Implementing Deep Learning for COVID-19 Segmentation within Low-Radiation CT Scans

Aayush Goradia
Duke University
`aayush.goradia@duke.edu`

Nathan Parikh
Duke University
`nathan.parikh@duke.edu`

Jason Stranne
Duke University
`jason.stranne@duke.edu`

Abstract

In the COVID-19 pandemic, computer tomography (CT) scans are being utilized to confirm and visualize the presence of COVID-19 in the lungs. A major health concern of CT scans is patients' exposure to radiation, and this paper investigates the effectiveness of image segmentation in identifying COVID-19 in lower-radiation CT scans of the lungs. Decreased radiation was modeled through sub-sampling the original images in the Fourier space, and the Fourier transform served as a proxy for the radon transform in CT image reconstruction. Two optimization techniques, annealing and one-hot relaxed categorical distribution, proved effective in improving segmentation precision, intersection-over-union, and recall scores when compared to naive sub-sampling techniques. As expected, image segmentation on these sub-sampled images performed slightly worse when compared to the original images, but there were promising trends in annealing precision scores and overall IOU patterns. Future work should expand on these preliminary findings to investigate the effectiveness of these optimization techniques on the radon transform.

1 Introduction

As one of the deadliest pandemics in history, the novel coronavirus (COVID-19) pandemic continues to pose a significant challenge to individuals worldwide. One of the primary tools implemented in order to halt the spread of the virus is robust diagnostic testing, which has come in forms such as molecular (PCR) tests and antigen tests. The role of imaging modalities as a first-line diagnostic test has been limited, mainly because of the ease and decent efficacy of the aforementioned tests. However, imaging modalities have been used to confirm COVID-19 diagnoses for patients who have already tested positive and to localize COVID-19 within the lungs.

One such modality is computerized tomography (CT). Computed Tomography (CT) scans are a common imaging modality utilized for examining the presence of disease in the chest. It does this by using x-rays at different angles around the area of interest to produce projections that, when combined, result in cross-sectional images in the body. It is also becoming increasingly common to have algorithms annotate CT scans in order to provide an initial screening for patients, or to help assist doctors [1]. CT scans are able to provide increased clarity when compared to chest radiographs, and are also able to avoid issues with motion associated with MRI scans [2]. In addition to providing high resolution scans, CT is also relatively fast, with the majority of scans taking less than 2 minutes. On the other hand, CT comes with disadvantages such as being expensive and having an increased potential for artifacts. Perhaps the most significant disadvantage from a health perspective, however, is CT's tendency to expose patients to high levels of ionizing radiation, which has been linked to higher risk of cancer [3].

Given this evident disadvantage to CT yet its utility in verifying COVID-19 diagnoses, this study aims to implement deep learning, and more specifically, learn a physical layer such that the optimized physical layer yields COVID-19 segmentation results for CT scanners with lower radiation that are comparable to results for normal CT scanners.

2   Related Work

Given the aforementioned danger that ionizing radiation poses, there have been considerable efforts made in finding ways to reduce the amount of radiation that a patient is exposed to for CT. One possible way of accomplishing this is reducing the number of projections that are taken for each CT slice [2], such that the resolutions of the resulting slices are worse.

For almost all medical imaging modalities, reducing the number of measurements that are taken lowers the quality of the resulting image, often introducing blur or aliasing. Recent work has aimed to utilize deep learning methods in order to improve the performance of undersampled MRI and CT data. One such paper showed the ability of neural networks to dynamically learn how to subsample MRI images and reconstruct them, resulting in improved performance compared to traditional reconstruction techniques [4]. Similar research has been conducted with CT scans, examining the ability of deep neural networks to learn how to reconstruct CT images using sparsely sampled data [1].

