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 an 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[16, 7, 18], 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[11]. 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.

The problem addressed in this paper is as follows: given a set of three MRI sequences (axial, coronal, and sagittal) of a patient's knee, can we predict the presence of injuries that will require surgery? In particular, we wish to predict whether the knee is healthy, has an ACL tear, has a meniscal 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. Past research has proven that MRI scans are accurate and sensitive tools for detecting these kinds of injuries in a non-invasive manner[3, 6, 22].

2. Related work

The current state-of-the-art for detecting injuries in MRI scans of the knee is MRNet[2]. Their method uses a deep convolutional neural network to predict a binary label per-sequence, ensembled via a simple logistic regression over the per-sequence probabilities to predict the final label.

For each image in each sequence, they use a pretrained AlextNet to extract features from the image. AlexNet is a deep convolutional neural network that achieves a good accuracy on the ImageNet dataset[13]. By using a

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

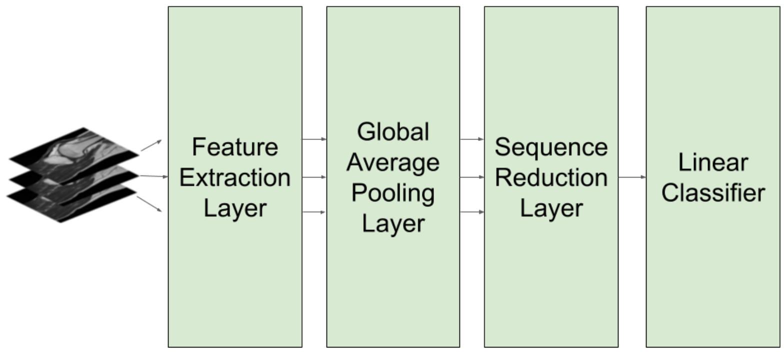

Figure 1. An abstract view of the sequence-specific networks.

achieved by Bien, et al's MRNet model [2]. They use a pretrained AlexNet [13] 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 [2].

2. Experiment with using different pretrained networks to extract features, specifically replacing AlexNet with GoogLeNet [19] as Chi, et al. did for ultrasound images [4], ResNet [8] as done in [12], and Inception-v3 [20] as done in [10].

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.

Figure 1 shows an abstract process for turning a single MRI sequence into an injury prediction. On the left side of the diagram, a series of n images, each a $\mathbb{R}^{3\times224\times224}$ matrix, are fed into the network as a single batch. The feature extractor converts each of these images into a $\mathbb{R}^{c\times w\times h}$ matrix, where c is the number of channels and w and h are the width and height, respectively. These features are then flattened into a series of $\mathbb{R}^{c}$ vectors in the global average pooling layer. Finally, the full batch of n vectors are flattened into a single $\mathbb{R}^{c}$ vector in the sequence reduction layer. This vector is then passed through a linear layer and a sigmoid activation to arrive at a probability for whether the entire MRI sequence indicates the injury in question.

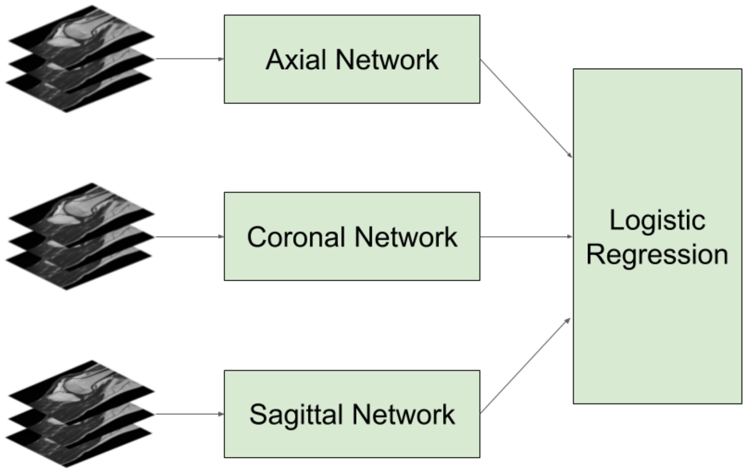

Figure 2. Ensembling predictions from the sequence-specific networks to generate an injury prediction.

Concretely, if we describe the feature extraction layer as a function, $f : \mathbb{R}^{n\times 3\times 244\times 244} \rightarrow \mathbb{R}^{n\times c\times w\times h}$, and the sequence reduction layer as a function $g : \mathbb{R}^{n\times c} \rightarrow \mathbb{R}^{c}$, then the entire network can be written as:

$$gap_{jk}(x) = \frac{\sum_{i=1}^{n} x_{j,k}^{(i)}}{n}$$

$$p = \sigma(g(gap(f(x)))) \tag{1}$$

where p is the predicted probability and σ is the sigmoid function: $\frac{1}{1+e^{-x}}$

Because we have access to multiple MRI sequences per patient, this entire process is repeated for each of the axial, coronal, and sagittal sequences, and the three probabilities are then fed into a logistic regression classifier to arrive at a final injury prediction, as shown in Figure 2.

