Attention is All You Knee

Shayne Miel
Stanford University
Stanford, CA, USA
smiel@stanford.edu

Abstract

Improving the accuracy with which automated methods can identify injuries in MRIs of the knee would lead to time and cost savings for doctors and patients. I show that by fine-tuning a pretrained neural network to extract features and then using a self-attention mechanism to combine features from multiple frames of the MRI scan, I can improve upon the current state of the art on the MRNet data set.

1. Introduction

Magnetic resonance imaging (MRI) is a method of obtaining three dimensional images of the inside of an object by using strong magnets to align the protons in the object and then radio frequency currents to disrupt and measure that alignment. An MRI sequence is a series of two dimensional images that can be stacked to recreate the three dimensional object. There are three sequence types, each corresponding to an orientation of the "camera" that captures the two dimensional slices. The axial sequence captures slices of the object that are horizontal to the ground; the coronal sequence captures vertical slices as viewed from the front of the object; and the sagittal sequence captures vertical slices as viewed from the side of the object.

MRI can be a useful tool when diagnosing knee injuries[9, 3, 10], however, analyzing the images is a time-consuming process and even with trained professionals, it is easy for a clinician to misdiagnose an injury based on an MRI reading[6]. Improving the automated identification of abnormalities in knee MRIs could help prioritize which MRIs to examine first, as well as provide better early results for patients whose scans appear normal. Model predictions could also provide a "second opinion" which would reduce the possibility of missed abnormalities. This could represent a large cost savings for hospitals and an increased level of care for patients.

Given a set of three MRI series (axial, coronal, and sagittal) of a patient's knee, we wish to predict the presence of injuries that will require surgery. In particular, we wish to predict whether the knee is healthy, has a meniscal tear, has an ACL tear, or has any other abnormality. Since these injuries can co-occur, we wish to predict three independent binary values: abnormal, acl tear, and meniscus tear.

2. Related work

Talk about the MRNet architecture. Maybe include a diagram. At least half a page. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

The main purpose of this study is to replicate the results achieved by Bien, et al's MRNet model [1]. They use a pretrained AlexNet [8] model to extract features from each 2D slice of the 3D MRI volume, followed by a global average pooling per slice to flatten the image, and a max pooling across the volume. Finally, a fully connected layer and sigmoid activation are used to predict a binary label for each series. Those three predictions (axial, coronal, and sagittal) are then used as features in a simple logistic regression to predict the final label in question. This process is repeated for each of the three independent labels.

Talk about SqueezeNet. Talk about ways that it differes from AlexNet. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Talk about self-attention. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Talk about other work on MRIs. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

3. Methods

My goals for this project are fairly simple:

1. Reproduce the results obtained by Bien, et al. in [1].

2. Experiment with using different pretrained networks to extract features, specifically replacing AlexNet with GoogLeNet [11] as Chi, et al. did for ultrasound images [2], ResNet [4] as done in [7], and Inception-v3 [12] as done in [5].

3. If time allows, try replacing the logistic regression ensemble with a direct concatenation of the features from each of the three series before going through the fully connected layer.

Explain that all experiments can be described as modifications to MRNet. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Replace AlexNet with SqueezeNet. Math Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Replace max pooling across the series with attention. Math Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Replace logistic regression with end-to-end ensembling. Math Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

All experiments use the same loss function as in MRNet. Math. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a

Diagnosis	Label	Train	Validation	Test
Abnormal	Positive	817	96	95
	Negative	193	24	25
	Total	1010	120	120
ACL	Positive	193	15	54
	Negative	817	105	66
	Total	1010	120	120
Meniscus	Positive	357	40	52
	Negative	653	80	68
	Total	1010	120	120

Table 1. MRNet data splits and label counts.

aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

4. Dataset

The MRI data provided in the MRNet challenge contains scans from 3 MRI types (sagittal plane T2, coronal plane T1, and axial plane PD) with 3 labels per MRI (abnormality, ACL tear, and meniscal tear) for 1,250 examinations.