In addition to examining these problems from a reconstruction standpoint, there has been a significant amount of research surrounding the segmentation of medical images. A commonly used segmentation model, the U-Net, was specifically designed for use in biomedical image segmentation. Specifically, the U-Net is able to use multiple up-sampling and down-sampling blocks to both create highly localized segmentation masks as well as learn abstract features about the images [5]. In addition, a network known as the HR-Net has recently achieved state of the art performance in many segmentation tasks [6]. This network connects high-to-low resolution convolution streams in parallel while maintaining a high resolution path through the network. Although it is not specifically designed for biomedical image segmentation, HR-Net has been shown to outperform the U-Net on various tasks.

3   Methods

3.1   Data and Pre-processing

The dataset used for this project is the COVID-19 CT scans dataset from Kaggle (`https://www.kaggle.com/andrewmvd/covid19-ct-scans`). It is composed of 20 patients diagnosed with COVID-19 and roughly 200 chest CT scan slices per patient. Experts have provided segmentations of lungs and infections in the form of masks as shown in Figure 1
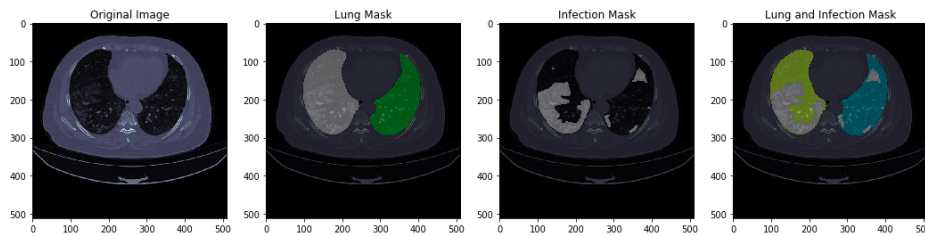

Figure 1: Examples of the images and mask labels that were provided in the dataset

In terms of pre-processing, the CT scan slices came in varying sizes (with most being 512 by 512). Thus, all slices were resized to 256 by 256 pixels. Additionally, the data was normalized to a value between 0 and 1.

No data augmentation was performed on the input dataset. Initially, horizontal and vertical flips of the input images and associated masks were considered. However, it was determined that this type of augmentation could potentially compromise the ability of the network to learn the ideal projection angles.

Finally, the training and test data were split in such a way that training was conducted on data for 16 patients, and testing was conducted on data for the remaining 4 patients. It was crucial that data from only a certain number of patients was considered for training and the remaining patients for testing so that the model could actually learn and make accurate predictions on additional slices if need be.

3.2 Physical Layer

Motivated by reducing the amount of radiation that a patient is exposed to, deep learning can be used to select the most relevant projection angles for COVID-19 segmentation. Through the projection slice theorem, there is a clear relationship between given CT projections and diagonal samples of the Fourier space. Therefore, choosing a subset of these diagonal lines in the Fourier domain is analogous to choosing a subset of the projections. Given the challenge of subsampling diagonal segments of the Fourier space, the problem was simplified to selecting vertical columns of the Fourier domain.

Therefore, the physical layer involved subsampling half of the columns of the Fourier space, thereby similarly reducing the amount of radiation by half. Given there were 256 columns in Fourier space, the optimized physical layer involved learning the ideal 128 columns to choose, while the unoptimized case simply selected every even column.

3.2.1 Annealing

The first method that was used to select the optimal 128 column is known as annealing. This method involves creating a vector of learnable weights for every column. To determine the factor each column is multiplied by, the weights are passed through the sigmoid function as part of the physical layer. The temperature function (which controls the slope of the sigmoid function) is gradually turned up every epoch, forcing the column weights towards one or zero. In order to constrain the model to only select half of the columns, a squared error loss term was appended to the loss function, reducing the difference between the sum of the sigmoid modified weights and the 128 value.

