COURSEWORK: OBJECT RECOGNITION

Alicia Addo 2220858

23 DECEMBER 2024

CSC345 BIG DATA AND MACHINE LEARNING

Introduction

The CIFAR-100 dataset, developed by the Canadian Institute for Advanced Research, is a subset of the 80 million Tiny Images dataset. [1] [2] It consists of 60,000 images, each measuring 32x32 in height and width with three colour channels. The dataset organises 100 distinct classes spanning 20 broader superclasses. Every image is precisely labelled with both a 'fine' label indicating its specific class and a 'coarse' label representing its broader superclass. The dataset is strategically divided into 500 training and 100 testing images per class, totalling 600 images per category.

This study investigates two reputable classification algorithms: a Convolutional Neural Network (CNN) and a Fully Connected Neural Network (NN), which will enable a comprehensive comparative analysis of the classification performance over the CIFAR-100 dataset, by evaluating their performance in both fine-grained and coarse-grained object detection on the CIFAR-100 dataset.

Preprocessing

The dataset was pre-processed by normalising pixel values to the range [0, 1] and applying one-hot encoding to the labels. Images were transposed as necessary to suit the models' input requirements. No additional feature extraction techniques were used as the models directly processed the image data.

Method 1

The fully connected Neural Network (NN) utilises dense layers to classify features extracted from the input images. It does so by transforming raw pixel values into feature vectors and passing them through densely connected layers.

To begin this approach, we flattened the 32x32x3 image into a 3072-dimensional vector. The first hidden layer consists of 256 neurons and makes use of ReLU activation function, which introduces non-linearity to model complex patterns in the data. The weights of this layer are initialized using the HeNormal method to endure efficient weight distribution during training. To stabilise training and reduce the sensitivity of the model. Additionally, a dropout rate of 0.3 is introduced to regularise the model and prevent overfitting by deactivating random neurons during training Following the first layer, a second layer is added with 128 neurons. Similarly repeating the same format to further generalise the models' capabilities for classification.

The model is compiled using the Adam optimiser, which adapts learning rates during training to converge the model efficiently. The performance of the model is evaluated using accuracy as the primary metric to make a standard comparison with the benchmark and method 2.

The model was trained over 20 epochs with a batch size of 64. During training, the network was able to different between classes by adjusting weights and biases to minimise the loss function. Due to the simplicity of the fully connected neural network, the results demonstrated the limitations of its ability to distinguish classify images with reasonable accuracy specially on the fine-grained labels.

Method 2

The Convolutional Neural Network (CNN) for the given task because of its performance in processing visual data. Unlike traditional neural networks not every node in a layer is connected in the subsequent layer, this works to reduce the number of parameters to a more computationally efficient and less prone to overfitting. Leveraging this connectivity CNNs can effectively extract meaningful image features, making CNNs ideally suited for the CIFAR-100 object classification task.

The first step we took was the 32x32x3 input images. The first convolutional block consists of two convolutional layers, each with 64 filters, a kernel size of 3x3, and ReLU activation. The HeNormal initialiser was used to optimise wight initialisation and zero-padding to preserve spatial dimensions, to ensure the output dimensions remain the same as the input, similarly to the neural network in method 1. Rather than start from scratch, method 2 uses method 1 and improves it dramatically. The block concludes with a max pooling layer with a pool size of 2x2, reducing the spatial dimensions and thereby compressing the feature maps and reducing the computational complexity. Furthermore, a dropout rate of 0.5 was employed to regularise the model and prevent further overfitting as seen in earlier implementations most prominently in the fine-grained data. The second block in the CNN, follows a similar structure to the first however, it replaces the filters with 128 for both convolutional layers. Increasing the filters in subsequent layers allows the network to capture more complex features from the images enabling better classification performance.

After the convolutional blocks the output is flattened into a one-dimensional vector, transitioning from a spatial feature extraction to fully connected layers for classification. A dense layer with 256 neurons, ReLU activation and initialised by HeNormal weights is followed by a dropout layer of 0.5 further regularising the model. The last layer consists of a dense layer with a parameter passing the number of neurons equal to the number of classes (100 for the fine label or 20 for the coarse label), using a softmax activation to output a probability distribution over the classes.

The CNN is compiled in a way that replicated methods one, using the Adam optimiser, categorical crossentropy loss and accuracy as the primary metric. However, to prevent overfitting on this more complex model the model trained for 20 epochs using early stopping with a patience of 5. The mechanism monitored the validation loss and restored the best-performing weights if the validation loss stops improving.

Results

The performance of the models was evaluated on both the fine-grained and coarse-grained classification data. The accuracy of achieved by each model is summarised and compared to the provided benchmark in Table 1 below.

Labels	Provided Average Benchmark	Neural Network	CNN
Fine	24.49%	21.6%	32.9.%
Coarse	39.43%	33.7%	47.3%

Table 1 Results

The CNN significantly outperformed the neural network on both classification tasks, achieving 32.9% accuracy for fine-grained labels and 47.3% accuracy for coarse-grained labels. The neural network did not exceed the benchmark, but its result hovers nearby, though still showing signs of overfitting in its slow convergence likewise for the CNN. However, through implementation and testing the results of the overfitting have been significantly reduced, and the CNN model demonstrates a strong performance improvement.

