Optimized Sub-Sampling for Fast MRI Segmentation Using U-Net Architecture

Ashutosh Raman
Department of Biomedical Engineering
Duke University
Durham, NC 27708
ashutosh.raman@duke.edu

Vani Yadav
Medical Physics Graduate Program
Duke University
Durham, NC 27705
vani.yadav@duke.edu

Abstract

Magnetic resonance continues to be one of the preferred modalities of imaging because of greater soft tissue contrast and ease of 3D imaging. However, MR scans can be expensive and take a long time, serving as a drawback for patients. Thus, there is a need for fast imaging and efficient ways of sub-sampling frequency domain data to reduce scan time without compromising feature resolution for segmentation tasks. Convolutional neural networks serve as an avenue for this sampling optimization, as they can be trained to choose the most important frequency (k-space) data, which can reduce average scan time. This paper implements a U-Net to develop an automatic way of segmenting brain images for tumors, while also optimizing sub-sampling of the frequency domain so that only those frequencies that are important to recreate accurate segmentation masks are chosen. The U-Net is progressively made to choose fewer and fewer columns of k-space in order to segment tumors from brain images. In this study, we found that the CNN prefers to sample areas around DC, as well as some higher frequencies, in order to segment accurately. We also found that segmentation ability is good until around ¼ of frequency data is taken, whereupon mask quality deteriorates quickly.

1 Introduction

Convolutional neural networks are used in a wide variety of tasks to optimize processes and solve problems that have not yet been solved either due to complexity of problem or lack of computational resources. Historically machine learning has been used for tasks such as segmentation, object detection and classification. Recently, there have been research studies that expand the functionality of CNNs from optimization of weights to also optimize the data collection process which better informs the network, thereby improving its performance. One of the most common applications of CNNs is segmentation of medical images. MRI is a common medical imaging modality utilizing magnetic properties of nuclei to create 3D images without delivering any radiation doses to patients.

MRI images are acquired by sampling data in the frequency domain which typically takes a long time and not all the data is useful for a particular task or diagnosis. Research studies have found that images reconstructed from selectively under-sampled frequencies can be just as good for many applications as those reconstructed from fully sampled k-space. This selective under sampling can be useful in reducing scan times without compromising feature resolution and improving the overall diagnostic goals. However, selective under sampling is a complex problem and varies from one application to the other.

In this study we implement a CNN using U-Net architecture to segment brain tumors in images reconstructed from optimally under-sampled MRI scans.

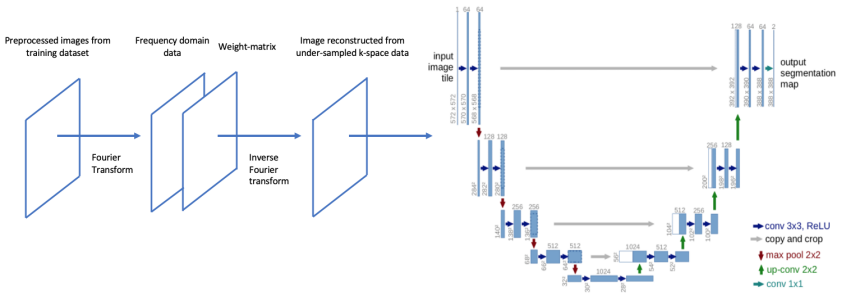

Figure 1: U-net with physical layer

2 Related work

A vast body of research studies investigating sub-sampling for task-adaptive compressed sensing exists. Studies examining model performance on reconstruction, classification and segmentation under stringent sub-sampling rates have been conducted [1,2,4]. There also exists vast literature on segmentation of MRI images of different anatomy using various pulse sequences, which can also be optimized to attain better images.

3 Methods

3.1 Dataset and pre-processing

The dataset used in this study is the subset of the Task1 Brain segmentation dataset from https://arxiv.org/pdf/1902.09063.pdf. It contains 750 brain MRI images segmented into four classes: native T1, Gadolinium weighted T1 contrast, T2, and FLAIR. This study utilizes a subset of 100 2D slices from the dataset to implement binary segmentation classifying pixels into tumor or non-tumor classes. Training and validation images were center-cropped to sizes (240,240). They were then normalized and converted to numpy arrays.