The current state-of-the-art system, MRNet - which was described in detail in Section 2, can be described as using parts of a pretrained AlexNet for the feature extractor layer, and an element-wise maximum to flatten the sequence in the sequence reduction layer.

3.1. SqueezeNet

SqueezeNet is a deep convolutional neural network that achieves high performance on the ImageNet challenge. It is composed of a series of "fire" layers, each of which consists of a layer of 1×1 convolutional filters (the squeeze layer), followed by a ReLU and then a mix of 1×1 and 3×3 convolutional layers (the expand layer) whose outputs are concatenated before going through a final ReLU activation. An illustration of a fire layer is shown in Figure 3. These layers are interspersed with max pooling layers before the 1st, 4th, and 8th fire layers, and the entire network begins with a traditional 7×7 convolutional layer.

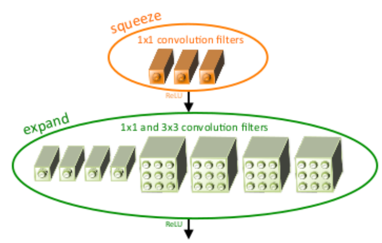

Figure 3. Illustration of a fire layer. Taken directly from "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size"[9]

Fine-tuning a SqueezeNet model that has been pretrained on the ImageNet data set is a method that has been used successfully to classify vehicles[1], crop disease[5], and cataracts[17]. By removing the final dropout, convolution, ReLU and adaptive average pooling layers from the pre-trained network, we can use it as a drop-in replacement for the feature extraction layer in Figure 1.

3.2. Attention

Attention is a mechanism that has traditionally been applied as a summarization method of the hidden states in a recurrent neural network. Given some query $\mathbf{Q}$, a set of keys $\mathbf{K}$, and a set of values $\mathbf{V}$, the attention score can be calculated as

$$s = (\mathbf{Q}\mathbf{K}^T)\mathbf{V}$$

$$\mathbf{A}_i = \frac{e^{s_i}}{\sum_{j=1}^{k} e^{s_j}}$$

as described by Tan, et al.[21], whose work was derived from Luong et. al's location-based attention[14].

In the case of an RNN, the query is often the final hidden state, the keys are either the input embeddings or the hidden states at each time step, and the values are the hidden states at each time step. A final representation of the entire sequence can then be generated by taking the weighted sum of the values times the attention vector, $result = \mathbf{A}\mathbf{V}^T$.

For the experiments in this paper, we eschew the use of an RNN and simply use the attention-weighted sum across frames of the MRI sequence to generate the final representation of the sequence. In this case, the query is a learned parameter of the network, while the keys and values are both the output of the global average pooling layer for each frame in the sequence. This allows us to use the attention-weighted sum as the sequence-reduction layer:

$$g(x) = \mathbf{A}x^T$$

Because of the imbalanced class sizes, we train all of the sequence-specific models with a weighted binary cross-entropy loss:

$$L = -\frac{1}{N}\sum_{i=0}^{N} w_i\big(y_i log(p_i) + (1-y_i) log(1-p_i)\big)$$

where w_i is the fraction of the instances in the data set that have label y_i, and p_i is the output of the network described by Equation 1 for sequence i.

4. Dataset

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

The released data has already been split into a training and validation set, and a test set has been withheld for leaderboard purposes. In order to evaluate my methods for this paper, I have further divided the training set into training and evaluation sets, and am using the validation set for all reported metrics. Counts of cases and labels for each set can be seen in Table 1.

4.1. Preprocessing

The data is provided as a collection of numpy[15] 3-dimensional matrices, one per sequence per case. They have already been extracted from the Digital Imaging and Communications in Medicine (DICOM) files and scaled to 256×256 images.

During training, we crop the images to 224×224 to match the expected input size for a pretrained ImageNet model. We then standardize each input sequence by subtracting the minimum pixel value and dividing by the range observed in the sequence, then rescale the image to a $0-256$ range of pixel values. Finally, we normalize the sequence by subtracting the data set mean (58.09) and dividing by the data set standard deviation (49.73).

Diagnosis	Label	Training	Evaluation	Validation
Abnormal	Positive	835	78	95
	Negative	175	42	25
	Total	1010	120	120
ACL	Positive	167	41	54
	Negative	843	79	66
	Total	1010	120	120
Meniscus	Positive	353	44	52
	Negative	657	76	68
	Total	1010	120	120

Table 1. MRNet data splits and label counts.

4.2. Augmentation

In order to prevent overfitting the small data set, we perform data augmentation during training. Every image 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. Note that the same transformation is applied to every image in a given sequence.

