Project 1: Image Classification Using Convolutional Neural Networks

Zhang Xiaoyang
Chalmers University of Technology
xiazhan@chalmers.se

Abstract

The goal of this project is to develop a convolutional neural network (CNN) model to classify tree bark images from the TRUNK12 dataset. The dataset consists of images from 12 different tree species. The project involves designing a CNN, training it, and implementing improvements such as data normalisation, batch normalisation, RELU, dropout layers and introducing more convolutional filters to enhance the model's performance. This report outlines the methodology, experimental evaluation, ablation study and results of the image classification task.

1 Method

1.1 Network Architecture

The basic CNN architecture used in this project consists of: an input layer for 40x40x3 colour images, two convolutional layers with 8 and 16 filters respectively, each followed by max pooling layers, a fully connected layer with 64 neurons, and an output layer with 12 neurons for 12 classes, followed by a softmax activation layer.

1.2 Improvements

Improvements to the model were introduced at each stage progressively, starting with normalising the training and test images to mean 0 and variance 1. The purpose is to normalise images to the range [0, 1] initially and then to mean 0 and variance 1. Different pixel intensity ranges can cause instability during training. Normalising the data ensures that the network processes inputs consistently, leading to faster and more stable convergence.

Subsequently, I have increased to 3 convolutional layers with increasing filter sizes (32, 64, 128), each followed by batch normalisation, Rectified Linear Unit (RELU) activation and max-pooling layers. Batch normalisation aims to address the problem of internal covariate shift, where the distribution of each layer's inputs changes during training. Batch normalisation normalises the output of each layer, stabilising and speeding up the training process by allowing higher learning rates and reducing sensitivity to initialization.

ReLU helps to introduce non-linearity into the model, allowing the network to learn complex patterns. Besides, it also addresses the vanishing gradient problem, where gradients become too small for effective learning. ReLU helps in mitigating this by allowing gradients to flow more effectively through the network, speeding up the training and improving convergence.

Dropout layer with a dropout rate of 0.5 was also added to prevent overfitting, where the model performs too well on training data but poorly on unseen data. It randomly deactivates a fraction of neurons during training, forcing the network to learn more robust features that generalise better to new data.

2 Experimental Evaluation

2.1 Training Perimeters

- Optimizer: SGDM
- Initial Learning Rate: 0.01 for SGDM
- Max Epochs: 20.
- Shuffle: Every epoch.
- test Frequency: 15 iterations.

2.2 Results

2.2.1 Basic CNN Model

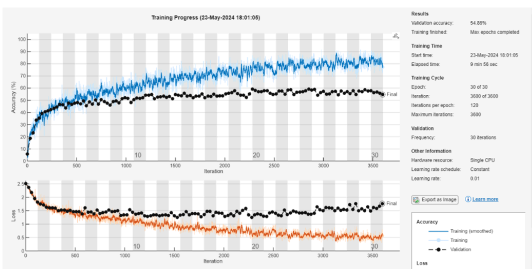


Fig 1

(Fig 1) Over 30 epochs of the training process, training progress increases more steeply at the beginning, followed by gradual increase. This indicates the training data is learning well by achieving an accuracy close to 80% at epoch 30. The test accuracy curve also shows improvement but at a slower rate. The final test accuracy is about 54.88%, indicating that while the model performs well on the training data, its performance on unseen test data is considerably lower.

This gap between training and test accuracy suggests overfitting, where the model learns the training data's specific patterns too well but fails to generalise to new data.

Similarly, the final test loss remains higher than the training loss, further indicating that the model may be overfitting the training data.

As such, this highlights the need for further improvements are required to enhance its generalisation capability. The subsequent experiments will focus on implementing and evaluating these enhancements.

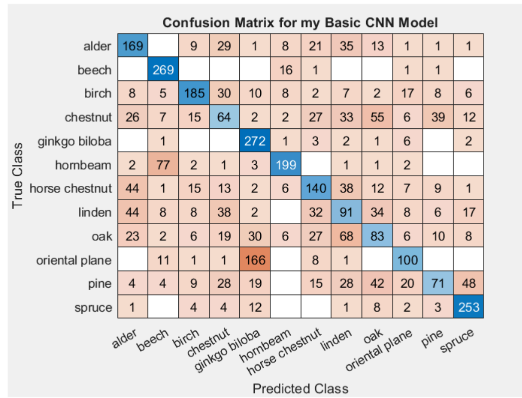


Fig 2

In addition, from the confusion matrix (Fig 2) we can see that although generally the model is able to predict the correct class more than half the time, there are instances whereby it is way off. For instance, for oriental plane, it was wrongly predicted as ginkgo biloba 166 times, which is much higher than the correct prediction count which is 100 only.

