Compare and Analyzing CNN Based Models' Performance

Archer Zhou, Junqiao Qu, Yilong Su

Faculty of Computer Science, Dalhousie University

Introduction

The fast growth of human population and the endless pursuit of economic development are making over-exploitation of natural resources, causing rapid, novel and substantial changes to Earth's ecosystems. The rapidly evolving computer vision can provide a powerful tool for biologist and animal photographers to classify different animals efficiently. In this project, we will explore the architecture different CNN based network, and use them to classify the animals in the data set. We will analyze the performance and try to make an explanation of relationship between performance and architecture.

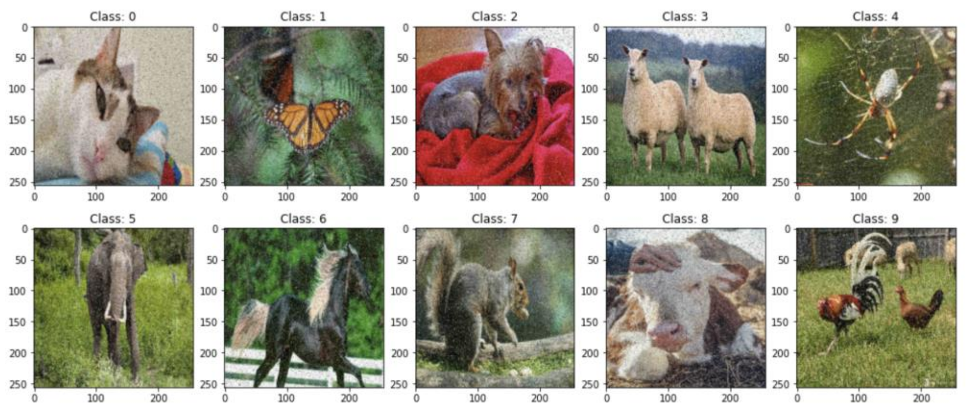
With the development of the neural network model, there are a lot of variants of the CNN model, the key difference is most of them has enhanced structure compared to the simple CNN model.

CNN-based model has played a very important role in deep learning. We will implement AlexNet and VGG model in our experiment, we will test and analyze why they can improve the overall performance. We will also refine a simple CNN model based on AlexNet and VGG respectively to validate the impact of certain structure. We focus on four points in our project, the "relu" activation function, the dropout layer, the deeper structure and smaller kernel size. We conclude that all these four points can improve the performance of CNN-based model.

Dataset and Features

The Animal 10:

- This original dataset comes form Kaggle for training
- Consists of 28,000 medium quality images
- With 10 categories of animal: dog, cat, horse, spider, butterfly, chicken, sheep, cow, squirrel, elephant.



The Bing animal:

- This dataset containing images crawled from BING with the same categories.
- Consists of 229 images:
  - 1-150: Original images
  - 151-224: Image with salt-and-pepper noise
  - 225-229: Random images.