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 **??** 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 5. 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 3 shows the AUC on the test set as reported in [2] 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

This paper explores the use of pretrained SqueezeNet feature extractors and self-attention mechanisms to predict injuries from MRI sequences of knees. The best performing model

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] A. S. Agoes, Z. Hu, and N. Matsunaga. Fine tuning based squeezenet for vehicle classification. In *Proceedings of the International Conference on Advances in Image Processing*, pages 14–18. ACM, 2017.

[2] 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.

[3] B. F. Boeve, R. Davidson, and J. E. Staab. Magnetic resonance imaging in the evaluation of knee injuries. *Southern medical journal*, 84(9):1123–1127, 1991.

[4] 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.

[5] H. Durmuş, E. O. Güneş, and M. Kırcı. Disease detection on the leaves of the tomato plants by using deep learning. In *2017 6th International Conference on Agro-Geoinformatics*, pages 1–5. IEEE, 2017.

[6] L. Felli, G. Garlaschi, A. Muda, A. Tagliafico, M. Formica, A. Zanirato, and M. Alessio-Mazzola. Comparison of clinical, mri and arthroscopic assessments of chronic acl injuries, meniscal tears and cartilage defects. *Musculoskeletal surgery*, 100(3):231–238, 2016.

[7] 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.

[8] 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.

