Is It Raining Cats or Dogs?

Leslie Juengst & Samantha Roberts

DATS 6203: Group 5 Final Report

**INTRODUCTION**

Image classification and computer vision have grown increasingly more popular as new technologies have emerged. The evolution of robust tools for machine learning has led to new boundaries of more intricate hypotheses and novel practical applications of results. For example, the ability to properly classify images of different species of animals is a complex and delicate problem. The refinement and satisfactory solution to a problem such as this has wide application, such as predictive capabilities for agriculture, health services and animal services (Motta, et al., 2019).

The goal of this project was to correctly classify a variety of cat and dog images according to species using the Caffe framework. An assortment of models has been adapted to handle the inconsistent nature of images and feature extraction, varying specifically in neural layers and input functions to capture the most specific degree of variation in multiclass problems (Taheri & Toygar, 2018; Sun, Sun, & Yeung, 2017). This project implemented two separate convolutional neural network (CNN) architectures with a varied combination of parameters to find the prime classification strategy for the selected data set.

**DATA**

The data set selected is “Cats and Dogs Breeds Classification Oxford Dataset” on Kaggle.com (dr. Avicenna, 2019). It is comprised of 37 categories of animals, with approximately 200 images per class for a total of 7,393 images. It was expected that this sizing would provide enough samples for each category to potentially identify unique feature maps for accurate classification purposes. The images themselves were stored as jpg files, which required no nuanced processing. The images were renamed and separated based on whether it was a cat (0) or dog (1). Premade training and testing groups were discarded in lieu of a custom 70/30 split train and testing sets. All files were resized to 80 x 80 for ease of use then converted into a Lightening Memory-Mapped Database (lmdb) for input into Caffe. The new training file contained 5,175 records and the new testing file contained 2,218 records.

**MODEL DESIGN AND INSTRUMENTATION**

The neural network model class best suited to address image classification was a CNN. The first network design selected for evaluation was LeNet (Lecun et al., 1998). Customization was introduced by attempting a variety of output layer transfer and optimization functions, although the basic structure of the CNN remained the same. The framework for implementing this network was Caffe due to its processing ability specific to this type of neural network. Caffe uses a mini-batch style update to learn the network.

The first implementation of the LeNet neural network used the new training and testing sets that were created with an architecture and solver identical to a recent mini-project made available through Github (Hagan et al., 2017). The experiment design was kept consistent to develop an initial baseline with the new data set based on a previously successful structure. The performance of the network was measured by loss function output and the test set accuracy rating. Overfitting was addressed primarily by using dropout layers. Additionally, a low number of layers is optimal for a less robust data set.

Model and parameter selection were used as a main source of manipulation to obtain the network’s maximum accuracy score. A baseline was developed using a combination of functions and, which can be found in Table 1 (models) and Table 2 (parameters). These values were chosen based on default values often selected when using CNN’s, were implemented in similar projects conducted by the authors, or were corroborated with other animal classification CNN studies when information was available (Taheri & Toygar, 2018; Motta, et al., 2019). An AlexNet (Krizhevsky et al., 2012) network with no customization efforts was also trained for result comparison.

**RESULTS**

The first network tested was the baseline LeNet model using previously established model functions and parameters. Without any amendments to the existing model architecture and parameters, initial results were poor. The loss value was large and completely stagnant as if no updating were occurring at all and the accuracy score was 100% which is highly suspect (figure 1). In response, parameters were adjusted such as increasing gamma which lowers the learning rate more significantly as the number of iterations increases, decreasing momentum and decreasing the starting learning rate. This resulted in a more varied loss and accuracy, however, the accuracy merely oscillated between 0 and 1 and the loss was still largely stagnant (figure 2). Further, tweaking of the parameters showed very minimal improvement (figure 3).

