Slide 1 + 2: First Page + Index

Hello and welcome, my name is Yassir Ibararh and today I will be doing a presentation for the final assessment of the machine learning module at Essex university.

We will start with a brief introduction, followed by some exploratory data analysis. Next the data partitioning strategies will be defined and the architecture of the model we are building. The Training strategies and the evaluation will be discussed. Then some possible challenges and improvements. Ending with a conclusion of the key findings and knowledge acquired.

Slide 3: Introduction

The recent rise in social media use around the world, and all the recent advancements in technology, resulted in an explosive growth in image data. It is nowadays one of the main ways that people use in their daily communications, thanks to the richness of its content. However, having this amount comes with its challenges when treating it manually.

With all this amount of data to explore. There was a rise in image recognition programs. They are built on neural networks that can classify images based on certain features and learn from mistakes to make more accurate predictions.

For the final assessment of the machine learning module, we were asked to develop a neural network able to perform image recognition using the CIFAR10 dataset.

CIFAR10 was made specifically for this purpose. It was collected by Alex Krizhevsky and others in 2008. The dataset is a subset of 80 million tiny images dataset that consists of a collection of 60,000 images. It is well varied and contains a total of ten categories such as Truck, automobile, or frog. As it is balanced there are 6,000 images in each of the ten categories.

The main objective here is to build a neural network model that can result in sufficient performance. The training strategy must be suitable, and the model should be evaluated and tested. This is to confirm whether it can accurately classify a certain image to its corresponding category.

Slide 4: Exploratory Data Analysis:

Before beginning to build the model, we started to explore the data and its characteristics. It is a very important process that helps us understand the amounts, quality and shape of the data before carrying out any pre-processing.

At first glance, we can see that the images in the CIFAR10 dataset have a dimension of 32x32 and 3 colour channels RGB that range from 0 to 255.

The graph on the right-hand side confirms the total distribution of the training and testing data on each class and how balanced it is. This will play a crucial role in avoiding overfitting.

Slide 5: Data Pre-processing:

Data pre-processing is a main step in building the model as it plays a key role in ensuring that the input data fed to the first layer is properly optimised and formatted to ensure effective learning.

An article published by Nazri Mohd in 2013 highlighted the effects of pre-processing on artificial neural networks. The research has shown that a good pre-processing can enhance the model's accuracy, prevent biases and errors, and speedup the training process.

The first step was to normalize the data to make the model able to learn effectively. Each pixel value of the image in every channel initially takes a value within the range of 0 to 255. The values were divided by 255 thus scaling the range to between 0 and 1. This makes the data more consistent and suitable for our model.

Additionally, categorical encoding was applied. The process was to convert integer class labels into a one-hot encoded format. This transformation will enable the model to correctly interpret the different classification categories and compute the probabilities for each class.

Slide 6: Data Partitioning:

Data partitioning is also one of the most important steps when building an artificial neural network. It ensures that the model is trained, validated and tested efficiently.

By splitting the data, we can assess the performance of the model on unseen data and observe the results. Having a validation data will be used to monitor the performance during the training phase, which allows us to do some hyperparameters optimization to reduce any overfitting and increase the accuracy.

The initial split from Kaggle included a collection of 50,000 for the training and 10,000 for the testing. An additional 10,000 was split from the training set to be used as the validation dataset.

Slide 7: Model Architecture:

Among the well-known architectures of artificial neural networks, we can find convolutional neural networks. They are the main ones that are used within the field of image and pattern recognitions.

The architecture of CNN is similar to the traditional ANNs. They are made of neurons that can self-optimize during training. Each neuron receives an input, process it and produces an output.

The network will follow a VGG architecture. It is made of 2 sets of layers that are fully connected. Each layer consists of a convolutional layer and a pooling layer.

The convolutional layer will be responsible for extracting the different features within the images. The process is done using filters or kernels that move across the image and identify its features. For our model, we decided to go with a total of 64 filters and a kernel size of 3x3.

The activation function chosen here is the RELU function. It is one of the most popular functions used in convolutional models. It will help our model by introducing non-linearity to be able to abstract the most complex patterns in the data. It will also play a key role in reducing the backpropagation complexity by having two simple outputs of 0 if it’s a negative or 1 if positive.

The output produced is then sent to the pooling layer whose sole purpose is to reduce the training parameters' complexity, by reducing the size of the feature maps. We will initiate the layer with a MaxPooling2D layer with a stride and filter of 2. The activation function chosen here is SoftMax. It will help us transform the output values into probabilities, making it simpler to interpret the model's predictions.

The following layers will be composed of a flattening layer that prepares the images and passes them to a dense layer. This layer contains 256 neurons that will collect the features learned throughout the previous layer and act like a decision maker to correctly classify the data. The activation function chosen here is also RELU due to its classification advantages.

The final layer is the output layer that consists of 10 neurons each corresponding to a class within the categories of that dataset.

Slide 8: Model Trainig:

Training the model is where all the magic happens. It allows the network to learn the different patterns and relationships within a certain dataset. For our model, his will involve adjusting the initial weights depending on the training data prediction starting from the first iteration. The final goal of the process is to minimize the loss function as much as possible and thus improve the accuracy on unseen data.

