Introduction

This paper will outline the experimentation and results of a movie genre classifier. Using information scraped from the Internet Movie Database (imdb), the model is trained on 10,932 movie posters, then validated on an additional 2,186 images. Finally, the model was tested on 2,499 separate images, with a few examples of each genre examined for anecdotal review. The genres are pulled from imdb, with multiple genres possible per movie-- this is a multilabel classification problem, as each movie can belong to multiple genres. The multiple genres assigned to a particular movie are treated independently.

For the purposes of this paper, I only used posters that could be assigned to at least one of seven different genres: comedy, drama, action, animation, romance, adventure, and horror. I ran three separate models to compare performance: Model 1, which ran over 50 epochs; Model 1a, identical to Model 1, but run for only 17 epochs (explanation for that choice to follow), and Model 2, also run over 17 epochs, but with fewer layers. To speed processing, all posters were reduced in size to 30% of their original size.

Model Specifications and Accuracy Results

Model 1

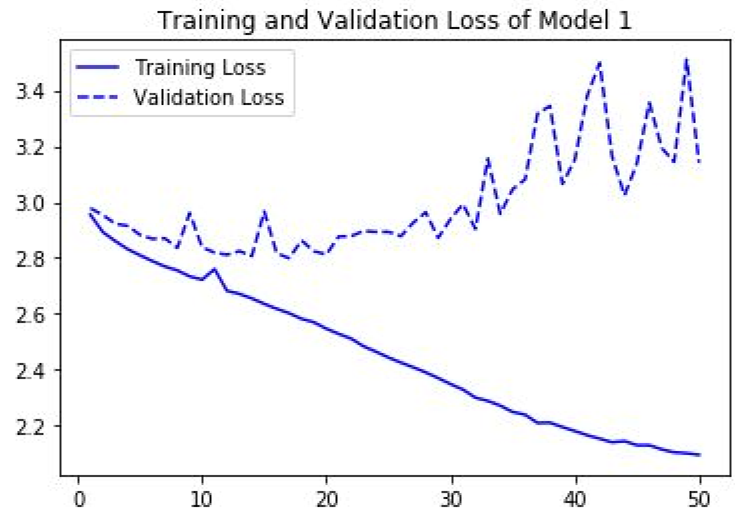
Model 1 has four convolutional layers, two max pooling layers, two dense layers, and three dropout layers. Details on each of the layers and their ordering can be seen in Table 1. There are 7.15 million total parameters in this model setup. Dropout layers are implemented after each max pooling layer and after the first dense layer. All three models in this experiment used dropout layers-- future experimentation in A4 will adjust the use of dropout layers to see the effect they have on overfitting.

Layer	Output Shape	Number of Parameters
Conv2D_1	(None, 55, 80, 32)	896
Conv2D_2	(None, 53, 78, 32)	9248
MaxPooling2D_1	(None, 26, 39, 32)	0
Dropout_1	(None, 26, 39, 32)	0
Conv2D_3	(None, 26, 39, 64)	18496
Conv2D_4	(None, 24, 37, 64)	36928
MaxPooling2D_2	(None, 12, 18, 64)	0
Dropout_2	(None, 12, 18, 64)	0
Flatten_1	(None, 13824)	0
Dense_1	(None, 512)	7078400
Dropout_3	(None, 512)	0
Dense_2	(None, 7)	3591

I will discuss results in more detail below, but the final validation accuracy for Model 1 was .5007, and the test accuracy was .4474.

Training and Validation Loss of Model 1

Figure 1 shows the training and validation loss at each of the fifty epochs of training. As one can see, the loss improvements diverge fairly quickly, with the validation loss reaching a minimum after only 17 epochs. As such, Model 1a, while structured identically to Model 1, learns for only 17 epochs. Model 1a produced a validation accuracy of .5007 (the same as Model 1, interestingly), and a test accuracy of .4752. While the test accuracy was slightly higher than Mode1, on far fewer epochs (and therefore much less computational expense), there are some anecdotal areas in which it appears to struggle that we will discuss later.



Model 2

For Model 2, two of the convolutional and one of the max pooling layers were removed, along with one of the dropout layers. The specifics of the model can be seen in Table 2, below.

Layer	Output Shape	Number of Parameters
Conv2D_1	(None, 55, 80, 32)	896
Conv2D_2	(None, 53, 78, 32)	9248
MaxPooling2d_1	(None, 26, 39, 32)	0
Dropout_1	(None, 26, 39, 32)	0
Flatten_1	(None, 32448)	0
Dense_1	(None, 512)	16613888
Dropout_2	(None, 512)	0
Dense_2	(None, 7)	3591

Model 2 produced a validation accuracy of .4597 and a test accuracy of .4088.

Results Discussion

From a pure accuracy standpoint, Model 1a is the clear winner, producing correct results 47% of the time, versus 45% and 41% for models 1 and 2, respectively. Overall accuracy may not be the best metric for this type of multilabel classification, however, as the model struggled with the animation genre-- it was able to correctly identify subgenres of animated movies, but only the 50 epoch model was able to correctly identify animated movies as “animated”.

Additionally, Model 1a was more likely than the other two models to predict subgenres that were not included in the original labels for a movie. It assigned the highest probabilities to the correct genres, but tended to muddy the waters a bit by also predicting unrelated genres. An interesting note: in the anecdotal examples, it seems that much of the time these genres could actually apply, so more research is needed to understand if this is a bug or a feature of Model 1a.

The reduction in epochs down to the point where Model 1 began to overfit certainly increased accuracy; however, it meant that the model had less exposure to less frequent genres, such as animation, which countered that rise in overall accuracy with struggles at subgroup accuracy. For Model 2, the scaling back of convolutional layers led to a decrease in accuracy overall, with the same issues with epoch reduction seen in Model 1a.

Future Research

It is important to note that this model is not actually pulling semantic context from these posters. It is focusing on color and shapes, without understanding that the shape is a car (signifying action movies, for instance) or two people dancing (signifying romance). Accuracy can likely be improved by adding semantic context, which will be attempted in A4. Additionally, I will experiment with adding and reducing dropout layers to understand their effect on overfitting, as well as using cyclical learning rates in place of RMSProp.