AIML 425 | Assignment 3
2 Dimensional Convolutional Neural Network (CNN)

Irshad Ul Ala | Student ID 300397080
Victoria University of Wellington

August 23, 2021

1 Introduction

In this paper, we explore the versatility of simple 2 dimensional convolution models that classify images with one object(shapes: triangle, rectangle or circle) in them only. After training and testing said models, they will be evaluated against images that contain 2 of the 3 objects/shapes.

2 Theory

2.1 Convolution

Convolution between 2 functions is often described as taking the reverse of one, and sliding it across the other, and calculating the sum of the 2. Doing this operation without reversing one of the functions is conversely called cross correlation.
A convolutional model similarly, takes the input matrix, and convolves it with a **rotated** smaller matrix (a filter or kernel), as shown in the example in Figure 1.

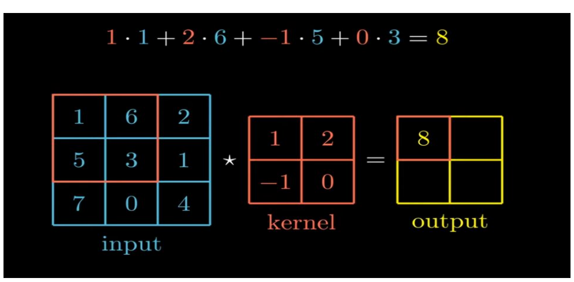

Figure 1: Convolving filter/kernel with sample input (Code, T., 2021)

The input layer shown above is typically called the receptive field, and the numbers in the kernels are called weights. One could conceptualise the kernel, as a "feature identifier". The weights in the kernels are initialized similar to how weight matrices are randomized. At the first epoch during training the network is only able to detect very basic, if any significant features, and with each epoch the CNN will gravitate towards more appropriate values for the features.

2.2 Pooling

Pooling reduces the dimensions of the feature map, while maintaining the most important noteworthy features. This is the one of the reasons why CNNs require less computation than standard networks.

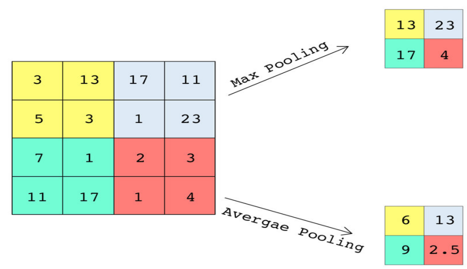

Figure 2: MaxPooling and Average Pooling method comparison (Aljaafari, N., 2018)

In our case, we are using MaxPooling, which takes the maximum value in the window used (2x2 in our case).

2.3 Softmax versus Sigmoid activation

The sigmoid function

$$f(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}} \tag{1}$$

takes any range real number and returns the output value which falls in the range of 0 to 1. It is used to return the total number of classes with a value greater than a predetermined threshold (significance level). This function would therefore be suitable for multi-label classification problems, since the sum of the values in the outputted vector are greater than 1. This is ideal to compare predictions against groundtruths such as [1,0,1] - triangle and circle present in image in our case.

In contrast, the softmax function

$$f(x_i) = \frac{e^{x_i}}{\sum_i^k e^{x_j}} \tag{2}$$

evident by the formula above, sums to one across the output vector. Such a configuration favours one-label classification problems. This means that this model would not be expected to be able to perform very well in classifying compound images, but better at singular images.

3 Results & Conclusion

A shuffled sample of 3000 images containing either a triangle, rectangle or circle was created. 3 convolutional models were trained using 1200 of these images, and validated against the remaining 1800. The first model was a standard CNN model with softmax output, the second, one with a batch normalization layer, and the final one, a CNN model with sigmoidal output.
These trained models were then evaluated against 30 sets of 1000 compound images containing 2 of the 3 shapes. Finally, they were evaluated against another 20 sets of 2000 images, consisting of 1000 singular and 1000 compound images of random order.
At 5 epochs, the simple CNN model achieved 83.1% categorical accuracy with validation data, the second achieved 95.9% accuracy, and the final sigmoidal CNN model achieved 83.7%.