3.2 Physical Layer

Our physical layer consisted of a matrix of trainable weights which would then be multiplied to input image to essentially mask the k-space data. Each column of this matrix was randomly initialized and tiled to create an optimizable matrix of size (240,240). The input to this physical layer is the Fourier transform of appropriately pre-processed data, as shown in Figure 1. Relaxed One-Hot Categorical distribution was implemented to constraint the weights of this matrix to be either 0 or 1 meaning either sample the corresponding frequencies to reconstruct an image or not [5,6]. An image was reconstructed by taking inverse Fourier transform of this under-sampled data, which was then passed as input to the U-net.

3.3 2D U-Net

A 2D U-Net model was created by implementing four encoding and decoding layers with two 2D convolutions in each layer [2]. The model consisted of ten batch normalizations and two instances of dropout at 0.5, in the bottleneck section. ReLU activation was implemented in all layers with Sigmoid function used in the final Dense layer for logits.

3.4 Evaluation metrics

We used accuracy and Binary Cross Entropy loss to quantify model performance. In addition, we also used pSNR (peak Signal to Noise Ratio) and mIoU (mean Intersection over Union) to quantify robustness of reconstruction with under-sampled k-space.

$$PSNR = 10.log_{10}.\frac{MAX^2}{MSE} \tag{1}$$

where MAX is the maximum value possible and MSE is the mean square error.

$$mIoU = \frac{target \cap prediction}{target \cup prediction} = \frac{TP}{TP + FP + FN} \tag{2}$$

where TP is True Positive, FP is False Positive and FN is False Negative

4 Results

In this section, we present results obtained from our study. Control outputs for the U-net without optimized subsampling are provided for comparison. Two methods of subsampling are evaluated here: a regular kernel, allowed to vary to any value in order to best sample the frequency data, and a kernel optimized through a relaxed one-hot categorical distribution, which implements a discrete distribution that pushes the kernel values to be 1 or 0, which is more realizable in MR.

4.1 Training of U-Net Alone

To establish a baseline for comparison of the optimized physical layer, it was important to first run the segmentation task with a basic U-Net. The image with ground truth overlay is shown in Figure 2, with the model prediction shown to its right:

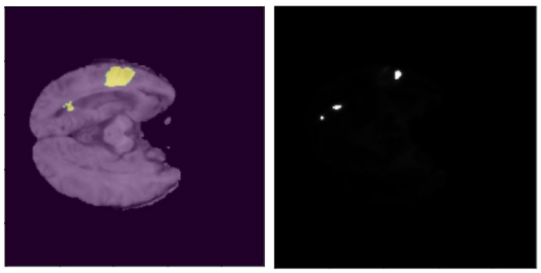

Figure 2: Brain Image with Ground Truth Mask Overlay, and Predicted Mask, for U-Net

The mask is recreated well upon qualitative observation, with both tumor instances showing up in the same locations, albeit with different shapes. To get a quantitative evaluation, model loss, accuracy and PSNR are included in Figure 3:

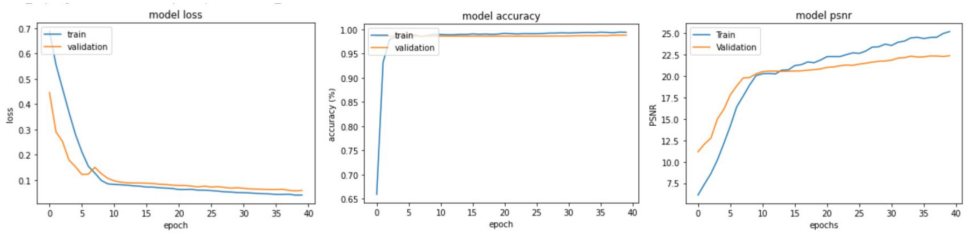

Figure 3: Training and Validation Loss and Pixel Accuracy, as well as Training PSNR, for U-Net

