# Capsule Networks for the Automated Segmentation of Left Atrium in Cardiac MRI

Joseph Chen
Grade 11
Queens High School for the Sciences at York College

# Abstract

Left atrium segmentation is a time-consuming, yet critical step for diagnosing the world's leading cause of death, heart disease. However, limited research has been done to automate this process using capsule networks—an emerging deep learning architecture. This study is the first in literature to apply capsule networks to left atrium segmentation and proposed a new neural network architecture, **U-CapsNet**, that combines a 2D U-Net feature extractor and a capsule network. Although it performed worse than the state-of-the-art with a net difference of 0.1 in dice score, it performed better under heavy class imbalance, and better than a corresponding similarly complex U-Net for larger patches. Therefore, despite capsule networks not being state-of-the-art architectures, their robustness to class imbalances and high performance for their network complexities underpins their potential for the automated segmentation of the left atrium.

# Contents

# 1 Introduction

Coronary artery disease has been the leading global cause of death for the past 15 years (World Health Organization, 2018). The segmentation—the pixel-wise masking of regions of interests in images—of cardiac structures, such as the left atrium, plays a critical role in detecting and diagnosing heart conditions (Zreik et al., 2017). Additionally, organ segmentation is especially important for radiotherapy planning because the locations of at-risk organs that move due to physical processes affects the radiation dosage volumes at any given point (Njeh, 2008).

However, segmenting medical images, such as scans from magnetic resonance imaging (MRI), manually is a time-consuming process. The reason is medical images are high-resolution and volumetric, as stacks of 2D images, and radiologists must accurately produce segmentation masks for each slice or 2D image (Suetens et al, 1993). Furthermore, Vorwerk found that segmentation accuracy varies heavily from physician to physician, which is known as the interobserver variation (Vowerk, 2009).

Artificial intelligence algorithms address the aforementioned time and interobserver variation issues because they provide more consistent performance and are unaffected by psychological factors that hinder humans from making correct decisions on the spot, such as fatigue (Sharma and Arggarwal, 2010). For example, recent successes with deep learning algorithms have underpinned their potential to automate the segmentation process, such as BioMind's AI beating top radiologists in performance and speed for brain tumour expansion classification (Yan, 2018). Moreover, the rise of startups using these algorithms to provide medical decision support services highlights their already successful capabilities for computer-aided daignosis (CAD) devices (Koios DS, 2018; Arterys, 2018).

One example is how Isensee et al. employed a convolutional neural network ensemble for the automated segmentation of various heart structures in cardiac MRI from the 2017 Automatic Cardiac Diagnosis Challenge (ACDC) dataset. Convolutional neural networks are deep learning artificial intelligence algorithms that extract increasingly complex features from images in layers with filters and then predict the images' class; for segmentation, the model would predict the class of each individual pixel of the input medical image. The researchers stacked U-Nets, which are convolutional neural networks that concatenate early layer outputs with later layers in a U-shape to combine information from low-level contextual features with higher-level and more distinct features. This stack, or ensemble, placed first in the ACDC segmentation challenge, with the highest dice scores across most structures, which demonstrates how convolutional neural networks are successful in the challenge of cardiac MRI image segmentation.

However, convolutional neural networks (CNNs) are flawed. They cannot inherently learn pose information, such as the orientation of the image or direction of the objects in the image, because this information is lost during an operation called max-pooling where the image is down-sampled to extract only the most prominent patterns. Consequently, when trained without a significant amount of data augmentation (i.e. rotations), they cannot classify augmented objects correctly on a consistent basis. To address this issue, Sabour et al. proposed capsule networks, a new deep learning algorithm that represents features with multidimensional capsules. These capsules inherently learn pose information because each capsule learns a feature and encodes the various properties of that feature, such as orientation, depending on the dimensionality of the capsule. Capsule networks also take into account the relationships between earlier, simpler features and later, more complex features with routing by agreement—the process in which connections are formed between capsules to develop a hierarchy of features. Sabour et al. demonstrated the potential of the new genre of neural networks by achieving state-of-the-art performance for multiple datasets, such as with a 2.7% test error rate on the smallNORB dataset (Sabour et al., 2017). Furthermore, since the architecture presented in the paper were fairly novel, the researchers illustrate the potential for capsule networks to be improved upon and be extended to different fields, such as left atrium MRI segmentation.