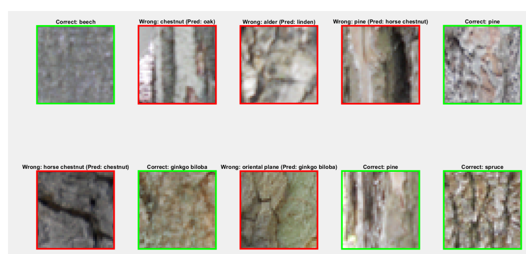


Fig 3

This figure (Fig 3) shows that for the given 10 samples, it seems like half of the predictions are wrong.

2.2.2 Basic CNN with Normalised Data

The same basic CNN model is trained after preprocessing the data through normalising the images to have a mean of 0 and a variance of 1.

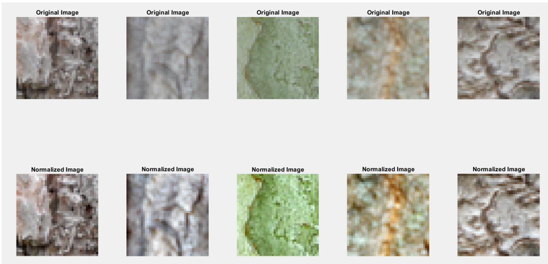


Fig 4

Through normalisation (Fig 4), we can see that normalised images appear more consistent in terms of brightness and contrast, whereby they have enhanced contrast, making features more discernible.

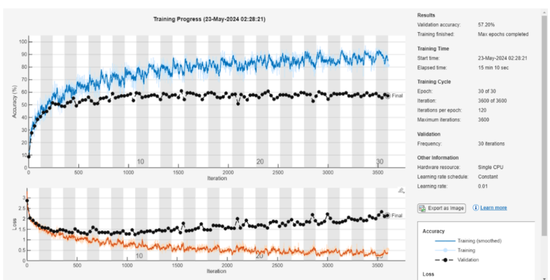


Fig 5

(Fig 5) The training accuracy reaches slightly higher than the result from the basic model without normalised data, and the test accuracy improves slightly to 57.2%. Although the normalised data model shows a modest improvement in test accuracy, the gap between training and test accuracy is larger in the normalised data mode, suggesting that while normalisation helps in some aspects, it may also lead to overfitting on the training data.

Training loss is similar for both, but in terms of test loss, for normalised data model, the test loss stabilises at a slightly higher value, further suggesting it is more prone to overfitting.

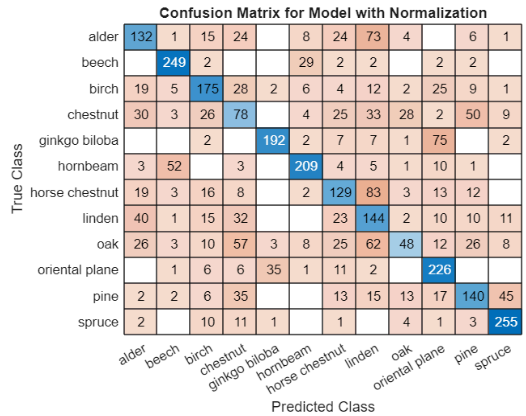


Fig 6

Overall for the confusion matrix (Fig 6), classes like beech, hornbeam, oriental plane, and spruce have high values on the diagonal, indicating that the model performs well in classifying these categories. However, alder is frequently misclassified as beech, and linden is often misclassified as oak.

Possible reasons for increased overfitting on normalised data could be due to increased sensitivity to noise. Although normalisation generally stabilises and speeds up training, it can also make the model more sensitive to small variations and noise in the training data. This increased sensitivity can lead the model to learn noise and specific patterns in the training data more effectively, resulting in overfitting. As such, subsequently, I introduced regularisation techniques in conjunction with other features to reduce overfitting and at the same time, improve performance.