Models

Baseline CNN

- Four convolutional layers
- Activation layer: Tanh

Refined CNN based on VGG

- Deeper structure
- Smaller Kernal Size

Refined CNN based on AlexNet

- Using Relu
- Using Dropout

AlexNet

VGG

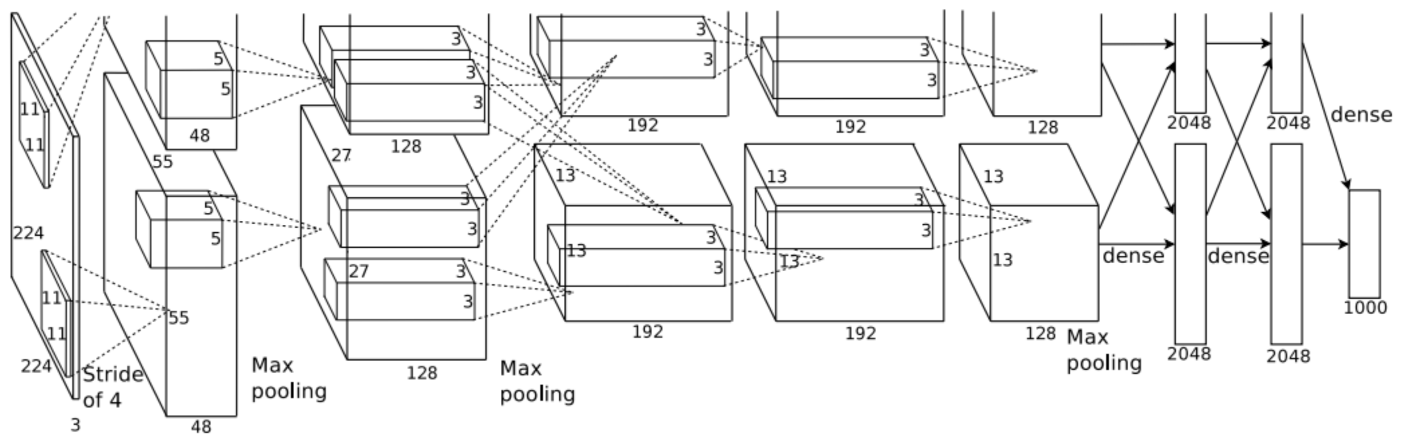


Fig Above : Sample AlexNet structure

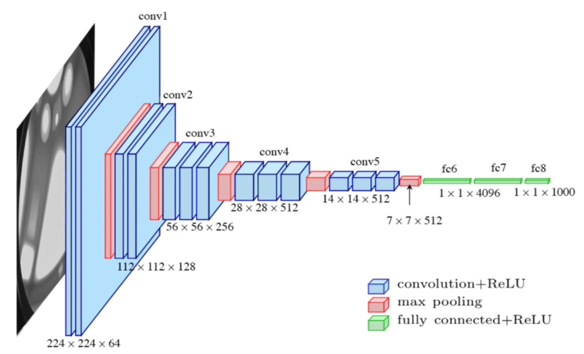
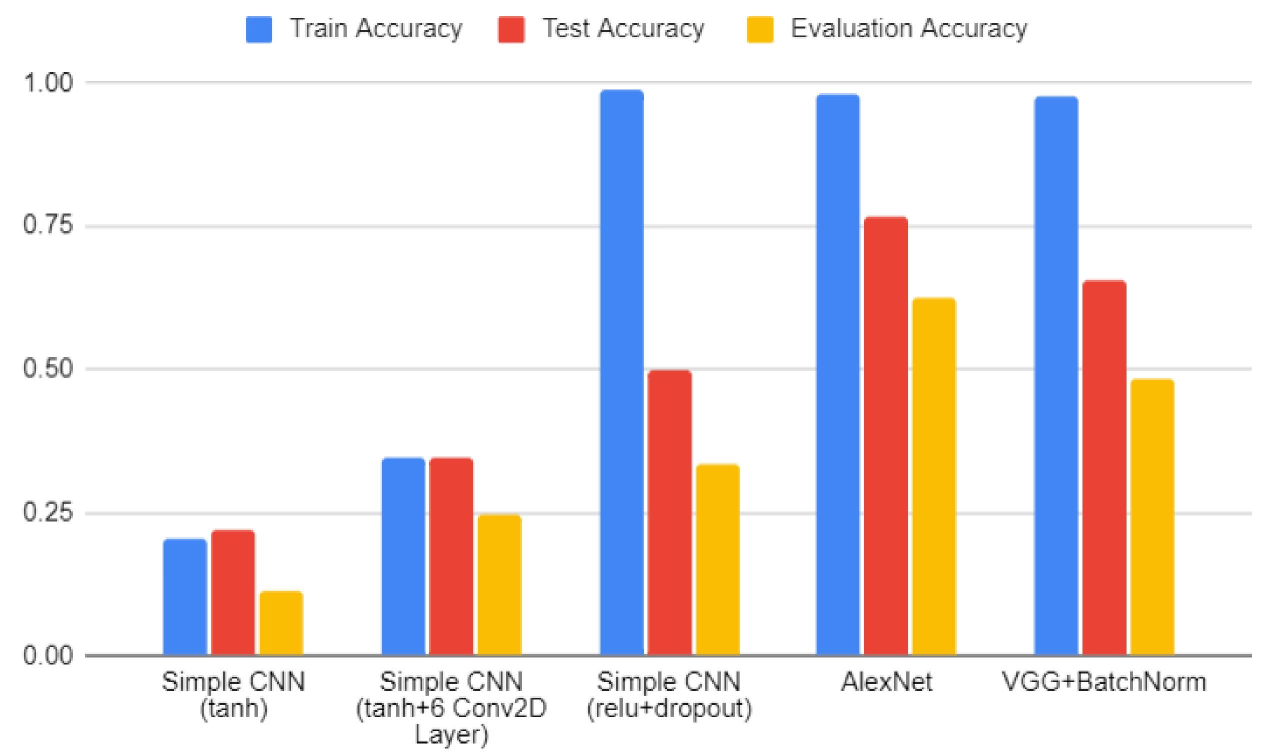