3.2.2 Relaxed One-Hot Categorical Distribution

The second optimization technique was a relaxed one-hot categorical distribution. This is a method that has recently been proposed to overcome the fact that subsampling is non-differentiable. When using this distribution, we encourage the model to learn the weights of a 128 by 256 one-hot probability distribution. Gumbel-softmax sampling is used to allow the subsampling of the distribution to be differentiable [7]. Every time the physical layer is used, a one-hot encoding is sampled from each row, allowing for 128 columns to be chosen. Since the physical layer would constantly change depending on the sample, the 128 most likely logits from the distribution were selected as columns for the Fourier filter after training the physical layer and model weights together for approximately 20-30 epochs. Every image was then passed through this constant physical layer, and the model weights were further refined.

3.3 CNN Architectures

The CNNs used for this project were the previously mentioned U-Net and HR-Net. Given that the task at hand is image segmentation, both CNNs are mostly convolutional in nature and thus have parameterizable layers that are primarily convolutions rather than fully connected layers.

Figure 2 represents a general U-Net structure. The specific U-Net implemented in this project had 4 up-sampling and 4 down-sampling blocks, as well as 32 initial filters and several skip connections.

The HR-Net implemented in this project was based off of an implementation that featured four stages (`https://github.com/niecongchong/HRNet-keras-semantic-segmentation/commits?author=niecongchong`). The first stage involves only high-resolution convolutions, while the remaining three stages add to those convolutions with varying resolution convolution to form high-to-low convolution streams in parallel, as referenced earlier. Figure 3 represents an example HR-Net.
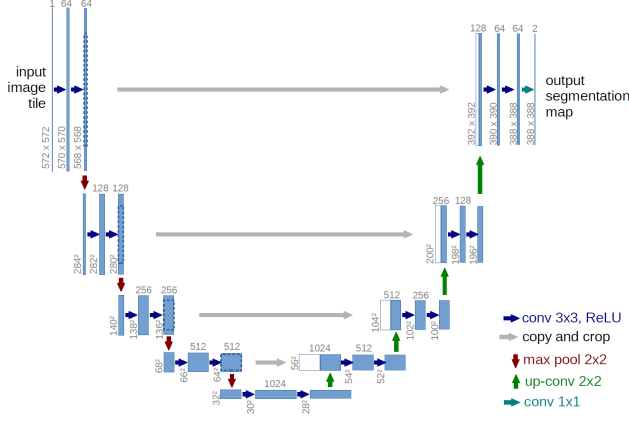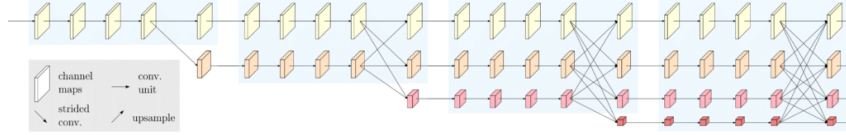
Figure 2: Visual representation of the U-Net architecture [5]

Figure 3: Visual representation of the HR-Net architecture [6]

3.4 Metrics

Several different metrics were used to evaluate the results. The precision and recall scores were recorded at many different thresholds for all of the models. This was chosen as opposed to the AUC score of an ROC curve in order to highlight the performance of the model on classifying true positives. This is primarily due to the fact that all patients in the dataset had already been diagnosed with Covid, and this task is primarily used to locate COVID-19 in the body. There are also multiple metrics that aim to analyze the similarity between the predicted COVID-19 masks and the groundtruth COVID-19 masks. One such metric, the IOU score is defined as the quotient of the intersection of the masks and the union of the two masks. A similar metric, the dice coefficient, was not reported in the results. As can be seen in the following section, the dice score was included in the loss function for training. For this reason, it would be improper to report the dice coefficient.

3.5 Training

All training was performed using the TensorFlow library and the Google Colab Pro resources for computation. Training was performed for 50 epochs in all experiments, and an Adam optimizer with an initial learning rate of 0.0005. Due to computational limitations, a batch size of 16 images was utilized during training.

