Analyzing the effect of additional AGs (Attention Gates) in U-Net for knee recess distention ultrasound segmentation

Jingyu Ye

University of Toronto St. George

STA299Y1 Y (LEC 0301)

Professor Pascal Tyrrell

Abstract

In the medical domain, the convolutional neural network (CNN) is the dominant choice for image segmentation due to its excellent representational power, fast inference, and filter-sharing properties. U-Net is a commonly used architecture because of its accurate performance and efficient GPU memory usage. However, recent studies have proposed that encoding attention gates to standard U-Net can improve its accuracy in a large variety of datasets without a significant cost of computational resources. Attention U-Net is proposed by Oktay and colleagues in 2018 which incorporates additive soft attention in standard U-Net. This paper explores the effect of attention gates in U-Net for knee recess distention segmentation with the aim to improve segmentation accuracy of ultrasound images by highlighting the salient areas while trimming unnecessary information. 3750 ultrasound images of the knee joint for patients with recess distention were used for training, validating, and testing both models. Although it is hoped in our hypothesis that the attention gates would bring more robust improvement, we only see a 1.3105% improvement on dice coefficient and 1.7431% on IOU for attention U-Net. Moreover, attention U-Net showed smoother and more complete segmentation which may be explained by the benefit of attention gates in obtaining global information features of recess distention and suppressing irrelevant background noise.

Introduction

Premise: The deep learning task—Image segmentation—is the process of extracting the desired object from an image (Pham et al., 2000). In the medical domain, the convolutional neural network (CNN) is the dominant choice for image segmentation due to its excellent representational power, fast inference, and filter sharing properties (Oktay et al., 2018). U-Net is a commonly used architecture because of its accurate performance and efficient GPU memory usage (Ronneberger et al., 2015). However, recent studies have proposed that adding attention gates to standard U-Net can improve its accuracy in a large variety of dataset without a significant cost of computational resources.

In image segmentation tasks, attention serves the role of highlighting information useful for prediction during training while ignoring unnecessary features (Siddique et al., 2021). Thus, the computational resources spent on irrelevant information would be reduced, resulting in a network with better generalization power. There are two forms of attention, one is soft attention whereas the other is hard attention. Soft attention places weights on different patches of the image to determine relevance: higher weights are multiplied to parts of high relevance and lower weights are multiplied to less relevant areas, these weights can be applied to multiple patches at a time (Schlemper et al., 2018). During training, the weights also get trained such that the model would pay more attention to relevant regions. On the other hand, hard attention highlights the relevant features by cropping the image. It can only pay attention to one region of the image at a time.

Attention U-Net is proposed by Oktay and colleagues (2018) which incorporates additive soft attention in standard U-Net. Soft attention is implemented at the skip connections which suppresses activation in less relevant areas. This helps reduce the number of redundant low-level feature extractions brought by skip-connection in standard U-Net. Oktay and colleagues (2018) has shown that Attention U-Net outperforms standard U-Net on CT pancreas segmentation (a challenging task where there is low tissue contrast and large variability in organ size and shape) while preserving computational efficiency.

Keeping this in mind, this paper explores the effect of attention gates in U-Net for knee recess distention segmentation because b-mode ultrasound images often have a large amount of noise, distortion, and shadow which causes some blurred local details and lots of dark areas and no obvious division (Jin & Long, 2018). Moreover, the recess distention area detected by ultrasonic signal is similar to the image background resulting in high rates of false identification. Thus, it is proposed that the additional attention gates in Attention U-Net, which can highlight the salient areas while trimming unnecessary information, would improve segmentation accuracy on knee recess distention ultrasound dataset.

Purpose: To compare the performance of Attention U-Net and Traditional U-Net on knee recess distention segmentation.

Research Question: How would the implementation of AG (Attention Gates) in a traditional U-Net (Attention U-Net) affect its segmentation performance on knee recess distention ultrasound dataset?

Hypothesis: The additive attention gate in standard U-Net architecture would improve segmentation performance on knee recess distention ultrasound dataset. Attention U-Net would outperform Traditional U-Net on knee recess distention segmentation.