Jimnez-Snchez et al. tested the robustness of Sabour et al.'s capsule networks for automatic medical image classification tasks in comparison to convolutional neural networks. Specifically, they targeted the the main issues with training CNNs for medical image tasks: class imbalances and small datasets. These two are problems for convolutional neural networks because CNNs require substantial amounts of data to learn, and class imbalances often cause the amount of data for one class to be insufficient. The researchers found that capsule network performance was superior to that of CNNs when training with imbalanced data, with higher F1 scores compared to the CNNs. Furthermore, they also found that capsule networks also performed better than CNNs when training with only 50%, 10%, and 5% of the original datasets. These two findings suggest that capsule networks are well-suited to handle medical image classification tasks. Therefore, they have potential for medical image segmentation because segmentation, as previously mentioned, is just a low-level or pixel-wise classification of an image. However, the study only included novel architectures in their assessment of the capabilities of capsule networks and do not address more complex architectures that would be needed for state-of-the-art performance. Additionally, the researchers found that the reconstruction images from the capsule networks on the medical images were much blurrier than easier datasets, such as MNIST for digit classification. They theorized that this may have been due to complex backgrounds in medical images (Jimnez-Snchez et al., 2018). This finding suggests how capsule networks may benefit from a feature extractor to remove the unnecessary features for capsule networks to learn only important representations of the data.

Building upon Jimnez-Snchez et al.'s study, Lalonde and Bagci proposed a deeper capsule network architecture for lung segmentation called SegCaps. This network combines the feature concatenation from U-Nets and the routing by agreement (dynamic routing) algorithm from Sabour et al.'s capsule networks. They augmented the dynamic routing algorithm to use less parameters and memory with their locally-connected routing algorithm, so their capsule network architecture could be extended to larger image resolutions. Additionally, the scientists proposed a new regularization technique for capsule networks where the network reconstructs the original image's region of interest with a separate decoder. In this case, this decoder reconstructed the lung from the original image. The SegCaps architecture had state-of-the-art performance, with a dice coefficient of 0.9848, and was slightly superior to the benchmark CNN that had a dice coefficient of 0.9841. These results confirm Jimnez-Snchez et al.'s finding that capsule networks perform well on imbalanced datasets and further support the idea that capsule networks would work well on left atrium segmentation. Nevertheless, the dataset the researchers used, the LUNA16 challenge dataset, had ground truth segmentation annotations generated by an automated segmentation algorithm instead of radiologists. The authors even acknowledged that some of the annotations were poor, and as a result, they reported dice coefficient medians instead of averages because of poor outlier segmentations. As such, more research is needed to evaluate capsule networks on more reliable medical imaging datasets with radiologist-produced segmentation ground truths. Furthermore, they only utilized full sized images instead of cropped patches (i.e. Jimnez-Snchez et al., 2018) to train their data because LUNA16 is a relatively large medical image dataset with 884 scans and they did not need to sample more data in the form of patches. Therefore, more research also needs to be done on capsule network segmentation performance when trained with patches because training with patches is inevitable for deeper and more memory-consuming models.

# 2 Materials and Methods

**Architectures Tested**

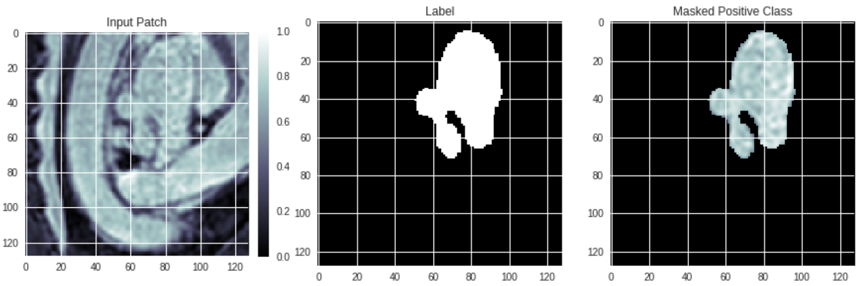
- *State-of-the-Art 2D U-Net*: This convolutional neural network serves as the overarching control for the experiment to evaluate the overall effectiveness of capsule networks for left atrium segmentation.