Multiple studies that have been conducted to evaluate the performance of different loss functions with semantic segmentation. Specifically, segmentation problems with unbalanced classes are of particular interest. For most of the images, the segmented COVID-19-19 mask only makes up a fraction of the CT scan. It has been observed that the focal loss and the dice loss perform well in cases of unbalanced classes, while the binary cross entropy loss tends to perform well on a pixel-by-pixel basis [8]. For this paper, a combined binary cross entropy and dice loss was used to balance these tradeoffs. The loss was defined as:

$$-\alpha[y_i \times log(\hat{y}_i) + (1 - y_i) \times log(1 - \hat{y}_i)] - 2(\frac{y_i \times \hat{y}_i}{y_i + \hat{y}_i})$$

The alpha parameter was chosen to be 10.0, and was determined through cross validation. There was an additional loss term for the annealing experiment, which was previously described in the physical layer section.

4

4 Results

When running the models, four cases were considered: (1) unmodified dataset with no physical layer, (2) subsampled dataset with unoptimized physical layer, (3) subsampled dataset with optimized physical layer via annealing, and (4) subsampled dataset with optimized physical layer via relaxed one-hot categorical distribution. For the unmodified dataset, the post-processed 256 x 256 images were fed into both the U-Net and HR-Net architecture. Sample infection masks for these trials can be found in Figure 4. The HR-Net achieved a maximum IOU of 0.439 at a threshold of 0.3 while the U-Net achieved a maximum IOU of 0.282 at a threshold of 0.3 (Figure 5).
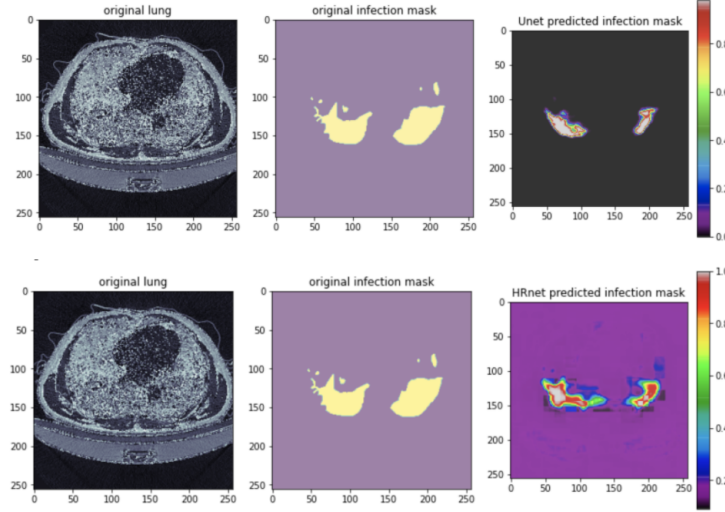
Figure 4: The predicted infection masks for both U-Net and HR-Net on the unmodified dataset. HR-Net has a wider mask while U-Net predicts values closer to the binary extremes, 0 and 1.
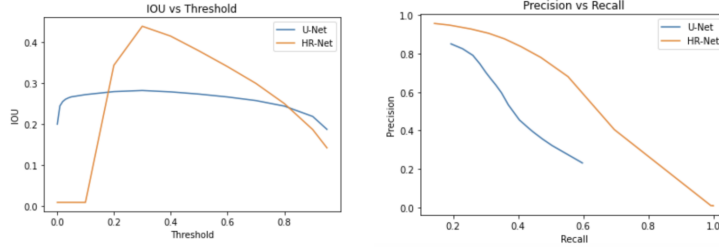
Figure 5: The IOU vs Threshold (left) and Precision vs Recall (right) plots for U-Net and HR-Net on the unmodified dataset.