Objectives:

1. To implement and describe Attention U-Net for knee recess distention ultrasound dataset.
2. To compare the performance of Attention U-net and Classical U-net on knee recess distention ultrasound dataset.
3. To propose and recommend when the addition of attentional layers would be beneficial.

Methodology

Dataset: The dataset comprises 3750 b-mode ultrasound images of the knee joint. These ultrasound images were collected electronically from various private independent healthcare facilities and are mostly from adult patients (estimated mean age of 50 years) who had a diagnostic ultrasound ordered between January 1, 2010 and December 31, 2018. Knee scans were performed by an undisclosed number of ultrasound devices and sonographers. The data were anonymized and encrypted before being sent to the lab archive; all patient demographics were deleted from the data in compliance with the center's rule. The collected data consists of an examination report and a clinical decision that states whether the patient presented a recess distention or normal joint. Natural language processing was used to process the examination reports for the identification of negative and positive examinations. The positive examinations were then analyzed by an expert to determine true positives and false positives. For the purpose of this study, only those images that showed recess distention were used for segmentation. To reduce the computational costs, the original images and masks are resized into equal sizes of 256*256 resolution.

Training environment and statistics: Both Attention U-Net and standard U-Net were trained for 200 epochs with a batch size of 30 on a twined NVIDIA GeForce GTX 1080Ti GPU. 566 images from the knee recess distention ultrasound dataset were held out for testing the model. The training dataset was then split into a 70:30 ratio for training and validation. Adam optimizer with a learning rate of 0.001 was used as an optimizer, Intersection over Union (IOU) loss was used as loss function, Dice coefficient and IOU coefficient were used as metrics for evaluation. The dice coefficient compares the pixel-wise agreement between the segmented region and the ground truth. The higher the dice coefficient, the better the segmentation result. The IOU coefficient is similar to the dice coefficient, it is the overlap rate between the predicted segmentation and the ground truth. The two metrics both range from 0 to 1 with 0 meaning there is no overlap between the predicted segmentation and the ground truth and 1 meaning there is a perfect overlap. Mean and standard deviation were used to compare the % coverage difference from ground truth for Attention U-Net and standard U-Net, they were also used to compare the number of pieces for segmentation by Attention U-Net and standard U-Net.

Results

The performance of the two models are evaluated to prove their feasibility on the knee recess distention dataset. Figures 1 and 4 show the learning curve of dice coefficients for both models. Both models have a low training and validation dice coefficient at the beginning which gradually increases