Here, we see that the loss decreased quickly in the first few epochs, after which the decrease was less substantial, and eventually it plateaued. Pixel-to-pixel accuracy appears to be quite high and stagnant throughout training. And PSNR appears to increase rapidly until five epochs, upon which it continues to moderately increase for the remainder of training. The Mean IoU was recorded as .4931

4.2 Training with Optimizable k-space Matrix

To gain an understanding of the k-space preference, we first implemented a k-space matrix that was allowed to vary in weight value. Weights were not constrained to 0 or 1, but this helped to establish how the tiled vector should look. This is shown in Figure 4:

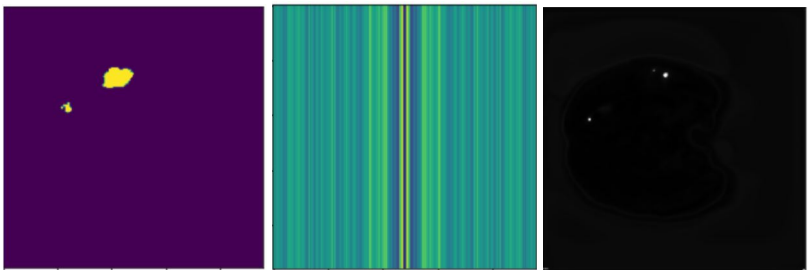

Figure 4: Ground Truth, Optimized k-Space Sampling Matrix, and Predicted Mask

As we see in the figure, the k-space matrix appears to focus on areas around DC, while heavily attenuating DC and some frequencies between preferred columns. There appears to be a ringing pattern, with an on/off sequence. Also, outer frequencies are slightly magnified, perhaps because of their ability to aid in edge-detection. The predicted mask appears to match the ground truth again, but to ensure this, graphs are shown in Figure 5:

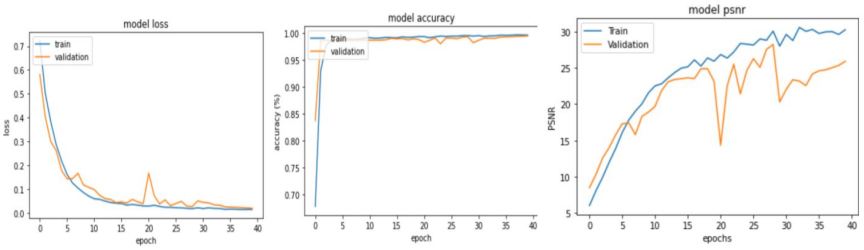

Figure 5: Training and Validation Loss and Pixel Accuracy, as well as Training PSNR, for k-space optimized model

This figure shows similar results to those of the original model. It is important to note a more dynamic validation loss and PSNR, likely due to the dataset changing. Nonetheless, the model appears comparable to the original. The mIoU after training was .4951, similar to the original U-Net.

4.3 Training after Application of Relaxed One-Hot Categorical Distribution

Next, we wanted to constrain the k-space columns to be close to 1 or 0, signifying the usage or omittance of a frequency band. Using a relaxed one-hot categorical distribution, we constrained the model to choose a certain number of k-space columns to take, which is specified as an argument for the physical layer class. Figure 6 shows the k-space matrices optimized for the model when it is forced to pick 120, 60, 30 and 15 samples, with resulting predicted masks also outputted:

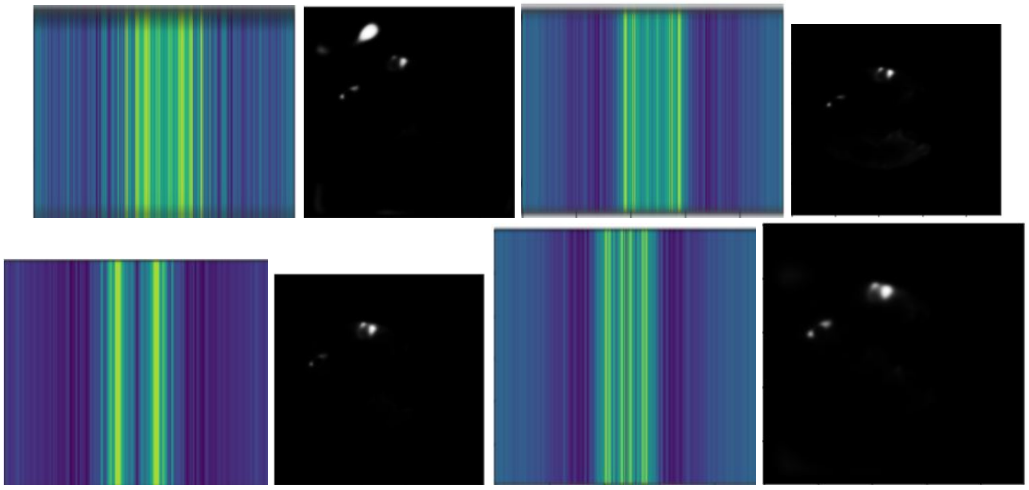

Figure 6: (L to R) 15, 30, 60, and 120 Optimized k-space Matrices and Predicted Masks, after Application of Relaxed One-Hot Categorical Distribution

Once again, the same general k-space preference is observed. The masks seem to indicate a trend of first choosing the areas around DC, then if more samples are allowed, to focus on other important frequencies further from DC. Finally, at 120 samples, we begin to see some preference for outer frequencies. It is important to note that colorbars were not included for lack of space, but the lower sampling matrices had less range in value. Additionally, it can be observed that as number of samples decreases, the predicted mask is less and less accurate, with more false positives and negatives.
To understand this mask change, a comparison of Mean IoU was done in Figure 7 below:

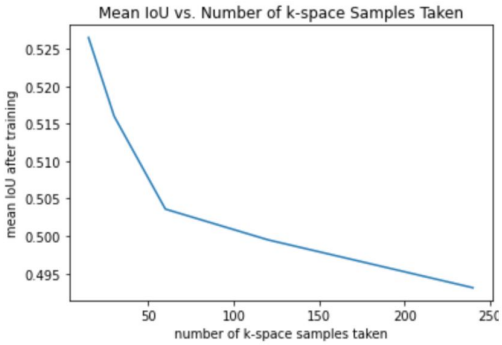

Figure 7: Mean IoU in relation to number of k-space columns sampled

The figure seems to show a decrease in Mean IoU as number of samples increases, which is somewhat counterintuitive for the mask observation above. It should be noted, however, that the decrease is not substantial, and no Mean IoU is too different from the control IoUs.

5 Discussion and Next Steps

Our model trained quite well, and the control case of k-space subsampling predicted closely to this model, as seen in the IoU and predicted mask.
There definitely appears to be a preference for certain frequencies when segmenting the brain for tumors. Because DC is usually oversampled, it makes sense that k-space matrices would attenuate it, since it is unnecessary data. It also makes sense that areas around DC would be preferred, since they contain most of the image signal. High frequencies being slightly preferred has to do with their use