Although HR-Net performed better on the unmodified dataset, the U-Net architecture performed significantly better when the physical layer was introduced. The HR-Net physical layer appeared to maintain a random distribution of column weights rather than learning a more ideal column weighting pattern. This led to low IOU values for most cases that involved a physical layer, and thus, we decided to move forward with the physical layer implementation in solely U-Net. In analyzing the physical layer, the first simulation consisted of an unoptimized physical layer where every other column was selected in the Fourier space, which represents 50% sub-sampling. Analyses were conducted with 25%, 50%, and 75% sub-sampling of the original Fourier transform, and 50% sub-sampling optimized the trade-off between precise results and over-sampling the Fourier space. After choosing 50% sub-sampling, the physical layer and U-Net were optimized together, and Figure 6 depicts these results.

At the threshold of 0.3, the IOU was 0.030 for the unoptimized physical layer, 0.129 for the annealing-optimized physical layer, and 0.149 for the one-hot-optimized physical layer. Higher IOU scores

were seen at lower thresholds for both optimized physical layers. The scenarios for sub-sampling, optimization, and type of optimization can be seen on the same plot in Figure 7. Precision-recall curves were calculated across a range of thresholds from 0.001 to 0.95.
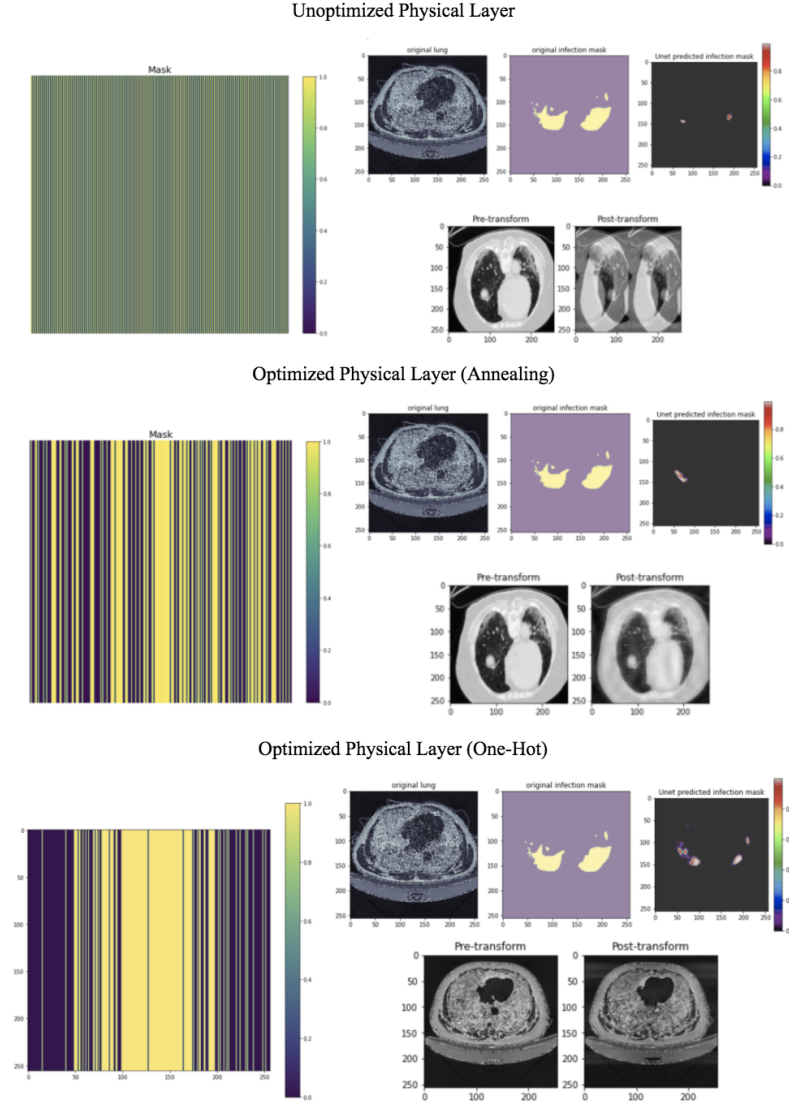

Figure 6: The mask (left), mask predictions (upper right), and transformed images (lower right) are shown for the unoptimized, annealing-optimized, and One-Hot-optimized scenarios.