Figure 3: Test Accuracy and Training Loss results of LeNet 2 model

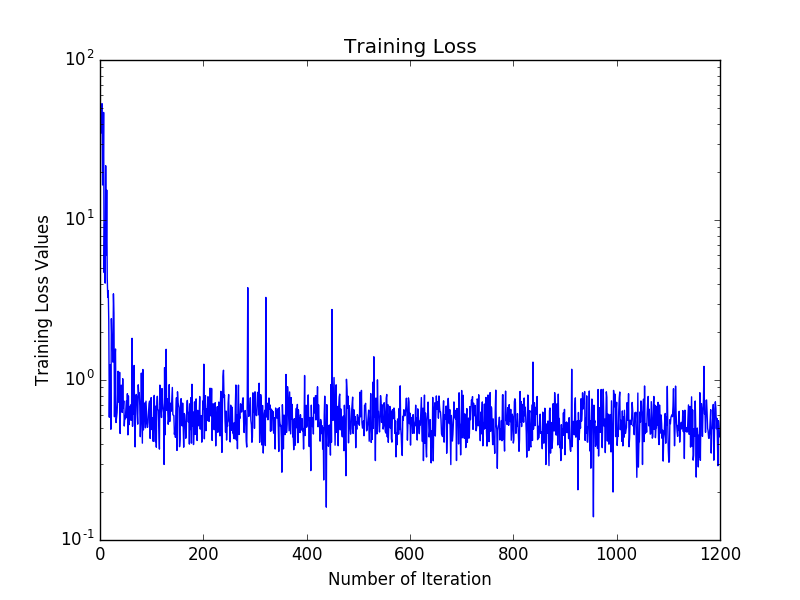
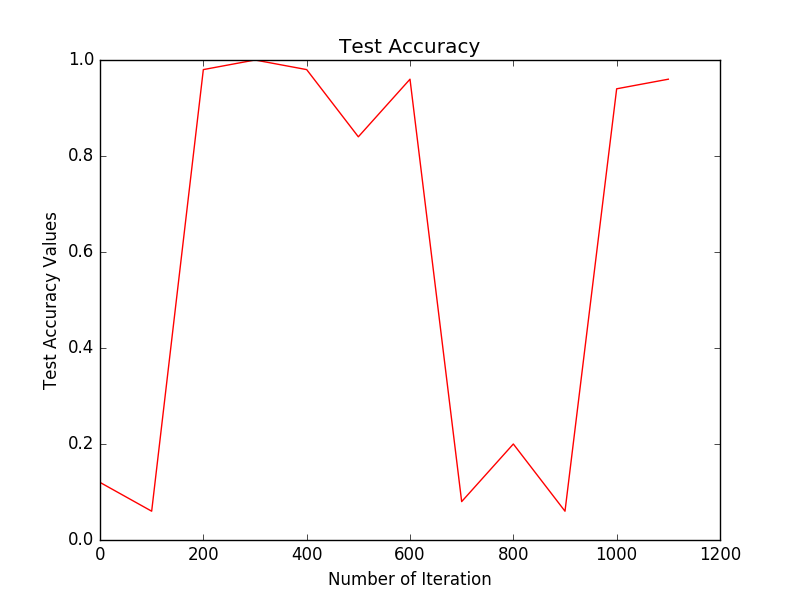