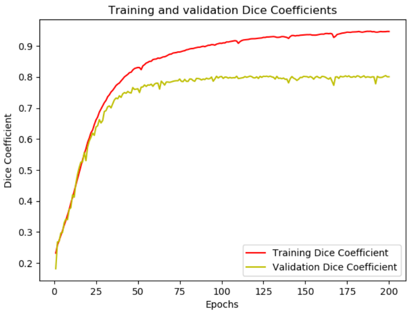


Fig 1. Attention U-Net Training and Validation Dice coefficient

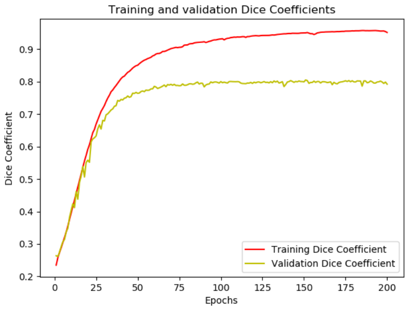


Fig 4. U-Net Training and Validation Dice coefficient

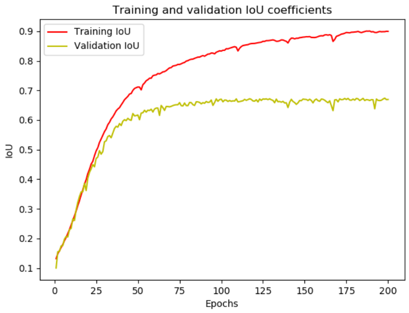


Fig 2. Attention U-Net Training and Validation IOU coefficient

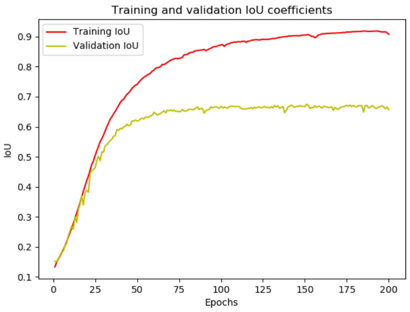


Fig 5. U-Net Training and Validation IOU coefficient

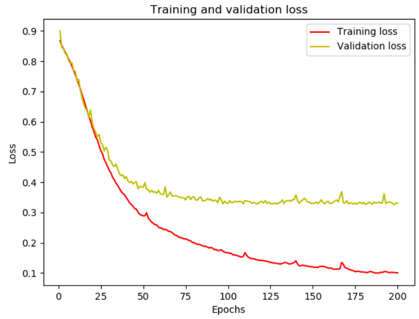


Fig 3. Attention U-Net Training and Validation IOU loss

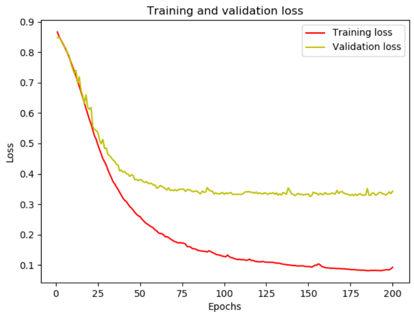


Fig 6. U-Net Training and Validation IOU loss

upon adding more epochs and flattens gradually indicating the addition of more epochs won't improve the models' performance on training data or unseen data. It is also observed that the dice coefficients of both models will almost always be lower for the validation dataset than for the training dataset. This gap between the models' performance in training and validation dataset is referred to as the generalization gap. Similar behavior is observed for both models' learning curves of the IOU coefficient (Figures 2 and 5). On the other hand, training and validation loss for both models (Figures 3 and 6) are high at the beginning which steadily decreases with the increase of epochs and flattens gradually denoting additional epochs won't improve the models' performance on training data or unseen data. From the above behaviors, it can be concluded that both Attention U-Net and standard U-Net trained for 200 epochs were neither underfitting nor overfitting and the models are both training well and generalizing well.

The trained models were evaluated on the 566 knee recess distention testing images in terms of mean dice coefficient and mean IOU. The mean Dice coefficient for Attention U-Net is 0.7817 and 0.7686 for standard U-Net. The mean IOU for Attention U-Net is 0.6430 and 0.6256 for standard U-Net.

Given the small difference in mean dice coefficient and mean IOU between Attention U-Net and standard U-Net, more analysis is needed to compare the performance of Attention U-Net and standard U-Net on segmenting knee recess distention. Thus, the predicted segmentation for Attention U-Net and standard U-Net is overlaid with the ground truth on the ultrasound image for all testing images. From the overlaid results, it is observed that both Attention U-Net and standard U-Net tend to over segment the recess distention area, that is, the area of the predicted segmentation for both models is often larger than the area of the ground truth. To compare the relative areas of the predicted segmentation and the ground truth, the percentage area covered by each segmentation and ground truth is calculated. On average, the difference between the percentage area covered by Attention U-Net segmentation and the ground truth is 1.3258% with a standard deviation of 2.5610%, indicating that Attention U-Net over segments the knee recess distention area by 1.3258% on average. The mean difference between the percentage area covered by standard U-Net segmentation and the ground truth is 1.0863% with a standard deviation of 2.5606%, reflecting that standard U-Net over segments the knee recess distention area by 1.0863% on average.

From looking at the overlaid predicted segmentation and the ground truth, it is also seen that while the ground truth is always in one piece, the predicted segmentation by both models is sometimes in multiple fragments. To investigate how this differs between Attention U-Net and standard U-Net, the average number of fragments for both models' predicted segmentation was calculated. On average, the segmentation by Attention U-Net consists of 1.1696 pieces with a standard deviation of 0.4560. For standard U-Net, the segmentation consists of 1.5088 pieces on average with a standard deviation of 0.9598.

Among the predicted segmentations by Attention U-Net and standard U-Net, 67.6678% of the segmentation by standard U-Net are in one piece whereas 80.0353% of the segmentation by Attention U-Net are in one piece. Moreover, 0.0070% of the predicted segmentation by standard U-Net failed to segment recess distention when it is presented but very small whereas 0.0212% of the predicted segmentation by Attention U-Net failed to segment recess distention when it is presented.