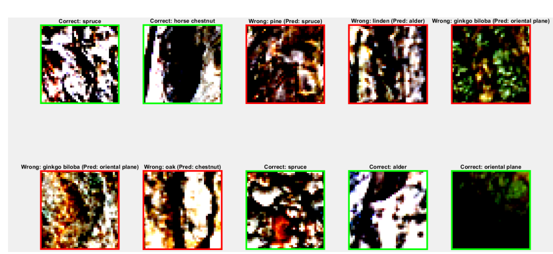


Fig 7

2.2.3 Enhanced CNN model with Normalised Data

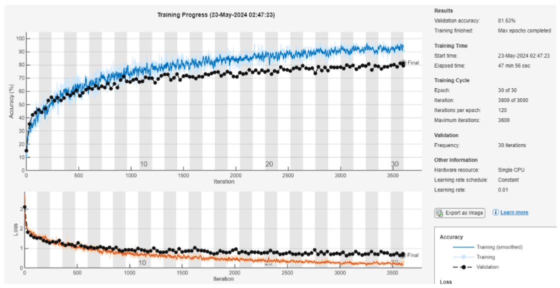


Fig 8

The improved CNN model incorporates several enhancements to address overfitting and improve generalisation. These enhancements include batch normalisation, ReLU activation, dropout layer, increased convolutional filters. For this enhanced CNN model, training accuracy reaches close to 90% whereas test accuracy reaches 81.63%, which is a significant improvement compared to the previous models. (Fig 8)

In addition, this enhanced model also shows lower and more stable training and test loss curves, indicating more efficient learning and better generalisation.

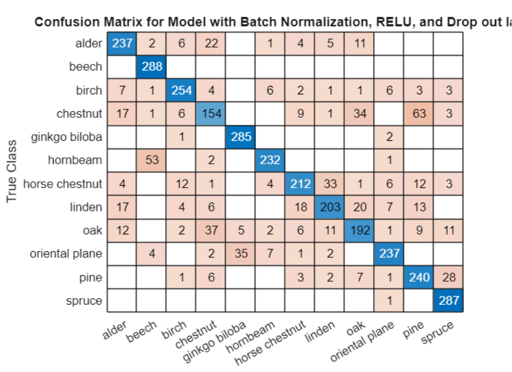


Fig 9

In terms of the confusion matrix (Fig 9), there is a high score across the diagonal of all classes indicating generally good performance, much better compared to the previous model.The enhanced model has met the improvements described in the "Method" section. The addition of batch normalisation, ReLU activation, dropout layers, and increased convolutional filters and layers collectively contributed to stabilised and accelerated training, mitigated vanishing gradient problem, reduced overfitting, and improved feature extraction and classification performance.

The actual results, including higher test accuracy, lower test loss, and a more balanced confusion matrix, validate the effectiveness.

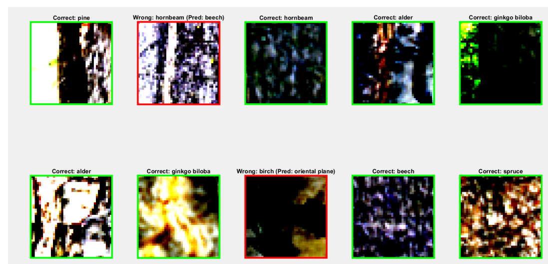


Fig 10

2.2.4 Ablation Study

To demonstrate the importance of each enhancement, we performed an ablation study by removing one improvement at a time and evaluating the resulting model.

2.2.4.1 Enhanced CNN Model Without Data Normalisation

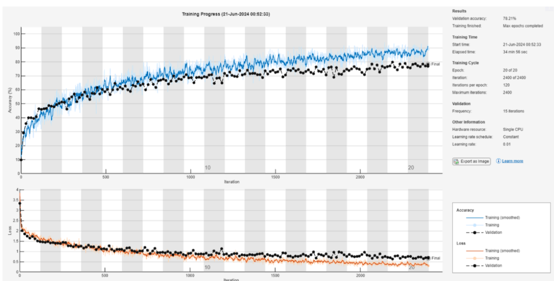


Fig 11

From the results (Fig 11), the model without data normalisation achieves a test accuracy of 78.21%, which is lower than the enhanced model's accuracy of 81.63%. Data normalisation helps stabilise the training process by ensuring consistent input ranges, which leads to faster convergence and better generalisation. Without normalisation, the model becomes sensitive to varying pixel intensities, leading to poorer performance. In this case, the difference in accuracy not being high could be due to the inherent characteristics of the dataset that makes it somewhat resilient to variations in pixel intensity ranges. Since the images have consistent lighting conditions and contrasts, lack of data normalisation does not significantly affect the model's ability to distinguish features.