in enhancing image contrast, which helps in segmentation, since that involves delineating borders around scan features.
In the discrete distribution enhanced masks, the pattern mentioned in Results follows this thinking. Signal is the most important metric in CNNs, as it allows images to vary enough that gradients can backpropagate and create expected losses without blowing up or going to 0. For this reason, when constrained to pick lower sample numbers, the model chooses to focus on areas around DC. Without much variation in the k-space tiled matrix, these columns are preferred closely to DC.
Once more samples are allowed, we see a clear deviation from DC, likely due to more possibilities being available. A drawback of the one-hot categorical method is that occasionally, certain columns can be sampled more than once, since their probabilities are more likely than others. This appears to be what is happening here, as areas around DC need to be magnified to enhance tumor features.
Finally, at high sample sizes, we see the ringing pattern of the original k-space matrix again. Since more samples are possible, the model can refine more details. The support of high frequencies is also beginning, since the contrast is necessary for good image reconstruction.
Naturally, we expect mask quality to deteriorate with less samples taken, as image will possibly begin to alias or blur, damaging segmentation power. The increase in false positives and negatives is in agreement with this thought process. It appears that segmentation ability begins to truly deteriorate at 60 samples taken, which when compared to the original image, with 240 columns, is ¼ of the k-space columns. At this point, problems ensue.
This increase in false positives and negatives for tumor location is unfortunately not reflected well in Figure 7. One would expect a curve that increases as number of samples are increased, since IoU is inversely related to number of false positives and negatives. We would expect the curve to stagnate around 60 samples as observed, and only make minor increases from that point, since 60 samples appears to recreate the mask well.
We can link these unexpected results to a number of possible explanations. One has to do with the use of mean IoU. IoU calculation could have been applied incorrectly, causing values to be measured wrongly. Tensorflow's MeanIoU metric subclass was utilized to monitor this metric, but perhaps we should have defined a custom IoU function. Also, it is possible that the CNN is overfitting to account for the lack of data in the lower sampled images, which can improve IoU erroneously. Another potential explanation has to do with the images themselves. Perhaps the images contain certain frequencies of noise, which corrupt reconstruction ability, and these are intentionally filtered out at lower sample sizes.
Perhaps the most plausible explanation, however, is that this decrease in MeanIoU is purely coincidental. The values for each sample size are pretty much equal, hovering around .5. This is already a very low Mean IoU, and minor deviations around this are not of much note.
Compared to the original model, the Mean IoU is pretty similar as sample number increases. The Mean IoU begins to deviate by a full percentage point once samples are decreased. Thus we can confidently say that the optimized physical layer included in the model allows for segmentation on par with the original model.
It would be interesting to further investigate this Mean IoU problem. Another interesting thing to investigate would be to now implement this specific phase and frequency encoding gradient combination in MRI, to generate physical images closely matching those simulated here. This could further inform on new ways to optimize the MRI scan process, in order to decrease scan time while maintaining segmentation fidelity. It could help to generate better initial values for the k-space tiled matrix, so training can converge faster, and to potentially more accurate values.

6 Conclusion

Overall, our results show an ability of the U-Net to accurately segment brain tumors from the brain image data. Additionally, the U-Net can be used to optimize a vector to select the most important columns of k-space to sample, in order to recreate a mask well. We found that even when the U-Net is forced to take fewer and fewer samples of k-space, it can still segment well, up to a certain level of subsampling (about ¼ the samples), before the mask quality begins to deteriorate. Pixel-to-pixel accuracy remains high for all cases, indicating that it may not be the best metric for comparison. Loss decreases and PSNR increases, as any optimizable mask is applied, potentially indicating that the mask helps to focus on important image frequency data. However, there is a trend in decreasing Mean IoU as number of samples increases, indicating that the model is overfitting, or it is possible the metric is not deviating much regardless.

7 Acknowledgments

We would like to thank Dr. Roarke Horstmeyer and Colin Cooke for their help in figuring out how best to implement this idea, as well as their guidance on the project.

References

[1] Y. LeCun, Y. Bengio and G. Hinton, “Deep Learning", Nature **521**, 436–444 (2015).

[2] O. Ronneberger, P. Fischer and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation", arXiv:1505.04597 (2015)

[3] A. Chakrabarti, “Learning Sensor Multiplexing Design through Back-propagation", arXiv:1605.07078v2 (2016)

[4] I. Huijben, B. Veeling and R. van Sloun, “Deep Probabilistic subsampling for task-adaptive compressed sensing", published as conference paper at ICLR (2020)

[5] E. Jang, S. Gu and B. Poole, “Categorical Reparameterization with Gumbel-Softmax", arXiv:1611.01144 (2016)

[6] C. Maddison, A. Mnih and Y. Teh, “The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables", arXiv:1611.00712 (2016)