3.1 Evaluation against sample of 1000 compound images | 30 times

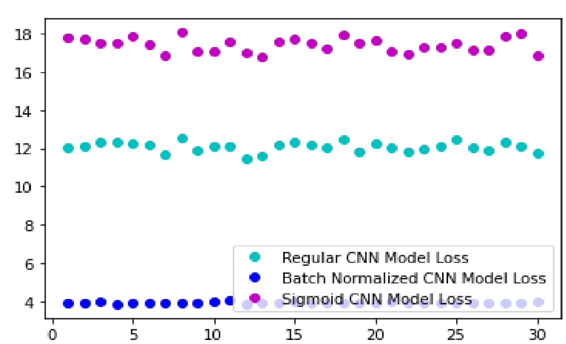

Figure 3: Categorical Crossentropy loss for all 3 models against sample of 1000 compound images, repeated 30 times

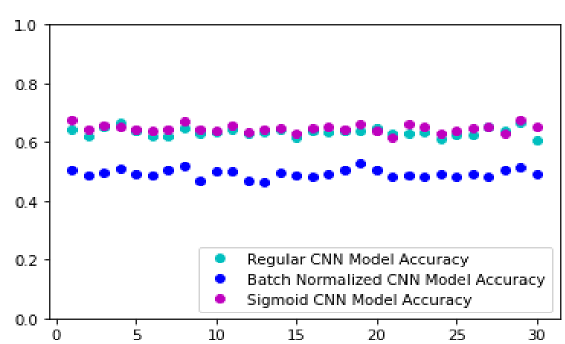

Figure 4: Categorical Accuracy loss for all 3 models against sample of 1000 compound images, repeated 30 times

3.2 Evaluation against sample of 2000 compound & singular images | 20 times

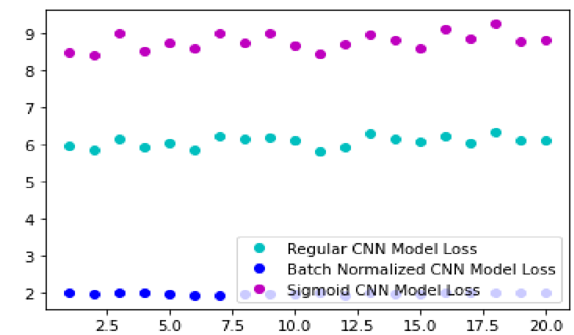

Figure 5: Categorical Crossentropy loss for all 3 models against sample of 2000 compound & singular images, repeated 20 times

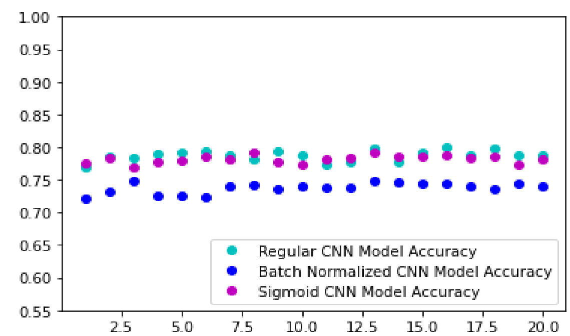

Figure 6: Categorical Accuracy loss for all 3 models against sample of 2000 compound & singular images, repeated 20 times

3.3 Conclusion

All models were trained using only singular images (images with one of the 3 shapes). The loss of the 3rd model was noticeably the highest among the 3. This is likely due to the nature of sigmoidal output not being capped to 1, like in softmax, which is why its loss might not reflect its proficiency.
The concept behind sigmoidal output implied that a trained model would be more flexible and better categorise compound images, contrary to the results. Instead, training said models against compound images beforehand may produce different results.
Batch normalization appears to make a singular image CNN perform slightly worse than a standard model for multi-image classification, but better at singular image classification (95.9% accuracy vs 83.1% at 5 epochs).
It is worth considering normalizing the input shapes (by dividing by 255) beforehand, for it could have produced different results.

4 Code

Workspace on Colab (`https://colab.research.google.com/drive/1SS_NyLw_TrYyzMJJh7WjfezB7dd8ApfN?usp=sharing`) .

5 References

Aljaafari, N. (2018). Ichthyoplankton Classification Tool using Generative Adversarial Networks and Transfer Learning. Researchgate. Retrieved 23 August 2021, from https://www.researchgate.net/publication/332092821_Ichthyoplankton_Classification_Tool_using_Generative_Adversarial_Networks_and_Transfer_Learning.

Code, T. (2021). Convolutional Neural Network from Scratch | Mathematics & Python Code [Video]. Retrieved 23 August 2021, from https://www.youtube.com/watch?v=Lakz2MoHy6o.

Kleijn, B. (2021). AiML425 on 7/14/2021 (Wed) [Video]. Retrieved 22 July 2021, from https://vstream.au.panopto.com/Panopto/Pages/Viewer.aspx?id=762fabd3-326f-4d71-8aa2-ad590165213a.