Effects of Data Augmentation on Equus Classification

Miguel Corona

April 2020

1 Introduction

Overfitting is the phenomena in which the model memorizes the characteristics of the training data and fails to generalize to yield unfavorable results to unseen samples. The models developed to identify the Equus members were capable of yielding a strong accuracy on the training data, but did not perform well on neither the validation nor the test set indicating that the training images were within the model's capacity. The capacity of a model is tied to the number of trainable parameters which can be interpreted as the model's ability to memorize characteristics of the training data. An increase in the number of samples in the training data may be utilized to combat the overfitting by exceeding the model's capacity.

2 Data Augmentation

The evaluations from the model indicate that the training data is within the model's capacity and that increasing the number of samples shall alleviate the overfitting that occurs on the limited training set. The dataset for the model is composed of 1044 images with 610 images dedicated to training, 204 to validation, and 230 to testing. A model can learn from additional training samples, but no new samples were collected nor were the validation and test samples redistributed into the training set. New samples were provided to the model via data augmentation in which a new sample was derived from an existing sample. The training data generator will yield transformed batches while the validation and test generators remain unchanged as these will serve as the data to use for evaluation.

The training set was doubled with the data augmentation. The 'new' samples were acquired from an additional pass through the training data as each batch is a transformed batch. The training data size was limited to be an integer multiple of the original training set as to not introduce an imbalanced class representation. Three sets of transformations were applied onto the training data for observation; translations only defined by the width_shift and height_shift arguments in the ImageDataGenerator, rotations only defined by the rotation_range, and a combined transformation of rotations and translations. The zoom_range was omitted due to the dataset containing samples with zoom applied.