- *Baseline 2D U-Net*: This convolutional neural network serves as the control for the experiment to evaluate the effectiveness of capsule networks for left atrium segmentation for their complexity level, so it has approximately the same number of parameters or complexity as the proposed **U-CapsNet**. This specific architecture was only added for **Section 3.3** to replace the *basic capsule network* and act as a secondary control.

- *Basic capsule network*: Inspired by Sarbour et al.'s capsule network architecture and augmented for segmentation outputs. This network also serves as another baseline for evaluating the **U-CapsNet**'s performance.

- *SegCaps*: The aforementioned convolutional-capsule network hybrid proposed by Lalonde and Bagci (2018).

- **U-CapsNet**: This new proposed capsule network architecture utilizes a shallower two dimensional U-Net as a feature extractor for a basic capsule network. The idea is that the U-Net would extract the important features and



(a) Example of 128 x 128 Patch Inputs

**Figure 1:** *Example of the cropped inputs into the neural networks.* The white pixels in the label mask represent where the left atrium is, while the black pixels represent the other structures in the MRI. The masked positive class is only fed into the capsule networks for the network to also learn the reconstruction of the positive class from the original image.

the capsule network would learn a hierarchy of features from the filtered images. Consequently, the capsule network would be less susceptible to the noisy backgrounds in MRI scans and would deal with the complex task of left atrium segmentation better (Jimnez-Snchez et al., 2018).

**Dataset** The dataset used to train the neural networks was the cardiac monomodal MRI dataset from the Medical Segmentation Decathlon challenge and King's College London. It is small, with only 20 patients and high variability, but the left atrium segmentation annotations were made by a radiologist and were peer reviewed. The dataset was split into three parts: training set (60%), validation set(20%), and test set(20%).

**Preprocessing** The MRI scans were resampled to 1 mm isotropic spacing because convolutional neural networks cannot inherently learn anisotropic spacings. They were then applied with z-score normalization and fed into the neural networks as 128 x 128 cropped patches using Keras, nibabel, Numpy, and SimpleITK.

**Methods** Python 3.5, Tensorflow, and Keras were utilized to program, train, and validate the capsule networks and convolutional neural networks in Google Colaboratory with their free Tesla K80 GPU. They were trained until the validation loss began to increase using a Early Stopping callback with a patience of 60 epochs. The segmentation performance of each neural network was judged by the Sorensen Dice Coefficient (F1 Score) and was computed on test set using sklearn. To record the data and figures, Overleaf and Evernote were used.

**Training and Sampling** Jimnez-Snchez et al. and Lalonde and Bagci both demonstrated that capsule networks perform well under heavy class imbalance. To test this theory, the neural networks were trained and tested when the data was randomly sampled (natural, imbalanced class distribution) and when the data was sampled so that at least one-third of the incoming batches had a patch with left atrium pixels. Also, the capsule networks were all trained using Sabour et al.'s routing by agreement algorithm and used shared weights instead of weights for every single

| *Architectures* | *Number of Parameters* | *Balanced Sampling* | *Random Sampling* |
|---|---|---|---|
| U-Net (SOTA) | 27,671,926 | **0.88** | 0.43 |
| Capsule Network (Basic) | — | 0.11 | 0.09 |
| SegCaps | 1,416,112 | 0.11 | 0.07 |
| U-CapsNet | 4,542,400 | 0.78 | **0.46** |

**Table 1:** Networks Performance. Dice coefficients were evaluated on a separate test set from the training and validation sets.

capsule—proposed by Lalonde and Bagci—to reduce the memory load(Sabour et al., 2017; Lalonde and Bagci, 2018). In addition to the crops, random elastic deformations, rotations and scaling were applied to the images on-the-fly, with each occurring on about 10% of incoming input samples.

The neural network with the highest F1 score was evaluated as being the best at left atrium MRI segmentation.