Figure 2: Test Accuracy and Training Loss results of LeNet 1 model

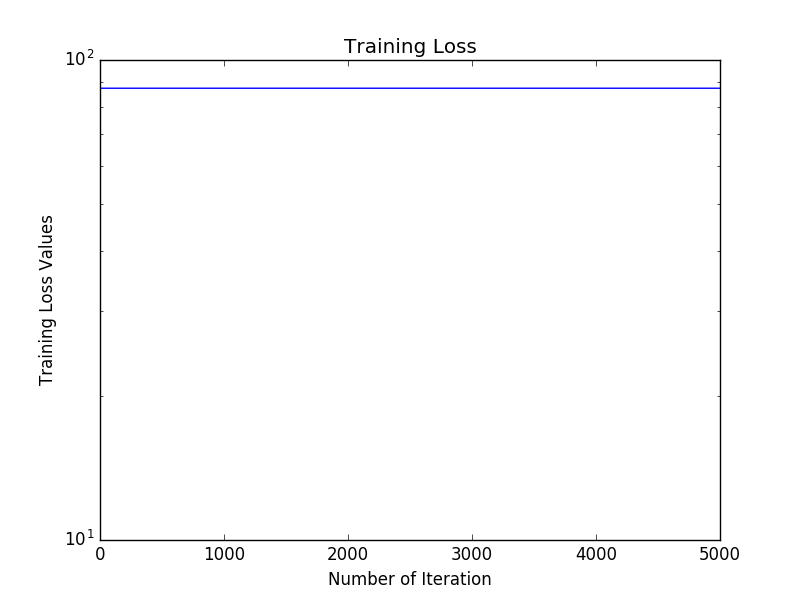
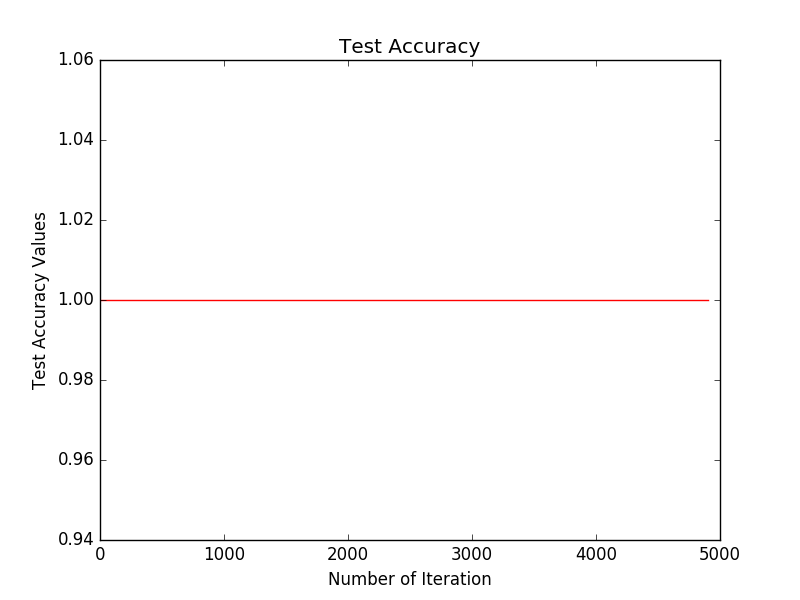