Discussion

From the result obtained, it is evident that Attention U-Net achieved a higher dice coefficient and IOU than standard U-Net on knee recess distention segmentation, although only by 1.3105% on dice coefficient and 1.7431% on IOU. Thus, further research is needed to determine whether the improvement is statistically significant.

Looking closer into each segmentation by standard U-Net and Attention U-Net and analyzing the statistics obtained from the predicted segmentations and the ground truths, it can be seen that both models tend to show increasing segmentation area, where there is a very small difference for the mean percentage coverage difference from the ground truth between Attention U-Net and standard U-Net (Attention U-Net showed more coverage of 0.2395% than standard U-Net). A possible reason for this is there is a higher uncertainty of annotations around the boundary recess distention. Although the knee recess distention dataset is annotated by medical experts, the annotations are not perfect for reproducibility because as the annotations were done manually, sometimes the annotation may be bigger than the region of interest whereas other times the annotation may be smaller. Consequently, there are much more uncertainties around the boundaries compared to the center of the targeted region. Therefore, both segmentation models generally are not so good at segmenting the boundary of the recess distention area and tend to segment more areas than expected around the boundaries. This reason may also account for the large variability in percentage coverage difference from the ground truth for both models (2.5610% for Attention U-Net and 2.5606% for standard U-Net).

Attention U-Net and U-Net also differ in the number of pieces each model tends to segment the recess distention area. The expected number of pieces for the knee recess distention segmentation should be one complete piece. However, only 67.6678% of the predicted segmentations by standard U-Net are one piece and 80.0353% of the segmentation by Attention U-Net are one piece. On average, the number of pieces for U-Net segmentation (1.5088) is also higher than Attention U-Net (1.1696). From the predicted segmentations, it is observed that standard U-Net showed less continuity in its segmentation (often broken down into multiple pieces) whereas Attention U-Net showed smoother and more complete segmentation. The additional attention gates implemented may be beneficial here. With the introduction of the attention mechanism, the global information features of the recess distention dataset is obtained and the irrelevant background noise is suppressed which in turn increases the model’s sensitivity and lends to smoother and more complete segmentation for Attention U-Net (Oktay et al., 2018).

Without requiring a large number of model parameters, Attention U-Net showed a slight improvement in knee recess distention segmentation than standard U-Net. Therefore, these results do suggest that future research in Attention Gates and medical imaging is worthwhile.

Limitations: The training of both Attention U-Net and standard U-Net was done on the twined NVIDIA GeForce GTX 1080Ti GPU from the remote lab server, the server is shared among students and researchers in Dr. Tyrrell's lab. Therefore, its memory availability is different when training for the two models resulting in incomparable computation time. Moreover, the experiment above was limited in computation and time. Thus, it is not possible to obtain optimal performance for Attention U-Net and standard U-Net by tuning hyperparameters. Another limitation is the single training and validation split that is used for training both models due to limited computational power and time. A single training and validation split may generate incorrect results if the random split is done inappropriately because it is fully dependent on one split. On the other hand, cross-validation would be a better option in future studies where the model is given an opportunity to train on different subsets resulting in a better indication of how well the model will perform on unseen data (Yadav & Shukla, 2016). The results from this study were meant to serve as a basis, future replications and additional trials would increase the confidence in the results.

Conclusion

To answer the research question, "How would the implementation of AG (Attention Gates) in a traditional U-Net (Attention U-Net) affect its segmentation performance on knee recess distention ultrasound dataset?", compared to standard U-Net we see that Attention U-Net achieved 1.3105% improvement on dice coefficient and 1.7431% on IOU. Although it is hoped that the attention gates would bring more robust improvement as stated in the hypothesis, we only see very little improvement and could not conclude whether it is significant.

Future work could improve upon the uncertainty with manual annotation of medical images as it may seem to influence the performance of both segmentation models. For instance, self-supervised could be leveraged with attention to build a reinforcement learning agent that learns representation guided by understanding what is salient in a scene for sequential decision making

(Wu et al., 2021). Hence, we do not need explicit information in the form of annotated datasets which not only reduces time expense and manual labor but also improves the problem with higher uncertainties of annotations around the boundary of targeted regions.