# 3 Experiments and Results

## 3.1 The Sorensen-Dice Coefficient or F1 Score

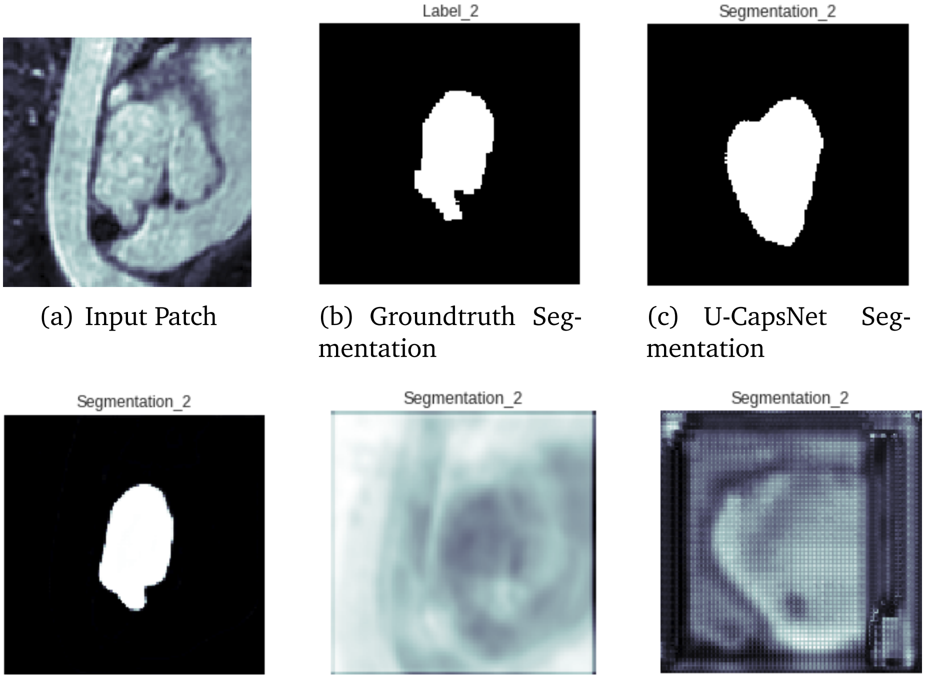
To evaluate each network's performance, their outputted segmentation masks of the test set were evaluated in comparison with the ground truth labels using the dice coefficient:

$$Dice(\boldsymbol{p}_i, \boldsymbol{g}_i) = 2\frac{\sum_i \boldsymbol{p}_i \boldsymbol{g}_i}{\sum \boldsymbol{p}_i + \sum \boldsymbol{g}_i}, \tag{1}$$

Where $\boldsymbol{p}_{\mathrm{i}}$ represents the predicted segmentation from the models and $\boldsymbol{g}_{\mathrm{i}}$ represents the labeled ground truth. It measures the overlap between the predictions and the labeled ground truth from zero to one, where zero indicates no overlap and one indicates perfection. This metric was chosen because it handles class imbalances well and is commonly used by by other researchers for segmentation evaluation (Milletari et al., 2016).

## 3.2 Smaller Patches for Capsule Networks

The proposed U-CapsNet slightly outperformed the U-Net when the inputs patches were randomly sampled because its dice coefficient of 0.46 was greater than that of the baseline's (0.43). This can be attributed to the U-CapsNet's basic capsule network decoder and how capsule networks are slightly more robust to class imbalances than convolutional neural networks. The trade-off was that the U-CapsNet was less effective at dealing with class imbalances (i.e. in **Table 1.** with a 0.32 difference in dice scores between sampling types) because of the presence of a less robust CNN feature extractor.



(a) Input Patch (b) Groundtruth Segmentation (c) U-CapsNet Segmentation

(d) 2D U-Net Segmentation (e) CapsNet Segmentation (f) SegCaps Segmentation

**Figure 2:** Examples of masked patch-wise predictions from the tested neural network architectures on a 128 x 128 patch.

However, the regular capsule network and SegCaps performed significantly worse than the baseline 2D U-Net with both the positive sampled data and randomly cropped data. The patch-wise approach may have been the cause because they often could not differentiate between the high contrast structures—which would require more contextual information to learn it. Nevertheless, their dice scores did not vary much when switching from positive sampling to random sampling, which indicates a small degree of robustness to class imbalance.

Additionally, the baseline U-Net was superior to all the capsule networks when trained with positive sampling. It led with a 0.1 dice score increase over the closest network in performance, the U-CapsNet. This demonstrates how the U-Net is more effective at the task of medical image segmentation than capsule networks. Nevertheless, only being 0.1 dice away from near state-of-the-art performance is normal considering how novel capsule networks currently are.