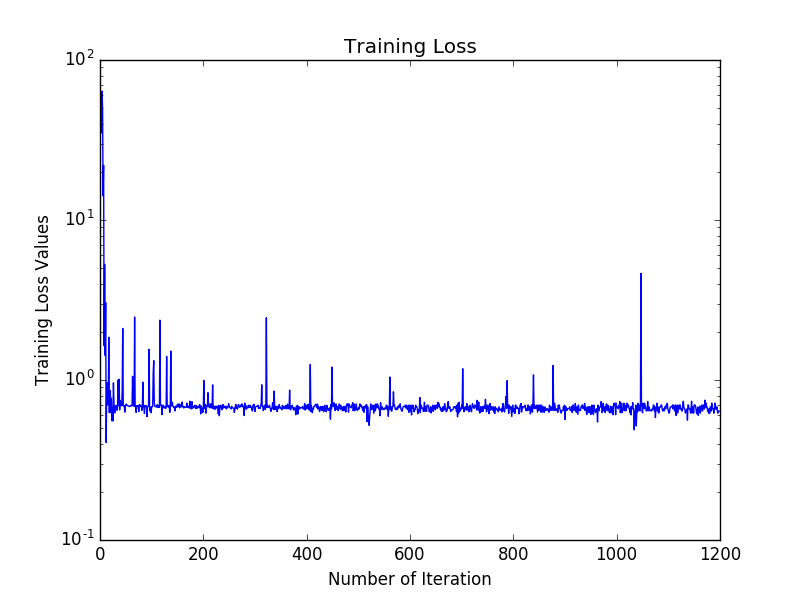
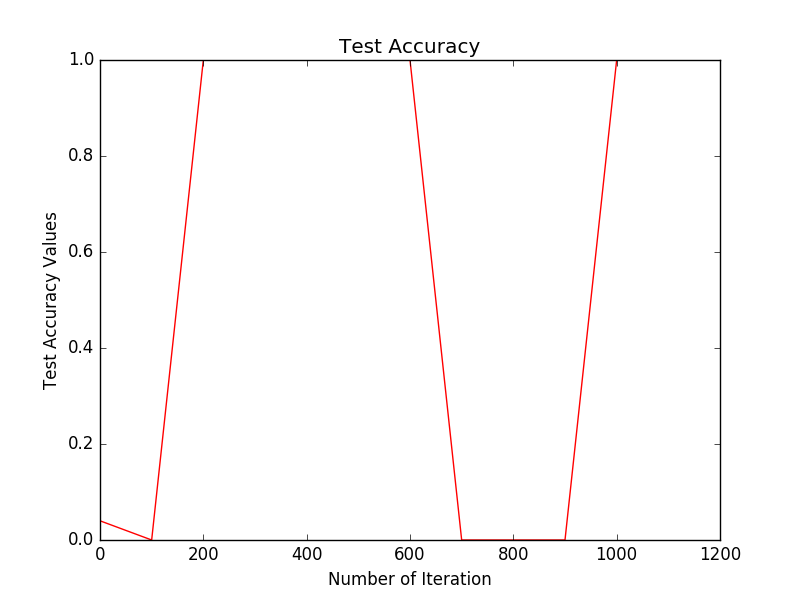
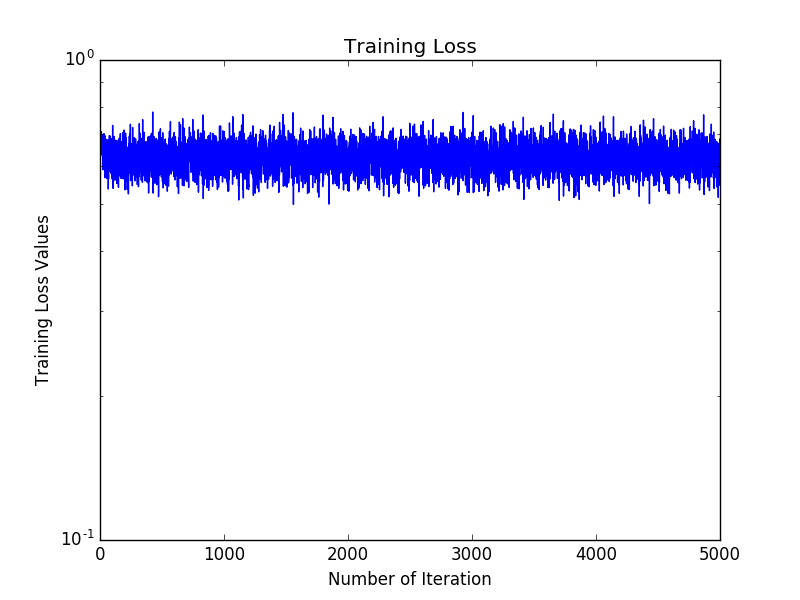
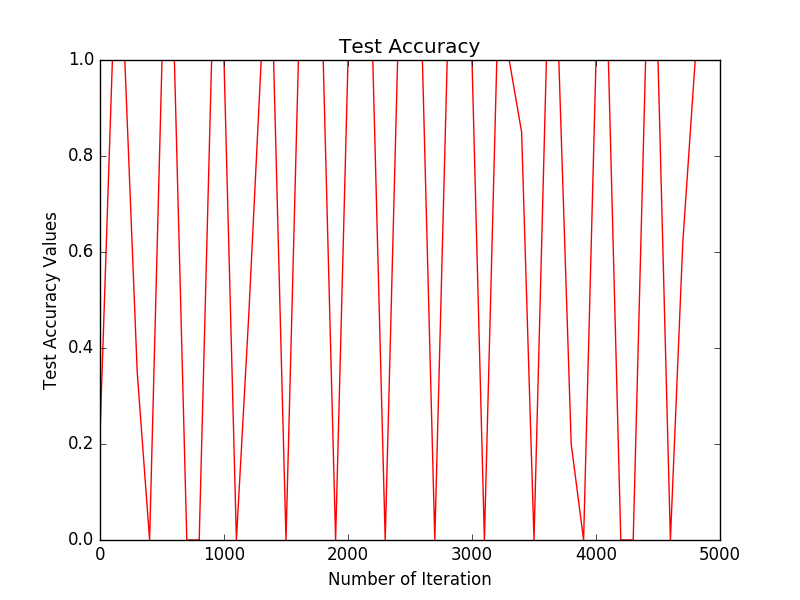