References

Jin, N., & Long, Z. (2018). Effusion area segmentation for knee joint ultrasound image based on atrous-FCN with snake model algorithm. *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. https://doi.org/10.1109/cisp-bmei.2018.8633127

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018, May 20). *Attention U-net: Learning where to look for the pancreas*. arXiv.org. Retrieved August 26, 2022, from https://arxiv.org/abs/1804.03999

Pham, D. L., Xu, C., & Prince, J. L. (2000). Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, *2*(1), 315–337. https://doi.org/10.1146/annurev.bioeng.2.1.315

Ronneberger, O., Fischer, P., & Brox, T. (2015, May 18). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. arXiv.org. Retrieved August 26, 2022, from https://arxiv.org/abs/1505.04597

Schlemper, J., Oktay, O., Chen, L., Matthew, J., Knight, C., Kainz, B., Glocker, B., & Rueckert, D. (2018, April 15). *Attention-gated networks for improving ultrasound scan plane detection*. arXiv.org. Retrieved August 26, 2022, from https://arxiv.org/abs/1804.05338

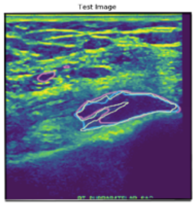
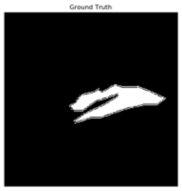
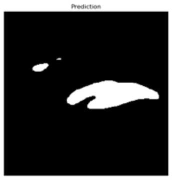
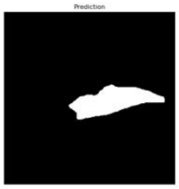
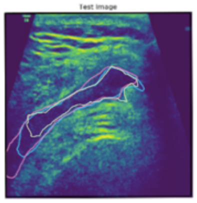
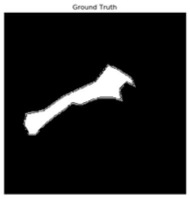
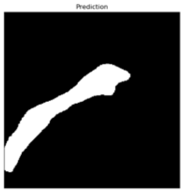
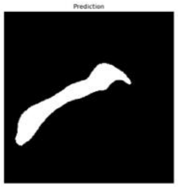
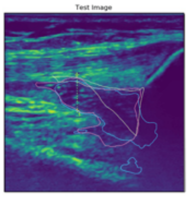
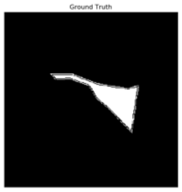
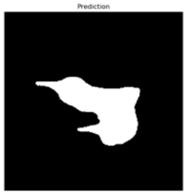
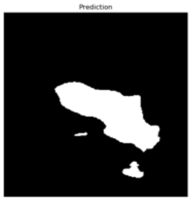
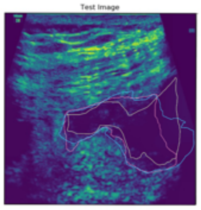
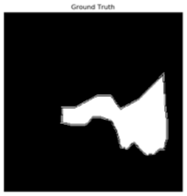
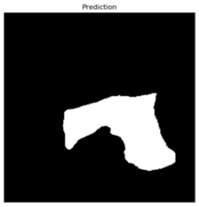
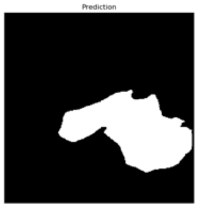
Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-Net and its variants for Medical Image Segmentation: A review of theory and applications. *IEEE Access*, *9*, 82031–82057. https://doi.org/10.1109/access.2021.3086020