As described in section 3, the columns of the Fourier domain were sampled due to the difficulty of propagating the loss through a true back projection method. However, the team continued to work on the radon transform and eventually utilized a combination of Python packages that allowed for the trainable layer to contain a backprojection. The Ram-Lak filtered radon transforms of all the images were created containing 256 artificially created projections. Further tests were run allowing the model to selectively learn 128 of the most important projections using the relaxed one-hot encoding method discussed earlier. The radon transform filter and the transformed images are shown in Figure 8.

5   Discussion

In order to create a functional physical layer, it was essential that the U-Net and HR-Net segmentations were effective in localizing COVID-19-19. The IOU curves of both hover in the 0.25-0.45 range,
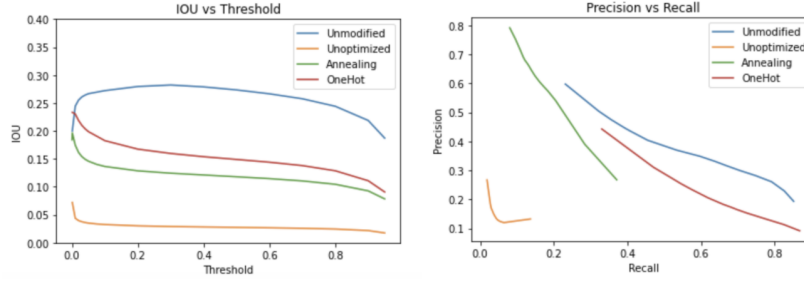
Figure 7: The IOU vs Threshold (left) and Precision vs Recall (right) plots for unmodified dataset, unoptimized physical layer, annealing, Relaxed One-Hot scenarios.
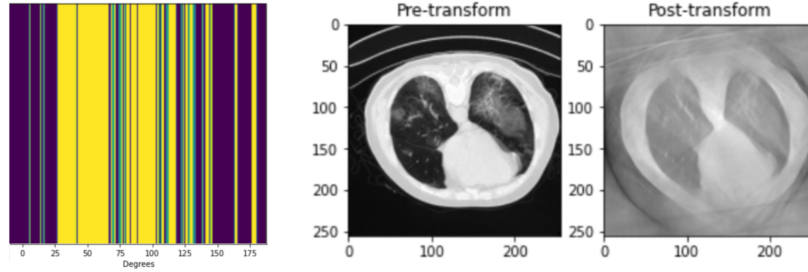

Figure 8: Learned radon transfer mask (left) and sample image shown before and after optimally sub-sampling the radon transform (right).

which are less than the commonly accepted 0.50 standard, but indicate potential given that the model is trained on only 20 patients' CT scans.

In comparing the two methods of physical layer optimization, the annealing mask samples a distribution of frequency ranges with a slight bias towards low-frequency columns while the one-hot mask samples primarily low-frequency columns in the middle. One-hot performs slightly better than annealing on the IOU scale, and both IOU scores decrease slightly as the threshold is increased. From a precision-recall standpoint, the annealing optimization tends to favor precision while the one-hot optimization favors recall. This is noteworthy, given that in a clinical setting where precision in COVID-19-19 detection is highly valued, annealing optimization may be favorable despite lower IOU scores and a lower recall rate.

Both physical layer optimization techniques drastically improve IOU and precision-recall performance when compared to the non-optimized physical layer that samples every other Fourier column. IOU improves by at least 0.10 across the 0.1 to 0.5 thresholds, and precision and recall values increase across all thresholds. These results indicate the potential of sub-sampling optimizations in the Fourier space and highlight patterns for optimization that utilize a combination of low-frequency and high-frequency samples.