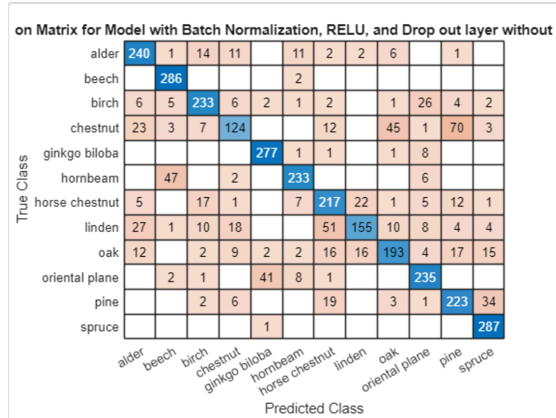


Fig 12

As for the confusion matrix (Fig 12), the accuracy is generally high across all classes. Compared to the enhanced CNN with normalised data, this model has similarly high correct classifications for beech, ginkgo biloba, and spruce, as well as common misclassification patterns, like Hornbeam misclassified as beech.

This suggests that the inherent features of these classes are the primary cause of confusion, rather than the choice of optimiser.

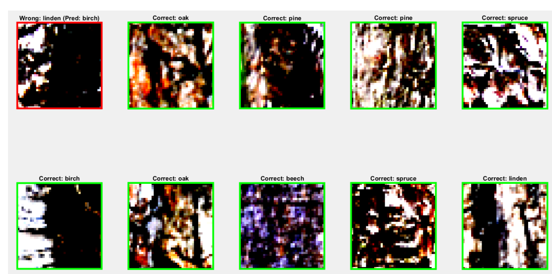


Fig 13

2.2.4.2 Enhanced CNN Model Without ReLU activation

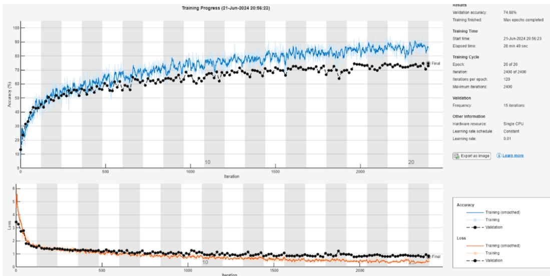


Fig 14

From the results (Fig 14), the model without ReLU activation layer achieves a test accuracy of 74.68%. The model without ReLU activation shows a significant drop in test accuracy to 74.68%, compared to the 81.63% of the enhanced CNN model.

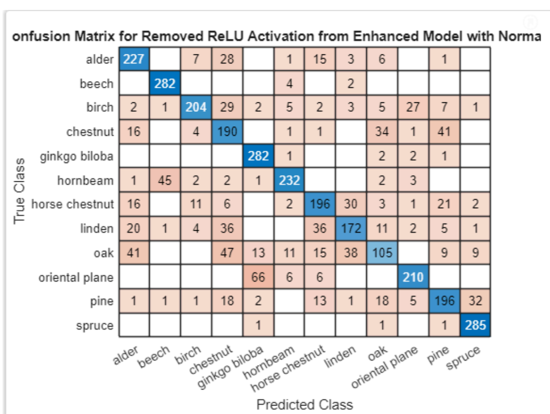


Fig 15

The confusion matrix (Fig 15) reveals more frequent misclassifications compared to the enhanced model, particularly when classifying oak and linden. ReLU activation introduces non-linearity into the network, allowing it to learn complex patterns. It also mitigates the vanishing gradient problem, which is crucial for deep networks. Without ReLU, the model struggles to capture these patterns, leading to reduced performance.

Fig 16

2.2.4.3 Enhanced CNN Model Without dropout layer

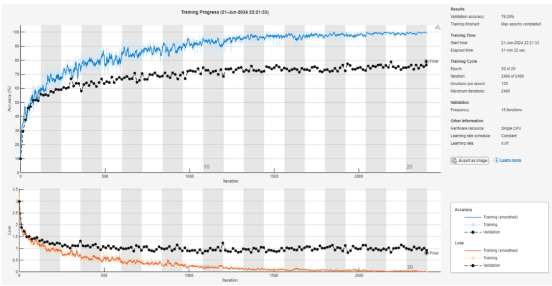
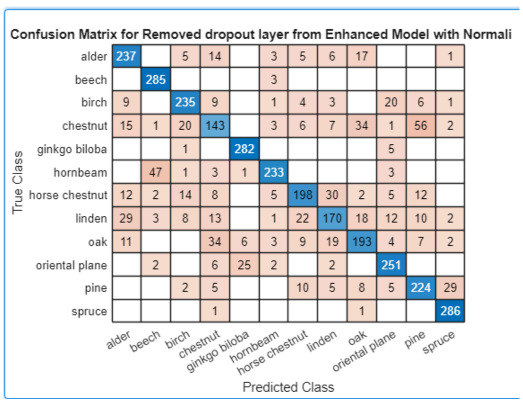


Fig 17

From the results (Fig 17), the model without dropout layer achieves a test accuracy of 79.2%. Although the test accuracy may seem to close to the test accuracy of the enhanced CNN at 81.63%, in this case the training accuracy is a lot higher than the test accuracy, indicating that there is greater overfitting.