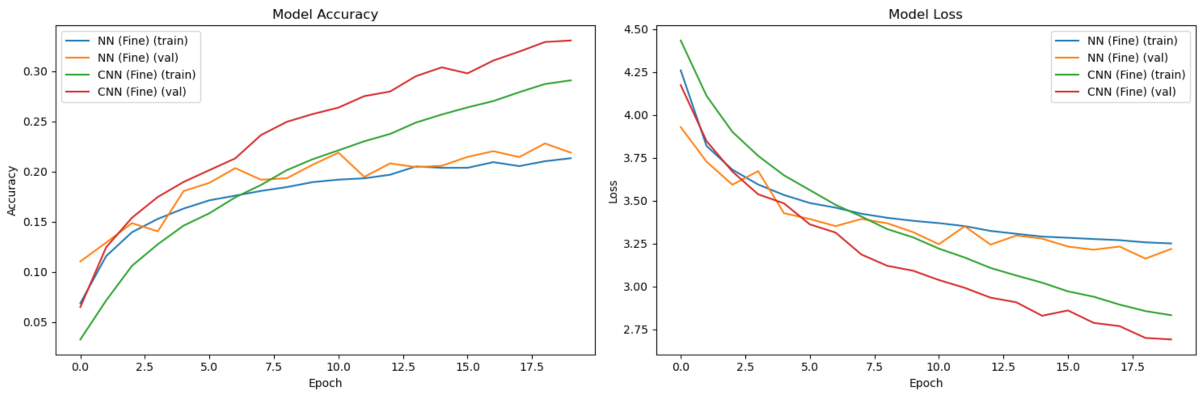

Figure 1 Metrics Comparison Graph

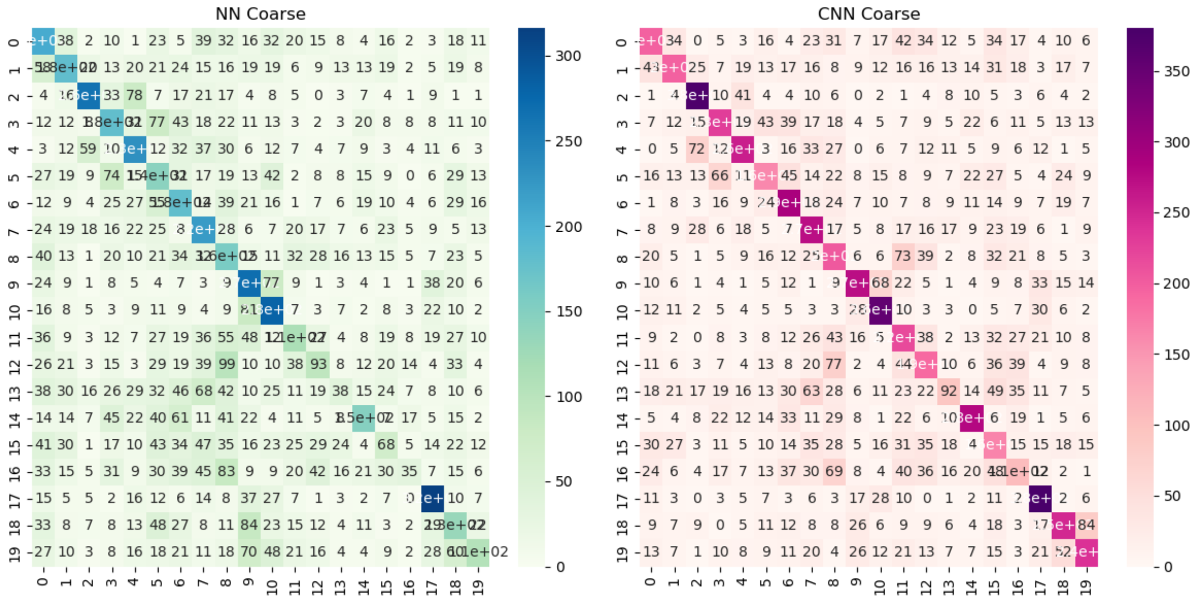

Figure 2 Coarse Confusion Matrices

Conclusion

The study compared the performances of two deep learning approaches, a Fully Connected Neural Network and a Convolutional Neural Network on image classification tasks using both fine-grained (100 classes) and coarse-grained (20 classes). The results demonstrated several key findings. The CNN significantly outperformed the basic Neural Network across both classification tasks. For fine-grained classification, the CNN achieved 32.9% accuracy compared to the Neural Network's 21.6%, exceeding the benchmark of 24.49%. Similarly, for coarse-grained classification, the CNN reached 47.3% accuracy versus the Neural Network's 33.7%, surpassing the benchmark of 39.43%. The superior performance of the CNN can be attributed to its architectural advantages in processing image data. The convolutional layers effectively captured spatial patterns and hierarchical features, as evidenced by the clearer diagonal patterns in the confusion matrices in *Figure 2* above.

This was particularly noticeable in the coarse-grained classification task, where the reduced number of classes and broader categorical distinctions allowed for better feature discrimination. Both models demonstrated better performance on coarse-grained classification, with the CNN showing a 14.4 percentage point improvement and the Neural Network showing a 12.1 percentage point improvement compared to their fine-grained results. This pattern suggests that while both models can effectively learn high-level visual features, the increased complexity of fine-grained classification presents a significant challenge. These findings validate the effectiveness of CNNs for image classification tasks and suggest that hierarchical approaches to classification might be beneficial for future implementations. While there is room for improvement, particularly in fine-grained classification accuracy, the CNN's performance demonstrates the potential of deep learning approaches in complex image recognition tasks.

References

[1] “Papers with Code - CIFAR-100 Dataset,” Papers with Code, [Online]. Available: https://paperswithcode.com/dataset/cifar-100. [Accessed 6 12 2024].

[2] R. F. B. F. Antonio Torralba, “80 Million Tiny Images,” 29 6 2020. [Online]. Available: https://groups.csail.mit.edu/vision/TinyImages/. [Accessed 6 12 2024].