Figure 1: Test Accuracy and Training Loss results of LeNet base model

An AlexNet architecture was also attempted with baseline architecture and parameters. It is expected that this model might exacerbate any overfitting issues that exist with these data since there are a greater number of layers and neurons in this network as compared to the LeNet. As before, the performance as determined by loss trends and accuracy was extremely poor. Once again, the accuracy oscillated between 0 and 1 and the loss did not improve over time (figure 4).

Figure 4: Test Accuracy and Training Loss results of AlexNet model



One possibility that was considered as possible cause for this behavior is overfitting. The data set is relatively small with only just over 5,000 training records. Further, the classes are imbalanced in these data. The dog pictures outnumber the cat pictures by about 2 to 1. This could certainly impact the accuracy of the learned model. However, the extreme nature of the outputs, at 1 or near 1 and 0, coupled with the lack of improvement of loss makes this unlikely. Another possibility is that despite multiple checks, there is an error in the data or the creation of the databases for inputs.

**CONCLUSION**

The results of this work were disappointing to say the least. There is clearly an error in either the implementation of these networks or perhaps the data that the networks are training on. This classification is fairly straight forward so the number one step for the future is to determine what is causing this issue. After that an additional improvement would be to expand the classification to breeds instead of the simple binary dog/cat classifier, though as mentioned, the first step is getting the binary classifier working.

One of the biggest difficulties in working with this framework is the lack of publicly available information and a fairy small community of users. This particular situation of a persistent error highlights the lack of resources available to novice users of the deep learning frameworks. Caffe in particular has a good deal going on “behind-the-scenes” and proper, detailed documentation of how certain things work would have been helpful.

**REFERENCES**

[dr. Avicenna](https://www.kaggle.com/zippyz) (2019). Cats and Dogs Breeds Classification Oxford Dataset. Retrieved from <https://www.kaggle.com/zippyz/cats-and-dogs-breeds-classification-oxford-dataset#annotations.tar.gz>

Hagan,M., Jafari,A, Wang, S. (2017). Caffe Mini-project. URL <https://github.com/amir-jafari/Deep-Learning/tree/master/Caffe_/Mini_Project>

Krizhevsky, A., Sutskever, I., and Hinton, GE..(2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NIPS)*.

Lecun, Y., Bottou, L., Bengio,, L. and Haffner, P. (1998). Gradient-based learning applied to document recognition. In *the proceedings of the IEEE*.

Motta, D., Santos, A. Á, Winkler, I., Machado, B. A., Pereira, D. A., Cavalcanti, A. M., . . . Badaró, R. (2019). Application of convolutional neural networks for classification of adult mosquitoes in the field. Plos One,14(1). doi:10.1371/journal.pone.0210829

Sun, T., Sun, L., & Yeung, D. (2017). Fine-grained categorization via CNN-based automatic extraction and integration of object-level and part-level features. Image and Vision Computing,64, 47-66. doi:10.1016/j.imavis.2017.06.003

Taheri, S., & Toygar, Ö. (2018). Animal classification using facial images with score-level fusion. IET Computer Vision,12(5), 679-685. doi:10.1049/iet-cvi.2017.0079