Confusion Matrix for Removed dropout layer from Enhanced Model with Normali

True Class \ Predicted Class	alder	beech	birch	chestnut	ginkgo biloba	hornbeam	horse chestnut	linden	oak	oriental plane	pine	spruce
alder	237		5	14		3	5	6	17			1
beech		285				3						
birch	9		235	9		1	4	3		20	6	1
chestnut	15	1	20	143		3	6	7	34	1	56	2
ginkgo biloba			1		282					5		
hornbeam		47	1	3	1	233				3		
horse chestnut	12	2	14	8		5	198	30	2	5	12	
linden	29	3	8	13		1	22	170	18	12	10	2
oak	11			34	6	3	9	19	193	4	7	2
oriental plane		2		6	25	2		2		251		
pine			2	5			10	5	8	5	224	29
spruce				1					1			286

Fig 18

The confusion matrix (Fig 18) also indicates increased overfitting, with more misclassifications in specific classes. Dropout helps prevent overfitting by randomly deactivating neurons during training, forcing the network to learn more robust features. Without dropout, the model tends to overfit the training data, resulting in poorer generalisation.

Fig 19

2.2.4.4 Enhanced CNN Model Without Increased Convolutional Filters

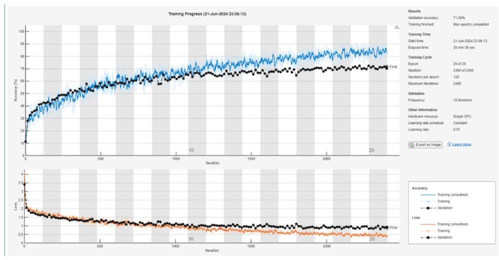
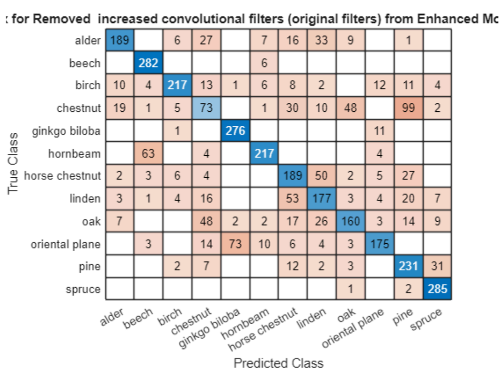


Fig 20

From the results (Fig 20), the model without increased convolutional filters exhibits the lowest test accuracy of only 71.5%.



The confusion matrix (Fig 21) shows significant performance degradation, especially in classes with complex textures. Increased convolutional filters allow the network to capture more detailed features and patterns in the images. By reducing the number of filters, the model's capacity to learn these features diminishes, resulting in lower accuracy and higher misclassification rates.

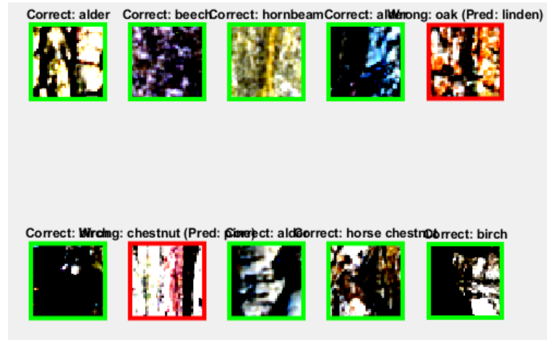


Fig 22

3 Conclusion

The ablation study highlights the importance of each enhancement in the final CNN model. Data normalisation stabilises training and improves generalisation. ReLU activation is crucial for learning complex patterns and mitigating the vanishing gradient problem. Dropout layers help prevent overfitting by promoting robust feature learning. Increased convolutional filters and layers enhance the model's ability to capture detailed patterns in the images. The enhancements led to significant improvements in model stability and accuracy, eventually reaching high test accuracy around 81%. Future work could focus on advanced feature engineering, data augmentation, and experimenting with more complex model architectures to further enhance classification accuracy and generalisation.