MRNet with Alternate Pretrained Features

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

1. Introduction

Reading and interpreting MRIs is a time-consuming process that requires trained professionals. 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.

2. Problem statement

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.

3. Related work

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 [2] model to extract features from each 2D slice of the 3D MRI volume, followed by a global average pooling per slice and then max pooling across the volume, and finally a fully connected layer and sigmoid activation to predict a binary label for the series. Each series prediction (axial, coronal, and sagittal) is then used as a feature in a simple logistic regression to predict the final label in question. This process is repeated for each of the three independent labels.

An important aspect of the MRNet model is the data augmentation done during training. They randomly flipped, shifted, and rotated each MRI volume every time it was seen in order to prevent the model from overfitting the small data set.

4. Technical approach

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, ResNet, and SENets.
3. If time allows, try replacing the global average pooling and max pooling layers with a 3D convolutional layer across the MRI slices.

Unfortunately, while the Bien, et al. did release code for the architecture of their MRNet model, they did not release the data loading, hyperparameters, data augmentation, ensembling, or training code. This means that more of the effort for this project will be focused on simply reproducing their results.[1]

However, they did release source code for training MRNet on a set of external validation data from the 2017 Štajduhar et al. study [3]. This provides a good start for trying replicate their pipeline on the MRNet dataset.

5. 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.

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]:

[1] I wrote to the authors hoping that they would share their code privately so that I could spend more time on enhancements, but they said that the code is not avaialble.

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.

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.

6. Preliminary results

I have been able to successfully recreate the data loading and training pipeline, but have not yet completed the data augmentation, inference and ensembling steps. Figure 1 shows the loss curves for training and validation for each series and diagnosis. My hope is that the data augmentation steps will fix the obvious overfitting that is currently happening.

I have also plotted the AUC values for training and validation at each epoch in Figure 2. The model is memorizing the training set and strong regularization is needed to reach the performance.

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] 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.

[3] I. Štajduhar, M. Mamula, D. Miletić, and G. Ünal. Semi-automated detection of anterior cruciate ligament injury from mri. *Computer methods and programs in biomedicine*, 140:151–164, 2017.

Model	Ensembled	Abnormal	ACL	Meniscus
MRNet (reported)	yes	0.937	0.965	0.847
MRNet (milestone)	no	?	?	?