Furthermore, the regularization reconstruction loss suggested by Lalonde and Bagci to help the network learn more global contextual information was not effective for patch-wise training. It also may have precipitated the severe underfitting of SegCaps and the basic capsule network, as shown in **Figure 2** and by their abysmal segmentation performance in **Table 1**.

## 3.3 Larger Patches for Capsule Networks

In this round of experiments, the patch size was increased to 256 by 320, and instead of using the Basic Capsule Network, a lower complexity U-Net (Baseline) was employed to serve as an additional and more useful benchmark for the U-CapsNet. Like the previous experiments, the networks were trained with balanced and random sampling, separately.

The introduction of larger patches for training the capsule networks greatly improved performances, as demonstrated by the 0.35 dice increase for the balanced sampled SegCaps architecture, the 0.27 dice increase for the random-sampled trained U-CapsNet, and the 0.05 dice increase for the balanced-sampled trained U-CapsNet in **Table 2**. The increased contextual information from having larger input images may have been the cause behind this increase, which suggests that capsule networks need fuller views of inputs to learn properly.

A more nuanced reason for the performance increase could be how taking 256 by 320 patches eliminated the occurrence of blank patches, inputs that were completely black. With 128 by 128 patches, the generators would occasionally sample from the black edges of the cardiac MRI's, and because of additional zero-padding to account for out-of-bound indices, the corresponding fed inputs occasionally came into the network as blank images. Furthermore, the small batch sizes (2-3) may have exacerbated the decreases in performance and hindered learning from occasional blank input images, which can explain the previous suboptimal capsule network performances in **Table 1**.

The SOTA U-Net stil performed better than the U-CapsNet overall with a dice coefficient of **0.89**. However, the U-CapsNet closed the gap a bit more (from 0.1 difference to only 0.06 difference in dice scores). In addition, compared to the Baseline U-Net with only 4.4 million hyperparameters, the U-CapsNet was able to actually outperform it (0.83 > 0.75 dice). Therefore, although the network was not able to meet state-of-the-art performance, the U-CapsNet definitely has potential considering how it only has about 4.5 million hyperparameters.

Since the random sampling results are not complete yet, we cannot say that U-CapsNet is the state-of-the-art in dealing with class imbalances for left atrium segmentation. However, the higher dice score in comparison to the Baseline U-Net and how it only deviated 0.1 dice rather than 0.32 like in **Table 1.** highlights the capsule networks' robustness to class imbalances and reiterates their potential for this task, if they can be scaled up.

| *Architectures* | *Parameters* | *Balanced Sampling* | *Random Sampling* |
|---|---|---|---|
| U-Net (SOTA) | 27,671,926 | **0.89** | — |
| U-Net (Baseline) | 4,434,385 | 0.75 | 0.58 |
| SegCaps | 1,416,112 | 0.49 | — |
| U-CapsNet | 4,542,400 | 0.83 | **0.73** |

**Table 2:** Networks Performance on larger patches of size 256 by 320. Dice coefficients were evaluated on a separate test set from the training and validation sets. The experiments for random-sampled training for the state-of-the-art U-Net and SegCaps have not been done yet; Hence, the dice scores were omitted for those rows.

# 4 Future Work

**3D Implementations** As previously mentioned, medical images are three dimensional in nature. However, the architectures used in this study only focused on capturing two dimensional information because of memory constraints. Moreover, if these implementations are extended to three dimensions, capsule networks especially may perform better because there is another dimension of contextual information for them to learn and infer upon.