As expected, the physical layer optimizations of the sub-sampled dataset did not reach the same IOU and precision-recall success as the unmodified dataset without any physical layer. However, the annealing-optimized physical layer did produce higher precision scores, albeit at lower recall rates, than the original dataset. This further validates the potential for annealing optimization in high-precision environments, such as the clinical setting. Since the IOU scores were within 25-50% of the unmodified dataset, these optimizations show promise in reaching similar segmentation success as the original dataset. Future studies should look into improvements in these optimization techniques, a few of which are described later in the paper.

Apart from the challenges mentioned above regarding the physical layer and the models, several factors relating to the actual dataset added an additional layer of complexity to this segmentation task. Specifically, the dataset was relatively small, there were some poor labels, and the CT scans were of variable quality. However, as the pandemic continues and more COVID-19-19 CT scan data is

collected, it is inevitable that an expanded dataset with better quality images will arise, which could render these dataset challenges obsolete.

Hence, overall, decreasing the radiation dose in CT scans leads to lower precision in COVID-19-19 segmentation. We developed two sub-sampling filters that improve performance for CT scans with lower radiation, but progress remains to be made in achieving similar success when compared to full-radiation CT scans.

5.1 Future Work

Given the poor accuracy observed during segmentation, it would be advisable to focus on using the physical layer to learn how to reconstruct CT images from subsampled CT scans using the U-Net. After this step, the HR-Net can be utilized to determine the highest quality segmentation masks. It is hypothesized that the U-Net would excel at reconstruction due to its high level of abstraction.

In addition, only preliminary results were able to be collected with regards to implementing the learnable parameters on the actual radon transform data. Given that this method is closest to how CT scans work in real life, future efforts should focus on running more tests with this methodology. It may be ideal to reevaluate this physical layer with a reconstruction task in mind, given the low performance that was observed on segmentation in the initial tests.

Finally, in regards to the radon transform, the U-Net appeared to struggle to learn how to decode the images using only a subset of the projections. However, there was a clear trend in the projections that were deemed to be more important. The model shows a clear bias towards selecting a group of projections centered on the horizontal direction. As can be seen in the Results section, the model found a way to select the projections to maximize the clarity in the middle of the image (where COVID-19 is often found), resulting in noise in the top and bottom of the image. This suggests a potential use case for such a subsampling method in medical applications where only a fraction of the CT scan is clinically relevant. Thus, future work should further investigate the implementation of the U-Net architecture with the radon transform physical layer.

Acknowledgments

References

[1] Donghoon Lee, Sunghoon Choi, and Hee-Joung Kim. High quality imaging from sparsely sampled computed tomography data with deep learning and wavelet transform in various domains. *Medical Physics*, 46(1):104–115, 2019.

[2] Takeshi Kubo, Pei-Jan Paul Lin, Wolfram Stiller, Masaya Takahashi, Hans-Ulrich Kauczor, Yoshiharu Ohno, and Hiroto Hatabu. Radiation dose reduction in chest CT: a review. *AJR. American journal of roentgenology*, 190(2):335–343, February 2008.

[3] Amy Berrington de González and Sarah Darby. Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. *Lancet (London, England)*, 363(9406):345–351, January 2004.

[4] Z. Zhang, A. Romero, M. J. Muckley, P. Vincent, L. Yang, and M. Drozdzal. Reducing Uncertainty in Undersampled MRI Reconstruction With Active Acquisition. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2049–2053, June 2019.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*, May 2015.

[6] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. Deep High-Resolution Representation Learning for Visual Recognition. *arXiv:1908.07919 [cs]*, March 2020.

[7] Iris A. M. Huijben, Bastiaan S. Veeling, and Ruud J. G. van Sloun. Deep probabilistic subsampling for task-adaptive compressed sensing. April 2020.

[8] Shruti Jadon. A survey of loss functions for semantic segmentation. *arXiv:2006.14822 [cs, eess]*, September 2020.