Table 2. AUC on the test set

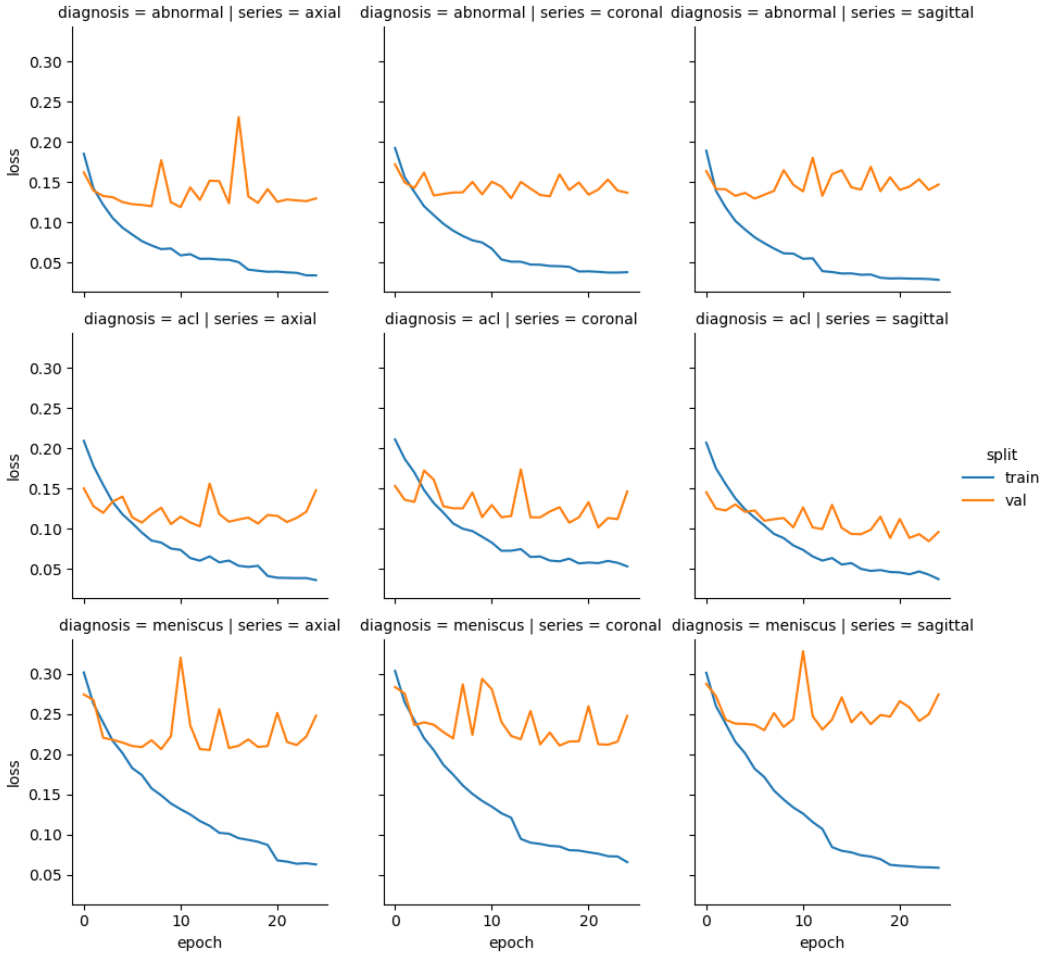

Figure 1. Baseline loss for each diagnosis and series.

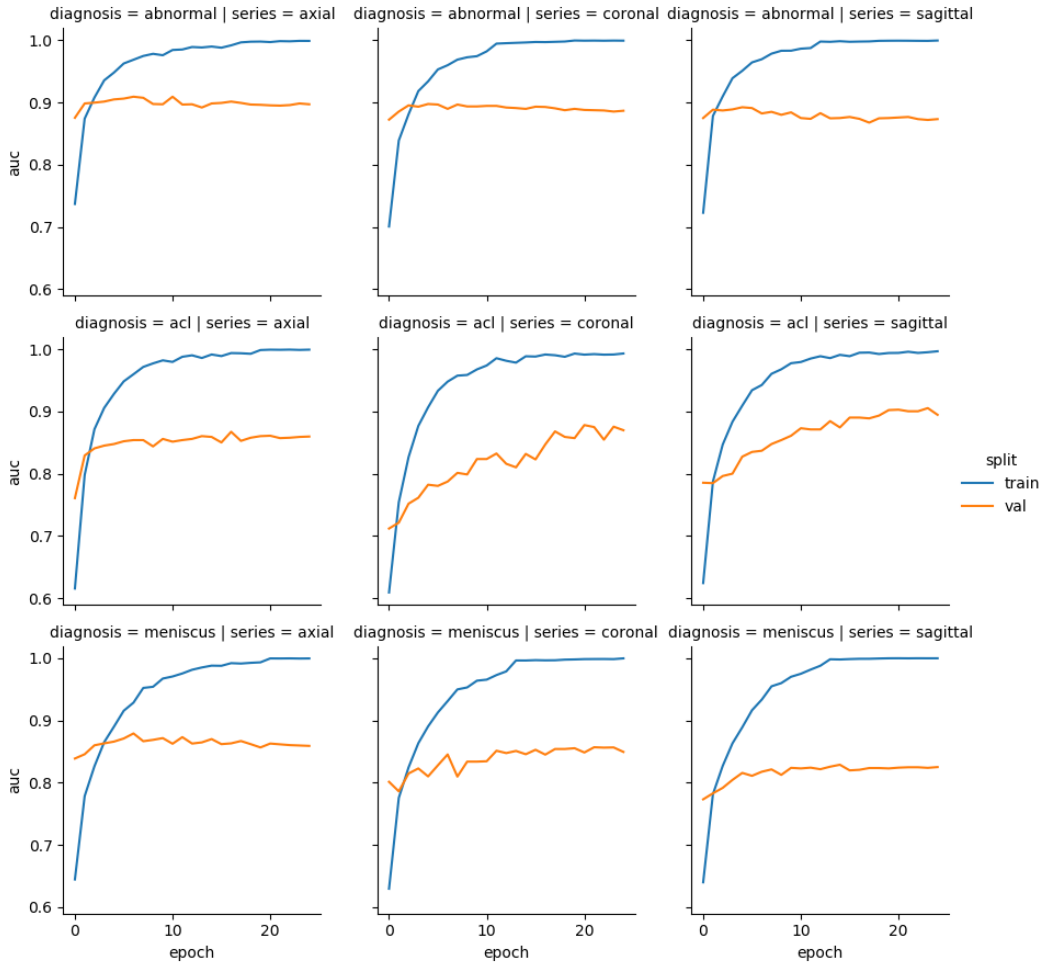

Figure 2. Baseline AUC for each diagnosis and series.