An important aspect of the MRNet model is the data augmentation done during training. Every volume is randomly flipped horizontally, shifted horizontally by -25 to 25 pixels, and rotated by -25 to 25 degrees each time it is seen during training. This helps prevent the model from overfitting the small data set.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

They have designated a training/validation split and have withheld the test set for leaderboard evaluation. In order to assess the changes I will be making, I am calling their validation set the test set, and splitting their training set into a training and validation set. Counts of cases and labels for each set can be seen in Table 1.

Note that this data has already been preprocessed as described in [1]:

> Images were extracted from Digital Imaging and Communications in Medicine (DICOM) files, scaled to 256×256 pixels, and converted to Portable Network Graphics (PNG) format using the Python programming language (version 2.7) and the pydicom library (version 0.9.9).
>
> To account for variable pixel intensity scales within the MRI series, a histogram-based intensity standardization algorithm was applied to the images. For each series, a representative intensity distribution was learned from the training set exams. Then, the parameters of this distribution were used to adjust the pixel intensities of exams in all datasets (training, tuning, and validation). Under this transformation, pixels with similar values correspond to similar tissue types. After intensity standardization, pixel values were clipped between 0 and 255, the standard range for PNG images.

5. Experiments/Results/Discussions

Hyperparameters and Adam - because MRNet paper. Quick experiemnts showed 1e-5 were a good learning rate. Quick experiments showed all learning was done by 40 epochs Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Primary metrics: - Average AUC over three diagnoses

For each diagnosis: - AUC - Specificity - Sensitivity - Accuracy

I have been able to successfully recreate the data loading and training pipeline, as well as the data augmentation steps, ensembling, and evaluation code. The hyperparameters and training process are the same for every model. It is unclear whether the current hyperparameters are the optimal ones, but the results look promising. Figure 1 shows the loss curves from the baseline MRNet model for training and validation on each series and diagnosis.

The AUC values for training and validation at each epoch can be seen in Figure 2. The models appear to be learning and generalizing well. The validation scores are noisy though, which is most likely due to the extremely small data set sizes and imbalanced classes.

Table 2 shows the AUC on the test set as reported in [1] and the test set (their validation set) with my reproduction of the MRNet model. I do not have access to the test set that they used for reporting results, so I do not expect to get exactly the same results. The AUC is fairly close for all three diagnoses. My scores are slightly lower on ACL and Meniscus tears, which is most likely due to the smaller data set size and non-optimal hyperparameters.

Show class activation map for MRNet Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Show 4x4 class activation map Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Show cross-method accuracy heatmap Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

6. Conclusion/Future Work

In this paper we... Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

The best performing model was... Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

Future work should include: - experimenting with more complex pretrained networks - using cross-series attention Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed magna non nulla tincidunt suscipit. Cras semper, sapien vitae maximus viverra, arcu ex efficitur massa, sodales finibus sem lorem id erat. Cras scelerisque, erat eu dignissim finibus, ipsum lacus tincidunt leo, nec fringilla nunc nisl eget diam. Nulla vitae interdum dolor. Vestibulum non finibus mauris, a ornare libero. Etiam a aliquet neque, ut dictum ante. Mauris ex massa, sagittis a viverra sed, lacinia vitae ante. Cras id arcu viverra turpis cursus porttitor in et ligula.

7. Contributions & Acknowledgements

Compute power generously donated by Turnitin.

References

[1] N. Bien, P. Rajpurkar, R. L. Ball, J. Irvin, A. Park, E. Jones, M. Bereket, B. N. Patel, K. W. Yeom, K. Shpanskaya, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of mrnet. *PLoS medicine*, 15(11):e1002699, 2018.

[2] J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, and M. Eramian. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *Journal of digital imaging*, 30(4):477–486, 2017.

[3] S. Figueiredo, L. S. Castelo, A. D. Pereira, L. Machado, J. A. Silva, and A. Sa. Use of mri by radiologists and orthopaedic surgeons to detect intra-articular injuries of the knee. *Revista brasileira de ortopedia*, 53(1):28–32, 2018.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] D. Kim and T. MacKinnon. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clinical radiology*, 73(5):439–445, 2018.