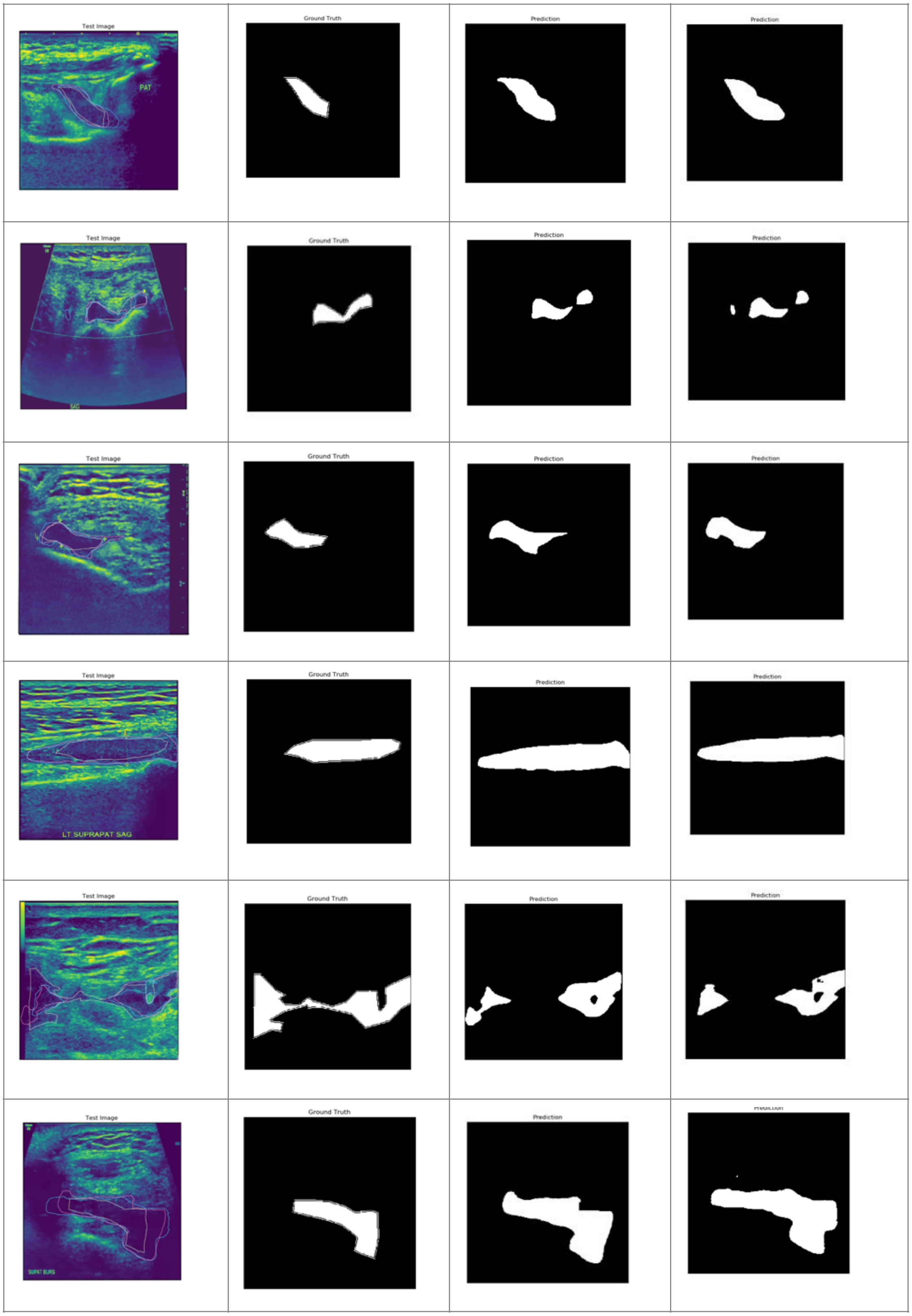
Wu, H., Khetarpal, K., & Precup, D. (2021). Self-Supervised Attention-Aware Reinforcement Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(12), 10311-10319. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/17235

Yadav, S., & Shukla, S. (2016). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for Quality Classification. *2016 IEEE 6th International Conference on Advanced Computing (IACC)*. https://doi.org/10.1109/iacc.2016.25

Appendix

Comparison of knee recess distention segmentation:

Overlaid segmentations	Ground Truth (yellow tracing)	Attention U-Net segmentation (pink tracing)	U-Net segmentation (blue tracing)
			
			
			
			


Test Image
PAT
Ground Truth
Prediction
Prediction
Test Image
SAG
Ground Truth
Prediction
Prediction
Test Image
Ground Truth
Prediction
Prediction
Test Image
LT SUPRAPAT SAG
Ground Truth
Prediction
Prediction
Test Image
Ground Truth
Prediction
Prediction
Test Image
SUPAT BURS
Ground Truth
Prediction
Prediction