The loss function used here is the categorical crossentropy which is well suited for multi-class classification. It measures the difference between the predictions and the actual labels, followed by a penalisation of incorrect values. The metric decided is the accuracy to follow the model's learning progress.

Choosing the number of iterations or epochs is quite a critical operation. The reason is that having a big number of epochs can lead to causing overfitting and a costly computation value. However, at the same time, having a very low number of epochs can cause underfitting. The number I have chosen for this model is 25 iterations which was thought to be well sufficient for our network to learn and act well on the data.

Although 25 iterations do not seem to be a very high number, there is always a risk of overfitting. To reduce that risk, we introduced an early stopping. It will be executed to end training if the validation loss value does not improve after a maximum of 3 iterations.

Slide 9: Model Evaluation:

Once the training iterations were completed, we could observe that it has highly increased the validation accuracy from an initial value of 54% to 71% with a loss value of 0.83.

From the image on the right-hand side, we can visualize the accuracy of the predictions based on a plot of 5X5 images. The network successfully predicts most of them right with a small percentage of errors, such as the airplane predicted as a bird or the dog as a deer.

The model then was evaluated on the 10,000 images testing data to see how it performs on unseen data. It has achieved a prediction score of 72% with a loss value of 0.82.

Slide 10: Training/Validation loss & accuracy:

As it can be seen here on both plots, the model has kept a close alignment between the training and the validation accuracy throughout the iterations. This is a good sign that our network is avoiding overfitting while increasing the correct predictions each time.

Slide 11: Confusion matrix:

A confusion matrix was generated to observe the classification performance by showing the predicted values versus the true values.

From the plot we can observe that some classes like class 1 which corresponds to the automobile category has the highest correct predictions. This can be explained by its features that are quite different from the other classes features.

On the other hand, classes like class 0 which corresponds to airplane is getting confused with class 2 that corresponds to bird. Again, this can be explained by both having very similar features.

Slide 12: Single Image test:

Finally, we tested the model on the unseen data to see whether it can predict a single given image. The image chosen was of cat which was successfully predicted to its correct corresponding class.

Slide 13: Strengths and weaknesses:

Although the iterations stopped after only 7 epochs the outputs showcase promising results. The Model was able to successfully extract the relevant features from the images and learn which label they correspond to while maintaining an overall good performance.

On the other hand, there are still some misclassifications to some images with similar features such as the one we showed earlier with the bird and the airplane. However, these could be a common misconception around humans as well as the features of both are very similar which are the tail, head and wings.

Slide 14: Improvements:

Although our first model displays good overall performance. There are a couple of fixes that we can try to improve it.

Data augmentation is a well know technique to increase the quality, volume and diversity of a certain dataset. It can help us here by increasing the model’s ability to pick on more variable features thus increasing robustness and accuracy. A survey made by Alhassan Mumuni & others in 2022 shown that different data augmentation techniques can all improve the performance in image classification tasks.

In additions, for our model we decided to use RELU activation function for the convolutional layer and the SoftMax function for the pooling layer due to how well suited they are for the task in hand.

However, there are a couple of other options to explore. For example, one of the weaknesses of RELU is that is that the neurons can become inactive when they consistently output zeros for all inputs. Ritesh Maurya and others in 2023 discussed how functions such as ELU and Leaky-RELU can be combined with it to ensure continuous learning and having a more efficient activation function.

Another way to improve model performance is the learning transfer. It can be done by leveraging pre-trained well know models such as Resnet. It is a residual learning framework trained on the large datasets such as ImageNet. The image on this slide shows the high accuracy of the Resnet model trained on a similar dataset.

Other suggestions to improve the overall performance can be hyperparameter optimization, by changing the values of the number of neurons in each layer or increasing the architecture depth by adding more layers.

Slide 15: Conclusion:

Image recognition models have highly impacted the modern world in various industries. Sectors like healthcare or transportation nowadays heavily rely on deep learning networks for image classification. For example, in healthcare, the main categories of applications involve treatment recommendation and diagnosis.

An article written by Thomas Davenport and others for the healthcare journal in 2019 showcased how radiologist employ image recognition models for tasks such as brain tumour detection or test analysis. Although the implementations are not completely autonomous yet due to the sensitivity of the sector.

For the future, as datasets become much bigger and hardware improves considerably, there is a very high potential for how these models can improve the overall service.

The artificial neural network we built here displays the potential of convolutional neural networks in object recognition. By utilizing the CIFAR10 dataset we were able to see how the network’s accuracy evolves. The model then was tested using unseen data which gave us an idea on how it can be employed for more complex data and visual patterns.

Although our model is not at a desirable level of accuracy yet, the improvements proposed can take the prediction's accuracy close to 99%. Minimising the loss value to a maximum level and making it scalable will be a key factor in making it applied to real-world applications.

Going through the experience of building a neural network and evaluating it helped me gain the necessary knowledge required to develop efficient image recognition systems.

It is safe to say that there is more appreciation for the complexity of the human mind and how it can process enormous amounts of information in detail and with great accuracy.

Thank you for watching. I’m happy to take any of your questions.