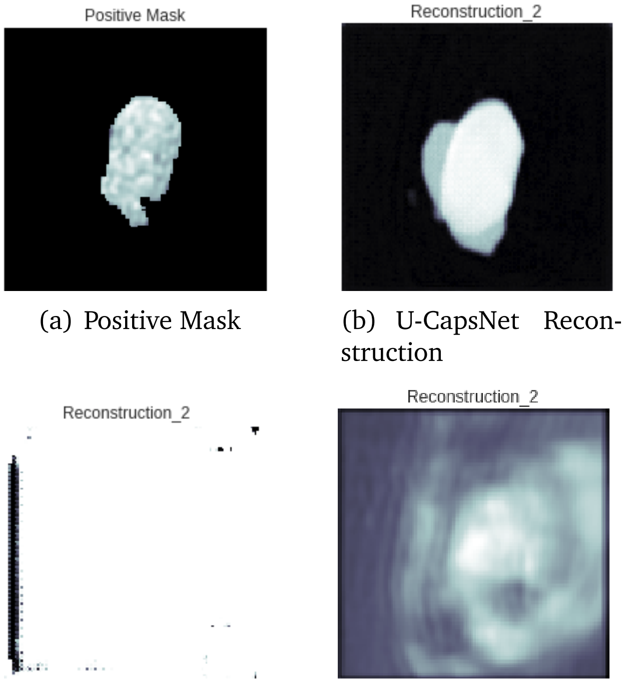
**More Data Augmentation** Data augmentation techniques other than random elastic deformations, random scaling, flips, and random rotations, could be tested, such as shears, random erasing, noise and color augmentation. However, for noise and color augmentations (i.e. Rician and Gaussian Noise and Gamma Correction), experiments illustrated that they often had detrimental effects on neural network performance and were, thereby, omitted. As such, additional experimentation and hyperparameter tuning must be done their parameters to make them useful (elaboration in **Section 6.1.1**).

**More Complex Feature Extractor** The feature extractor used in this study was only a simple 2D U-Net. This could be improved with the use of residual connections, atrous convolutions, or even sub-pixel upsampling in the place of transposed convolutions to further supplement the capsule network in the U-CapsNet.

**Larger Batch Sizes** The capsule networks were trained with batch sizes of only 2-3 compared to that of the state-of-the-art CNN with a batch size of 17. By obtaining larger computation resources and testing with larger batch sizes, the capsule networks could potentially perform better because they could have a stabler gradient during training and converge later.

**More Folds/Trials** Since the left atrium dataset used is inherently a high variability segmentation dataset, conducting more trials and computing confidence intervals will provide more accurate and comprehensive results.

**Different Datasets** Larger datasets, such as the BRaTs Brain Tumour Segmentation dataset, could be used to allow for increased diversity and for full image training. Additionally, capsule networks could also be tested on multi-class segmentation datasets, such as the aforementioned BRaTs dataset and the liver and cancer



(a) Positive Mask

(b) U-CapsNet Reconstruction

(c) SegCaps Reconstruction

(d) Basic Capsule Network Reconstruction

**Figure 3:** Examples of reconstructions of the original image's left atrium from the test capsule networks from a 128 x 128 patch.

segmentation task from the Medical Segmentation Decathlon challenge.

# 5 Conclusion

A new neural network architecture, U-CapsNet, was proposed to be the first capsule network architecture for left atrium segmentation with cardiac MRI scans. It was tested against two other capsule network architectures (Sabour et al., 2018; Lalonde and Bagci, 2018) and a state-of-the-art 2D U-Net. Although it performed worse than the U-Net with a dice score of 0.78 v. 0.88, it held up better than the network when strictly using random sampling. Additionally, when the networks were trained with a larger patch size (256 by 320), the U-CapsNet outperformed a new baseline 2D U-Net with similar complexity for both cases of balanced and random sampled training, which illustrates their proficiency in their specific network complexity level. Therefore, although this study highlights how capsule networks are not state-of-the-art architectures, it does underpin the potential they have because of their increased robustness to class imbalances in comparison to convolutional neural networks and their high performance for their low number of hyperparameters.

# References

Cardoso, Simpson, Ronneberger, Menze, van Ginneken, Landman, Litjens, Farahani, Summers, Maier-Hein, Kopp-Schneider, Bakas, Antonelli (2018). Medical segmentation decathlon. `http://medicaldecathlon.com/`. pages

Isensee, F., Jaeger, P., Full, P. M., Wolf, I., Engelhardt, S., and Maier-Hein, K. H. (2017). Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features. *CoRR*, abs/1707.00587. pages

Jiménez-Sánchez, A., Albarqouni, S., and Mateus, D. (2018). Capsule Networks against Medical Imaging Data Challenges. *ArXiv e-prints*, page arXiv:1807.07559. pages

LaLonde, R. and Bagci, U. (2018). Capsules for object segmentation. *arXiv preprint arXiv:1804.04241*. pages

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. *arXiv preprint arXiv:1606.04797*. pages

Njeh, C. (2008). Tumor delineation: The weakest link in the search for accuracy in radiotherapy. *Journal of medical physics/Association of Medical Physicists of India*, 33(4):136. pages