[6] G. Kolata. Sports medicine said to overuse m.r.i.'s, October 2011. [Online; posted 28-October-2011].

[7] P. Korfiatis, T. L. Kline, D. H. Lachance, I. F. Parney, J. C. Buckner, and B. J. Erickson. Residual deep convolutional neural network predicts mgmt methylation status. *Journal of digital imaging*, 30(5):622–628, 2017.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[9] N. Orlando Júnior, M. G. d. S. Leão, and N. H. C. d. Oliveira. Diagnosis of knee injuries: comparison of the physical examination and magnetic resonance imaging with the findings from arthroscopy. *Revista brasileira de ortopedia*, 50(6):712–719, 2015.

[10] T. Smith, M. Lewis, F. Song, A. Toms, S. Donell, and C. Hing. The diagnostic accuracy of anterior cruciate ligament rupture using magnetic resonance imaging: a meta-analysis. *European Journal of Orthopaedic Surgery & Traumatology*, 22(4):315–326, 2012.

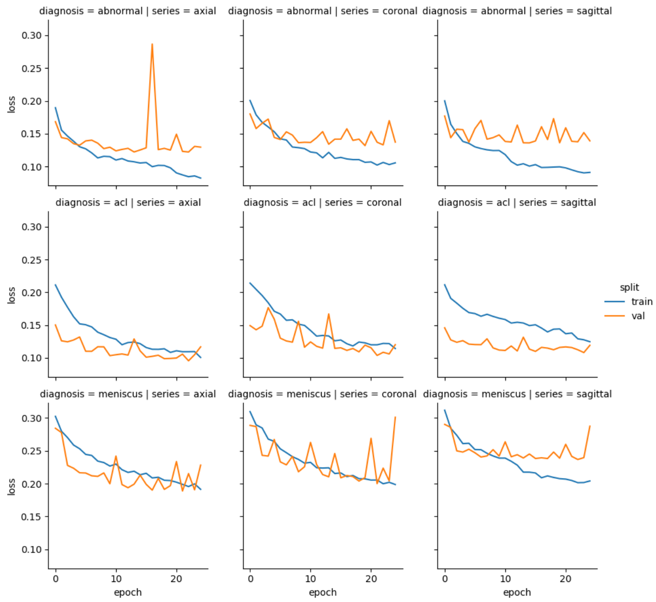

Figure 1. Baseline loss for each diagnosis and series.

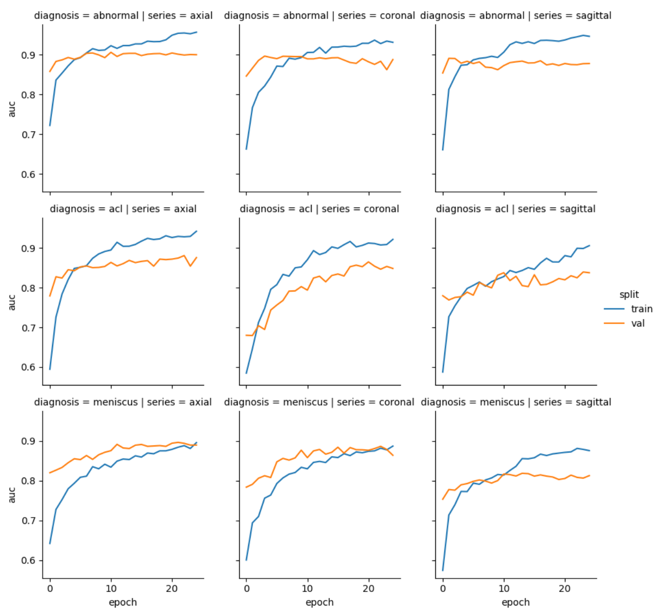

Figure 2. Baseline AUC for each diagnosis and series.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recogni-*

Model	Abnormal	ACL	Meniscus
MRNet (reported)	0.937	0.965	0.847
MRNet (milestone)	0.951	0.949	0.839

Table 2. AUC on the test set

tion, pages 1–9, 2015.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.