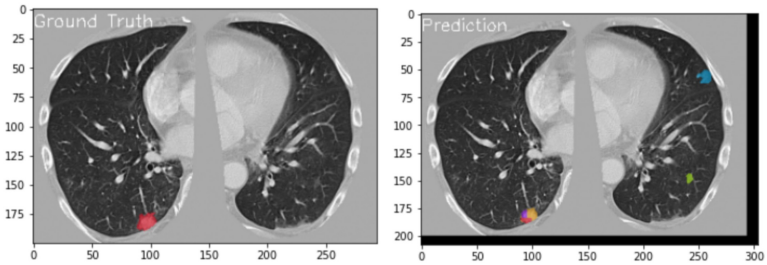


Figure 12: Performance of NoduleNet (Ground-Truth vs Prediction)

Values on the left, are ground truth of the annotated pulmonary nodule. On the right, are the predictions made by NoduleNet, where warmer the values, the higher likelihood of nodule existing in the location, and vice versa.

3 RESULTS

On top of NoduleNet, we develop postprocessing software and 3D reconstruction visualization software that accomplishes the goal of the project.

Post-processing: As NoduleNet produces location and probability of pulmonary nodules, we used that information to select the potential nodules with the highest likelihood, with a thresholding value of greater than 95%. We use the location to extract the contour of the potential pulmonary nodule at different slices of the prediction mask, resulting in a list of contours. With the help of existing matplotlib3d libraries, it was possible to create 3D reconstructions based on 2D contours. The final result of the potential lung nodule is shown below:

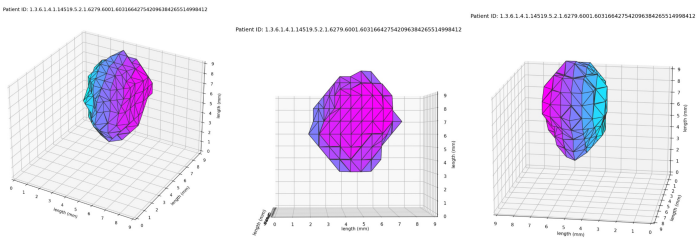


Figure 13: Final 3D Reconstruction of Pulmonary Nodules

CONCLUSION 4

The aim of this project is to create a three-step pipeline for identifying lung nodules, isolating them, and creating 3D reconstructions. The first step involves the identification of lung nodules using various architectures and hyperparameters. The second step involves the segmentation of these nodules, isolating and outlining them for further analysis. Finally, the resulting network is used to perform 3D reconstruction of the nodules, enabling more accurate diagnosis and treatment planning. The final solution employs 3D-CNN and multi-task learning to achieve high accuracy in both classification and segmentation.

REFERENCES 5

[0] https://github.com/Bardhitoo/lung_cancer_proj
[1] Kalpathy-Cramer, J., Zhao, B., Goldgof, D., Gu, Y., Wang, X., Yang, H., Tan, Y., Gillies, R., Napel, S.: A comparison of lung nodule segmentation algorithms: methods and results from a multi-institutional study. Journal of digital imaging 29(4), 476–487 (2016)
[2] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39(6), 1137–1149 (2017)
[3] P.-P. Ypsilantis and G. Montana, “Recurrent convolutional networks for pulmonary nodule detection in CT imaging,” 2016, arXiv:1609.09143. [Online]. Available: https://arxiv.org/abs/1609.09143
[4] Y. Onishi "Automated Pulmonary Nodule Classification in Computed Tomography Images Using a Deep Convolutional Neural Network Trained by Generative Adversarial Networks" BioMed ResInt: PMC6334309 (2019)
[5] K. Matao et al "Feature Representation Using Deep Autoencoder for Lung Nodule Image Classification" (2018), 3078374
[6] Hao Tang, Chupeng Zhang, Xiaohui Xie "NoduleNet: Decoupled False Positive Reductionfor Pulmonary Nodule Detection and Segmentation", CoRR, 2019, arXiv: 1907.11320

[7] https://github.com/uci-cbcl/NoduleNet
[8] "The LUNA16 dataset: Towards a standard for evaluating lung nodule detection algorithms," by F. Isensee, P. E. Golland, and K. Mori, Medical Physics, vol. 44, no. 6, pp. 2197-2207, 2017.
[9] Anton Morgunov "How to Train Your Own Object Detector Using Tensorflow Object Detection API," , Published on the Neptune Blog, (2022), https://neptune.ai/blog/how-to-train-your-own-object-detector-using-tensorflow-object-detection-api.
[10] Anton Morgunov "TensorFlow Object Detection API: Best Practices to Training, Evaluation Deployment", (2022), https://neptune.ai/blog/tensorfl
object-detection-api-best-practices-to-training-evaluation-deployment.
[11] Guilherme Aresta1, Teresa Araujo, Colin Jacobs, Bram van Ginneken, Antonio Cunha, Isabel Ramos, and Aurelio Campilho, "Towards an Automatic Lung Cancer Screening System in Low Dose Computed Tomography" Journal of Image Analysis for Moving Organ, Breast, and Thoracic Images, p321-p329, 2018
[12] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection" in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017

6 APPENDIX

Over the past 14 weeks, you have been working on a project to develop a three-step pipeline for detecting, segmenting, and 3D reconstructing lung nodules in medical imaging scans. The goal of this project was to create a solution that could accurately and efficiently identify lung nodules and determine their potential risk to patients.

You worked on average 10-15 hours per week to accomplish this project, and during that time, you completed the following tasks on a weekly basis:

Week 1: Conducted research on existing methods for lung nodule detection, segmentation, and 3D reconstruction. Week 2-4: Implemented and tested different architectures and parameters for 2DCNN lung nodule detection. Week 5-8: Implemented and tested different architectures and parameters for 2.5DCNN lung nodule detection. Week

8-9: Developed a 3D reconstruction step using 3D-CNN and multi-task learning. Week 10-11: Combined the detection and segmentation steps into a single pipeline and further optimized the parameters - NoduleNet. Developed and integrated software to enable smooth running of NoduleNet in my machine. Week 12-13: Integrated the 3D reconstruction step into the overall pipeline and further factorized code. Conducted testing on a larger dataset of medical imaging scans to evaluate the performance of the pipeline. Week 14: Finalized the pipeline and wrote a report on the results of the project, including the code used, which is available at [0].

This project you were able to develop a pipeline that could detect, segment and reconstruct lung nodules in medical imaging scans with high accuracy and efficiency. This pipeline will be able to analyze the nodules for signs of malignancy, such as irregular shape or size, and provide a potential risk to the patient.