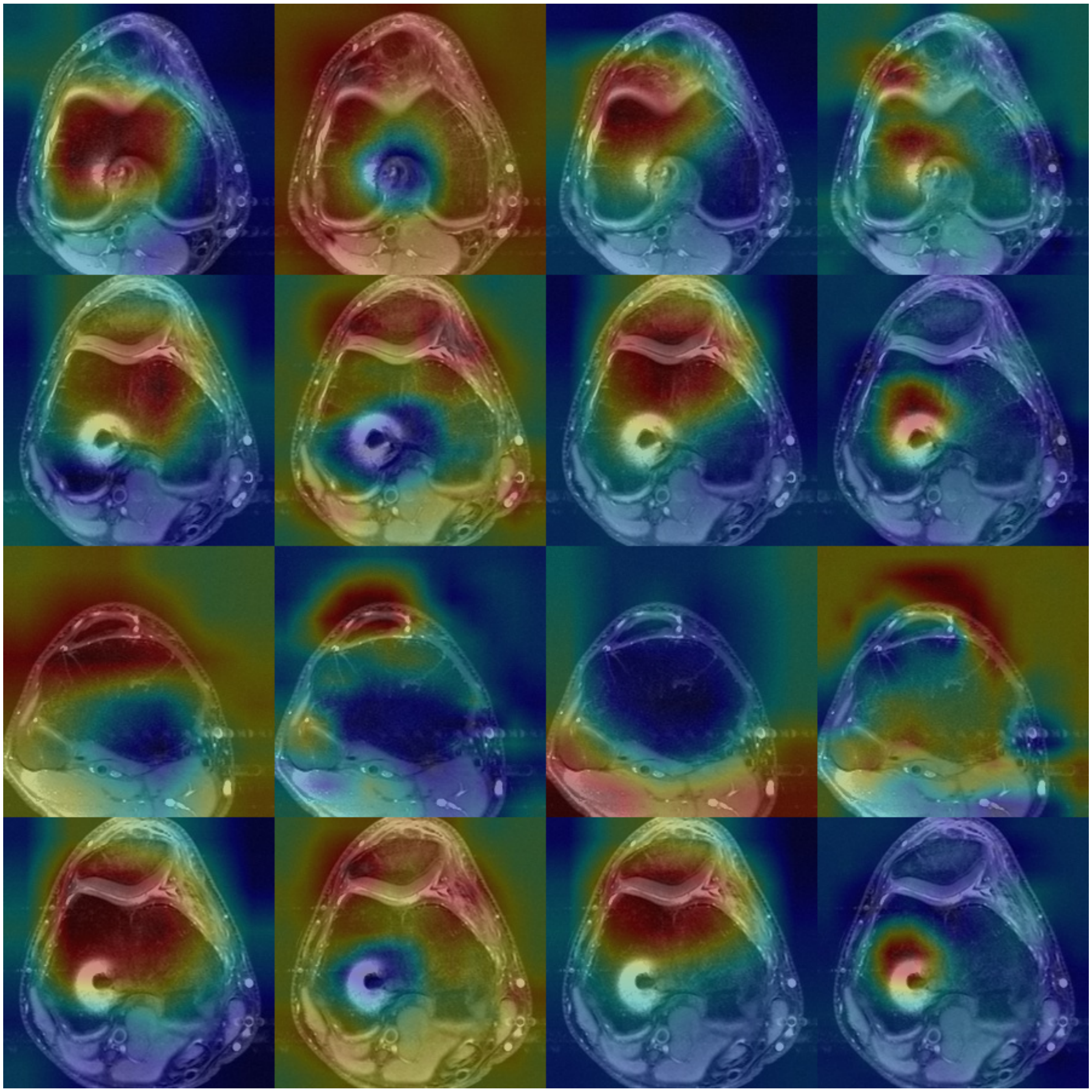

Figure 4. Class activation maps from the four networks for the axial sequence of case 1218, when asked to predict whether there is an ACL tear. From left to right, the columns are class activation maps for MRNet, MRNet-Squeeze, MRNet-Attend, and MRNet-SqueezeAttend. Each row represents the frame from the axial sequence that each network found most interesting.

Model	Abnormal	ACL	Meniscus	Average
MRNet (reported)	0.937	0.965	0.847	0.916
MRNet	**0.940**	0.960	0.839	0.913
MRNet-Squeeze	0.925	**0.974**	0.829	0.910
MRNet-Attend	0.925	0.910	0.838	0.891
MRNet-SqueezeAttend	0.936	0.925	**0.885**	**0.915**

Table 2. AUC on the validation set

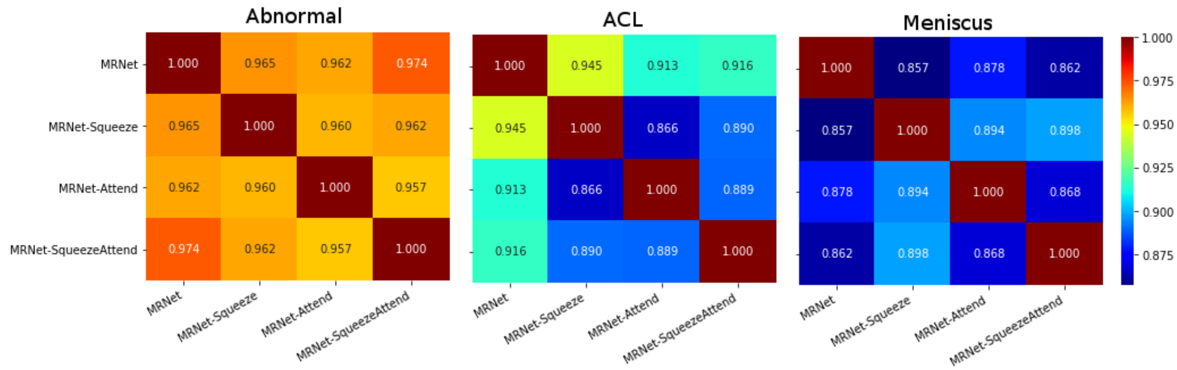

Figure 5. Correlation between predictions on the validation set.

[9] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.

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

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

[12] 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.

[13] 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.

[14] M.-T. Luong, H. Pham, and C. D. Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.

[15] T. E. Oliphant. A guide to numpy, 2006–. [Online; accessed ¡today¿].

[16] 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.

[17] X. Qian, E. W. Patton, J. Swaney, Q. Xing, and T. Zeng. Machine learning on cataracts classification using squeezenet. In *2018 4th International Conference on Universal Village (UV)*, pages 1–3. IEEE, 2018.

[18] 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-

Prediction	Specificity	Sensitivity	Accuracy
Abnormality			
MRNet (reported)	0.714	0.879	0.850
MRNet	0.440	0.968	0.858
MRNet-Squeeze	0.560	0.968	0.883
MRNet-Attend	0.480	0.979	0.875
MRNet-SqueezeAttend	0.440	0.968	0.858
ACL tear			
MRNet (reported)	0.968	0.759	0.867
MRNet	0.894	0.907	0.900
MRNet-Squeeze	0.909	0.963	0.933
MRNet-Attend	0.803	0.778	0.792
MRNet-SqueezeAttend	0.864	0.852	0.858
Meniscus tear			
MRNet (reported)	0.741	0.710	0.725
MRNet	0.721	0.788	0.750
MRNet-Squeeze	0.735	0.731	0.733
MRNet-Attend	0.691	0.846	0.758
MRNet-SqueezeAttend	0.794	0.750	0.775

Table 3. Metrics on the validation set

analysis. *European Journal of Orthopaedic Surgery & Traumatology*, 22(4):315–326, 2012.

[19] 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 recognition*, pages 1–9, 2015.

[20] 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.

[21] Z. Tan, M. Wang, J. Xie, Y. Chen, and X. Shi. Deep semantic role labeling with self-attention. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[22] J. Yaqoob, M. S. Alam, and N. Khalid. Diagnostic accuracy of magnetic resonance imaging in assessment of meniscal and acl tear: Correlation with arthroscopy. *Pakistan journal of medical sciences*, 31(2):263, 2015.