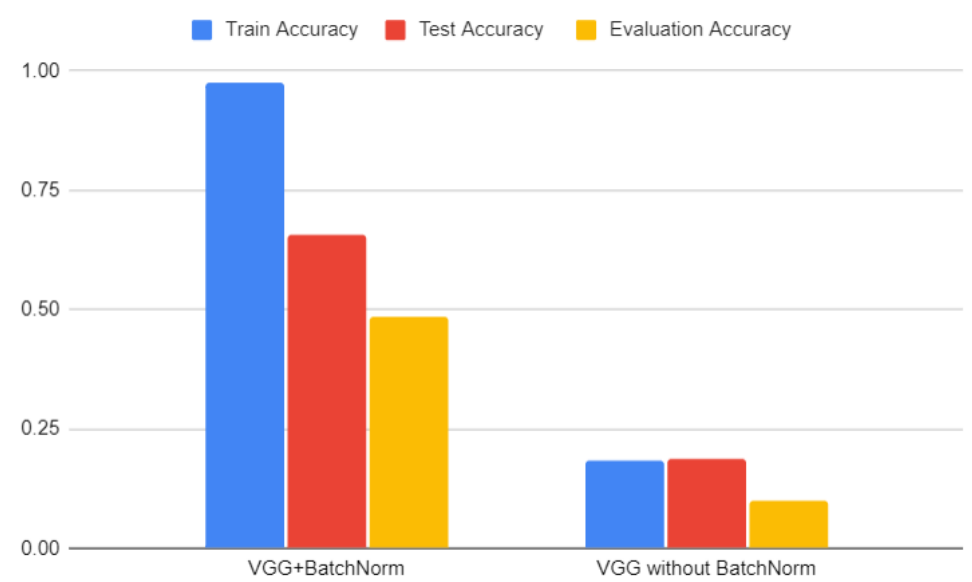
Fig Left: Sample VGG structure

Results and Evaluation



According to the figure, we can noticed that AlexNet, VGG and the refined CNN-based models all have better performance than our baseline CNN model. And AlexNet has the best performance in all the models. Our refined CNN model has a decent improvement compare to the baseline, but not as good as the AlexNet and VGG model. It means the "relu" activation function, the dropout layer, the deeper structure and smaller kernel size can improve the model overall performance.

According to the figure on the right , we found that batch normalization is very important to the VGG model, the performance drops a lot without batch normalization. The performacne of VGG without batch normalization is almost at the same level of our baseline CNN model.



Discussion

According to the two comparison, baseline and AlexNet, baseline and refined CNN based on AlexNet, we can see the model's performance improves. We address this to the "relu" activation function and dropout layer.

ReLU allows neuron to express a strong opinion, it is less sensitive to random initialization and the gradient doesn't saturate.

According to the two comparison, baseline and VGG, baseline and refined CNN based on VGG, we can see the model's performance improves. We address this to the deeper structure and replace large kernel size layer with smaller kernel size layer.

The deeper the structure, the better the model can fit the features. The deeper the structure, also the easier for each layer to do the job. For these reasons, a deeper structure can improve the performance. According to the research in the VGG paper\cite{VGG}, a smaller kernel size can improve the performance, we also validate this in our experiment. Most of the useful features in an image are usually local and it makes sense to take few local pixels at a time to apply convolutions.

Another important thing we have found during our experiment is the importance of batch normalization to VGG model. When we are building and testing the VGG model, we tried to build it without the batch normalization and the result turn to be terrible. The batch normalization plays an important role in the VGG model, since this is not our main goal of the project, we will leave it to future work.

Future Work

We will further compare the models' performance with pure noisy data and pure clean data. In our experiment, we combine noisy and clean data to get a more generalize result. We can experiment with the baseline model and our refined model to see how performance varies between pure noisy data, pure clean data and combined data.

We can also experiment with some other CNN based data to see the structure's impact on the model's performance, we can use model such as ResNet.

We noticed that batch normalization is very important to VGG model, we can further test the impact of batch normalization on different models.

References

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep
convolutional neural networks. Advances in neural information processing systems, 25,
2012.
[2] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning
applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.
[3] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-
scale image recognition. arXiv preprint arXiv:1409.1556, 2014.