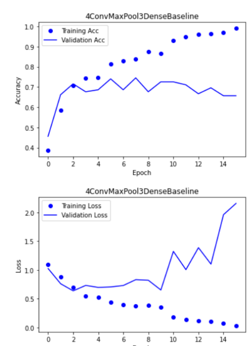
3 Experiments

3.1 Benchmark

The network that was previously observed to yield the best performance on the validation set is given by the following architecture.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 254, 254, 12)      336
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 127, 127, 12)      0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 125, 125, 12)      1308
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 62, 62, 12)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 60, 60, 32)        3488
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 30, 30, 32)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 28, 28, 32)        9248
_________________________________________________________________
flatten (Flatten)            (None, 25088)             0
_________________________________________________________________
dense (Dense)                (None, 24)                602136
_________________________________________________________________
dense_1 (Dense)              (None, 12)                300
_________________________________________________________________
dense_2 (Dense)              (None, 3)                 39
=================================================================
Total params: 616,855
Trainable params: 616,855
Non-trainable params: 0
_________________________________________________________________
```

The model was trained on the original dataset to acquire the benchmark results:



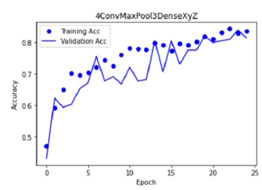
Val Loss	Val Acc	Test Loss	Test Acc
0.6401	0.7059	0.6130	0.7087

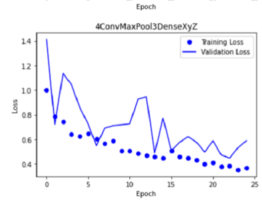
The benchmark execution indicates that the model is capable of learning the training set, but the validation accuracy plateaus at approximately 70%. The test loss and validation loss are comparable and the validation loss diverges through the training iterations. The model does not appear as though it would improve through additional epochs.

3.2 Translations Only

Observations were performed on images in which the data was augmented to contain translations along the horizontal and vertical axis. The translations for the images were restricted to be within 20% along each axis with nearest as the fill mode. The initial expectation for the results on translation only transformations was that neither the accuracy nor the loss would exceed those from the unaltered images due to the network's use of convolution layers. Convolution layers are capable of learning translation invariant properties suggesting that translating the images shall not have a noticeable impact on the model's performance.

The model was trained on the translations only images to yield the results:





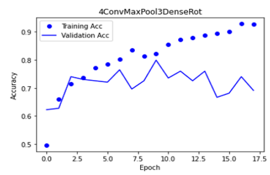
Val Loss	Val Acc	Test Loss	Test Acc
0.4707	0.7843	0.4263	0.7870

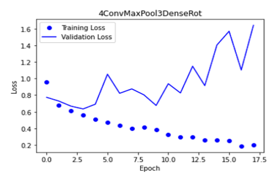
The translations only transformations yield promising results as the loss and accuracy improved for both the validation set and the test set. The model appears to benefit from the additional samples as it is capable of training across additional epochs prior to ceasing. The learning curves for the validation accuracy and the validation loss are also more aligned with those of the training set; No divergence in the learning curves indicate that overfitting was mitigated. The increase in performance may be attributed to the number of samples as the patterns the convolution layer recognizes are the same, but only differ in position. Adding more translated images may yield stronger performance as these images may be similar to those in the validation and test sets.

3.3 Rotations Only

Observations were performed on images in which the training samples were augmented to contain rotations about the normal axis. Rotations were limited to 30 degrees and nearest as the fill mode. The initial assumption was that rotations would have a more significant impact on the learning process than translations due transformations applied to the patterns in the data set. The patterns may be more difficult for the convolution layer to recognize due their the manner in which they are depicted and whether the each layer contained enough filters to recognized the transformed patterns.

The model was trained on the rotations only images to yield the results:





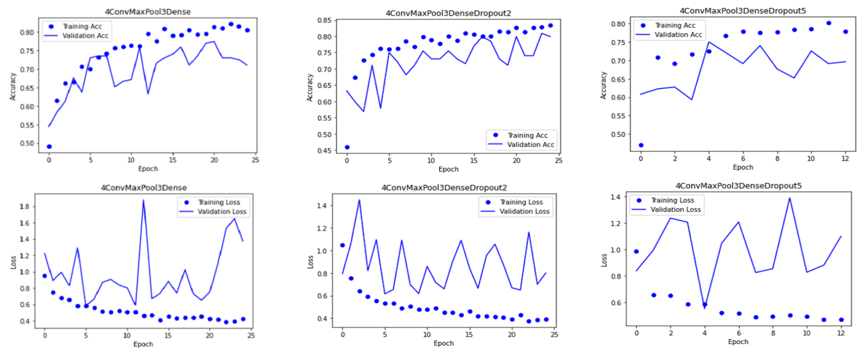
Val Loss	Val Acc	Test Loss	Test Acc
0.6425	0.7353	0.5506	0.7522

The rotated transformations do not yield the same benefits as those found in the translations only transformation even with both training sets being of the same size. The accuracy on the validation and test set improved compared to the benchmark, but the learning curves indicate that the rotations were not sufficient to prevent overfitting as the loss on the validation set diverged from the loss on the training set. The accuracy on the validation set hovers around the mark established by the benchmark. The model may not benefit from additional rotated images unless the rotations are adjusted to a lesser extent. It is possible that the minimal improvement is due to the patterns being too different in the training set than in the validation and test sets.

3.4 Rotations and Translations

The effects on performance of the model based on translations and rotations in the training samples were observed. The transformed samples contain the same configurations as the translations only and the rotations only with an additional parameter to mirror the images horizontally. Dropout layers were included into the model based on the literature's recommendation and due to the additional variability in the generated batches.

The model was trained on the transformed images to yield the results:



Dropout	Val Loss	Val Acc	Test Loss	Test Acc
None	0.5683	0.7353	0.5798	0.7174
0.2	0.6067	0.7598	0.5438	0.7435
0.5	0.6002	0.7304	0.5083	0.7696

The combination of translations and rotations yields less loss on the validation set and the test set than both the benchmark and the rotations results, but do not match the level of the translated images. The learning curves indicate that the dropout plays a role in determining whether the validation loss and validation accuracy align with the loss and accuracy of the training set. The dropout rate has a noticeable impact on the final evaluations of the validation and test sets, but the learning curves vary drastically across the learning process which may be expected due to the simulated noise.

4 Conclusion

The results indicate that acquiring additional samples for the training set may prevent overfitting even if the the new samples are not original images. Additional samples may be derived from the existing samples so long as the derived image has a transformation that makes it different to the model. It was observed that not all transformations yield the same performance improvement on the validation set and test set after training. It is speculated that the most significant improvements were with images containing transformations that best represent the validation and test sets.