Sabour, S., Frosst, N., and Hinton, G. E. (2017). Dynamic routing between capsules. *CoRR*, abs/1710.09829. pages

Sharma, N. and Aggarwal, L. M. (2010). Automated medical image segmentation techniques. *Journal of medical physics/Association of Medical Physicists of India*, 35(1):3. pages

Suetens, P., Bellon, E., Vandermeulen, D., Smet, M., Marchal, G., Nuyts, J., and Mortelmans, L. (1993). Image segmentation: methods and applications in diagnostic radiology and nuclear medicine. *European journal of radiology*, 17(1):14–21. pages

van Herk, M. (2004). Errors and margins in radiotherapy. *Seminars in Radiation Oncology*, 14(1):52 – 64. pages

Vorwerk, H., Beckmann, G., Bremer, M., Degen, M., Dietl, B., Fietkau, R., Gsänger, T., Hermann, R. M., Herrmann, M. K. A., Höller, U., et al. (2009). The delineation of target volumes for radiotherapy of lung cancer patients. *Radiotherapy and Oncology*, 91(3):455–460. pages

World Health Organization (2018). The top 10 causes of death. `https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death`. pages

Zreik, M., Leiner, T., de Vos, B. D., van Hamersvelt, R. W., Viergever, M. A., and Isgum, I. (2017). Automatic segmentation of the left ventricle in cardiac CT angiography using convolutional neural network. *CoRR*, abs/1704.05698. pages

| *Probabilities* | *Precision* | *Recall* | *Dice* |
|---|---|---|---|
| (0.5, 0.7, 0.3, 0, 0) | 0.47 | 0.61 | 0.53 |
| (0.53, 0.47, 0.08, 0, 0) | 0.67 | 0.82 | 0.74 |
| (0.53, 0.47, 0.08, 0.95, 0.35) | 0.51 | 0.56 | 0.53 |
| (0.1, 0.1, 0.1, 0.95, 0.35) | 0.74 | 0.65 | 0.69 |
| (0.1, 0.1, 0.1, 0, 0) | 0.83 | 0.83 | **0.83** |

**Table 3:** *Various metrics for each associated set of probabilities ($p_{elastic}$, $p_{scaling}$, $p_{rotation}$, $p_{rician}$, $p_{gamma}$) for the U-CapsNet.*

# 6 Appendix

This section is intended for those seeking to further improve upon and successfully implement capsule networks for medical image segmentation

## 6.1 Data Augmentation

The overall impact of data augmentation parameters, particularly the probability as illustrated in **Table 3**, was not mathematically formalized in this study. However, from **Table 3**, the general trend is that decreasing the amount of data augmentation per batch or the probabilities for the occurrence of each augmentation, $p$, led to increases in performance. The most likely reason behind this is that the batch size was small, so when there was higher probabilities, the capsule networks barely had the chance to actual learn from an unaugmented training set. Additionally, since the left atrium was often located near the center of the cardiac MRI's, they also benefited from slight overfitting (similar to that of self-driving car computer vision models), which would have been undermined with more data augmentation.

Surprisingly, the addition of noise and gamma correction led to decreases in performance across the board in comparison to when they were excluded. This could have been due to the specific fold that I trained and tested on where the test set may not have matched the distribution that the two augmentations created. It could have also been just due to the capsule network not having enough complexity and the regularization from data augmentation caused it to underfit.

## 6.2 The Importance of Mindful Preprocessing

For left atrium segmentation, the input-output (IO) heavily influenced the overall performance the networks. It was essentially the difference between converging to 0.05 and converging to 0.53. Here are some notable best practices (employed in this study, but not mentioned):

- *Resample to the median voxel spacing*: In doing so, you have the least amount of padding as possible and you have isotropic spacing. The result is that you get less blank patches. For 2D applications, isotropic spacing is not that important, but the voxel spacing should be consistent throughout each slice (the z-axis can be anisotropic).

- *Avoid excessive intensity clipping*: Since most left atrium pixels and surrounding structures are high intensity, clipping the intensities too much (i.e. in the percentiles (0.1, 0.9)) would make the learning process in differentiating the left atrium from other structures much more difficult for the neural networks.
- *Allow for overlap in patch extraction*: When the overlap was set to zero for random patch extraction, the occurrence of blank or majority black images was much higher than when the overlaps were set to greater than or equal to half the patch size.