Liz Simpson, 951428

951428@swansea.ac.uk

m45J: Object recognition

Comparing the use of AlexNet and ResNet Convolutional Neural Networks to classify objects

Introduction

The aim of this paper is to discuss ways for machines to accurately classifying images. Convolutional Neural Networks (CNN) are a machine learning technique used to create a model to accurately predict the classification unknown objects (Yang, Choi and Lin, 2016). However, there are many different architectures that can provide powerful frame works for classification. In this paper we will discuss which CNN architecture creates the most effective machine leaning model. We implement and compare two architectures using a subset of the CIFAR-100 dataset. Firstly, we implement an AlexNet CNN model, we train the model using fine and coarse labelling sets and compare its effectiveness to a more recent architecture, the ResNet CNN. Both models were scored on their learning time, accuracy of results using coarse labelling, accuracy of results using fine labelling and, on the accuracy of predicting objects from an unseen dataset. ResNet performed better than AlexNet on all fronts.

Method

In this section we will discuss the dataset and the implementation, training, and testing of the AlexNet and ResNet Networks. We will discuss how features are extracted, normalised and how the models were trained and tested. All CNN’s will be written in jupyter notebooks usning the tensorfloew and kerkas framework.

The Dataset

A large dataset is needed to build a CNN. In this instance I used a subset of the CIFAR-100 dataset. The CIFAR-100 subset contains 32x32 images of 50000 objects, ranging from plants to tractors. As CNNs are a supervised learning technique, each image is labelled with a “fine” (more specific) and “coarse” (more general) label (Krizhevsky, Nair and Hinton, 2009). Fine labels identify 100 classes while the coarse dataset identifies 20 super classes ranging from fish to trees. We will analyse the effectiveness of AlexNet and ResNet models using both label variations.

Pre-processing of the data is needed to ensure all the data is within the appropriate format. Before using this data in a model, the data will be shuffled as to avoid unnecessary bias by the ordering of the images. For each model, the dataset is split into 2 sections. 80% of the dataset will be used as training data while the remaining 20% will be withheld and used to test the effectiveness of the models. All the input images must be 32x32 pixels and normalised so that there is no degeneration of gradients.

CNN Architecture

I chose to use CNNs for this project as it is a supervised and can provide high accuracy results for image recognition. The similarities between AlexNet and ResNet is that they both use layers. However, ResNet employs skip connections with batch normalisation which significantly lowers the complexity of previous networks and can achieve human level accuracy.

Every model will have a: batch\_size= 100, epochs=100, learn\_rate=.001

AlexNet

AlexNet was chosen because it is a relatively simple and early implementation of a CNN, made in 2012 (Khan *et al.*, 2020). The implementation of the AlexNet CNN closely follows the demonstration by Kumar. The difference between architecture comes from the differences in layers (Sharma, Jain and Mishra, 2018). AlexNet is made up of layers of convolutional (partial) Conv, Normalisation and Pooling layers before 3 layers of fully connevcted layers and finally an output layer that uses softmax to reduce the outputs into our 20 classes (Kumar, 2020).

Features are extracted by …

Feature processing is …

how they are trained and tested.

ResNet

Resnet is short for Residual Network and is a much deeper model, meaning this model has more layers. Generally speaking deeper models are more accurate but are more complex and thus take longer to train. Deeper networks can also start to overfit the data so the deeper the network, the more likely your accuracy will plateo and state to degrade (Sharma, Jain and Mishra, 2018).

However, it is observed that the main thrust in CNN performance improvement came from the restructuring of processing units and the designing of new blocks. (Khan *et al.*, 2020)

* Extract features using HOG’s
* 2d CNN trained and tested
* 3d CNN trained and tested
* Graph plotted of the learning and accuracy and scored on unseen data
* The CIFAR-100 dataset has 50000 training data of 100 different categories with a smaller selection of 10000 images to be used as tests for the models
* There are 2 sets of labels, fine and coarse, both will be usedx

and ResNet(2015)(Das, 2017)

The initial hypothesis is that the ResNet will be the most accurate model.

Hypothesis

I expect that ResNet will outperform AlexNet.

The proposed method shows clear understanding of the material. Multiple comparative methods are presented, and the reasoning behind their selection is well presented. There is deep, critical reasoning behind the choices.

Results

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **AlexNet** | **ResNet** |
| **Fine Labels** | **Accuracy** |  |  |
|  | **Accuracy\_val** |  |  |
|  | **Time to Train** |  |  |
|  | **Score on unseen** |  |  |
|  |  |  |  |
| **Coarse Labels** | **Accuracy** |  |  |
|  | **Accuracy\_val** |  |  |
|  | **Time to Train** | **4.5 hours** |  |
|  | **Score on unseen** |  |  |

* Time to Train from AlexNet, ResNet
* Accuracy, accuracy\_val readings
* Accuracy best of 5
* overfitting
* Confusion matrix

A well-presented and thorough evaluation. The results provide a clear insight into the experimentation proposed within the methodology section. A clear understanding of the results is evident.

Present your experimental results in thissection. Explain the evaluation metric(s) you use and present the quantitative results (including the confusion matrix).

Conclusion

ResNet was the most accurate over the coarse labels and the fine labels..

A concise summary of the report. The critical analysis shows clear understanding of the materials and findings. Shows well considered suggestions for future work.

Provide a summary for your method and the results. Provide your critical analysis; including shortcomings of the methods and how they may be improved.

References**.**

Das, S. (2017) *CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more… | by Siddharth Das | Analytics Vidhya | Medium*. Available at: https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5 (Accessed: 13 April 2021).

Khan, A. *et al.* (2020) ‘A survey of the recent architectures of deep convolutional neural networks’, *Artificial Intelligence Review*, 53(8), pp. 5455–5516. doi: 10.1007/s10462-020-09825-6.

Krizhevsky, A., Nair, V. and Hinton, G. (2009) *CIFAR-10 and CIFAR-100 datasets*, *https://www.cs.toronto.edu/~kriz/cifar.html*. Available at: https://www.cs.toronto.edu/~kriz/cifar.html (Accessed: 13 April 2021).

Kumar, V. (2020) *Hands-on Guide To Implementing AlexNet With Keras For Multi-Class Image Classification*, *Analytics India Magazine*. Available at: https://analyticsindiamag.com/hands-on-guide-to-implementing-alexnet-with-keras-for-multi-class-image-classification/ (Accessed: 13 April 2021).

Sharma, N., Jain, V. and Mishra, A. (2018) ‘An Analysis of Convolutional Neural Networks for Image Classification’, *Procedia Computer Science*, 132(Iccids), pp. 377–384. doi: 10.1016/j.procs.2018.05.198.

Yang, F., Choi, W. and Lin, Y. (2016) ‘Exploit All the Layers: Fast and Accurate CNN Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers’, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2129–2137. doi: 10.1109/CVPR.2016.234.