Assignment 2

Build and improve a CNN for Image Classification

Student Name: Abdullah Jamal AlHarriem

Registration Number: 12217437

Report on CNN Modifications for CIFAR-10 Classification

Introduction

This report summarizes the modifications applied to a Convolutional Neural Network (CNN) for image classification on the CIFAR-10 dataset. The objective was to build, train, and improve the performance of the CNN through various enhancements. The evaluation criteria include model accuracy, validation loss, and overall generalization.

Modifications and Their Effects

Several modifications were applied to the baseline model, and their effects on performance were analyzed.

1. Baseline Model

- Architecture: 2 convolutional layers, max-pooling, fully connected layers with ReLU activation, and softmax output.
- Performance:
  - Training Accuracy: 77.35%
  - Validation Accuracy: 68.58%
  - Test Accuracy: 67.99%
- Observation: Overfitting was noticeable as training accuracy was higher than validation accuracy.

2. Model with Dropout

- Modification: Added dropout layers (25% after convolutional layers, 50% before final dense layer) to reduce overfitting.
- Performance:

    - Training Accuracy: 64.23%
    - Validation Accuracy: 67.95%
    - Test Accuracy: 67.31%
- Observation: Dropout slightly reduced overfitting but also led to lower training accuracy, indicating that more epochs might be needed to compensate for the regularization.

3. Deeper Model with Batch Normalization
- Modification: Increased the number of convolutional layers and added batch normalization for better gradient flow.
- Performance:
    - Training Accuracy: 93.44%
    - Validation Accuracy: 74.98%
    - Test Accuracy: 73.64%
- Observation: The model showed significant improvement in accuracy, suggesting that deeper architectures and batch normalization stabilize training.

4. Model with Data Augmentation
- Modification: Used image augmentation (rotations, shifts, flips) to improve generalization.
- Performance:
    - Training Accuracy: 57.53%
    - Validation Accuracy: 63.33%
    - Test Accuracy: 64.39%

- Observation: The model had lower training accuracy, but generalization improved slightly. Training for more epochs could yield better results.

5. Combined Model (Regularization + Augmentation + Deeper Layers)

- Modification: Added L2 regularization, batch normalization, dropout, and data augmentation.
- Performance:
  - Training Accuracy: 62.57%
  - Validation Accuracy: 66.57%
  - Test Accuracy: 65.94%
- Observation: Regularization prevented overfitting, but accuracy was still lower than the deeper model with batch normalization.

6. Best Model (Deeper Network + Optimized Hyperparameters)

- Modification: Used a deeper architecture with 4 convolutional blocks, dropout, batch normalization, and AdamW optimizer.
- Performance:
  - Training Accuracy: 92.78%
  - Validation Accuracy: 89.16%
  - Test Accuracy: 89.29%
- Observation: This model achieved the best results, demonstrating that deeper networks with proper regularization and learning rate scheduling improve performance.

Challenges Encountered

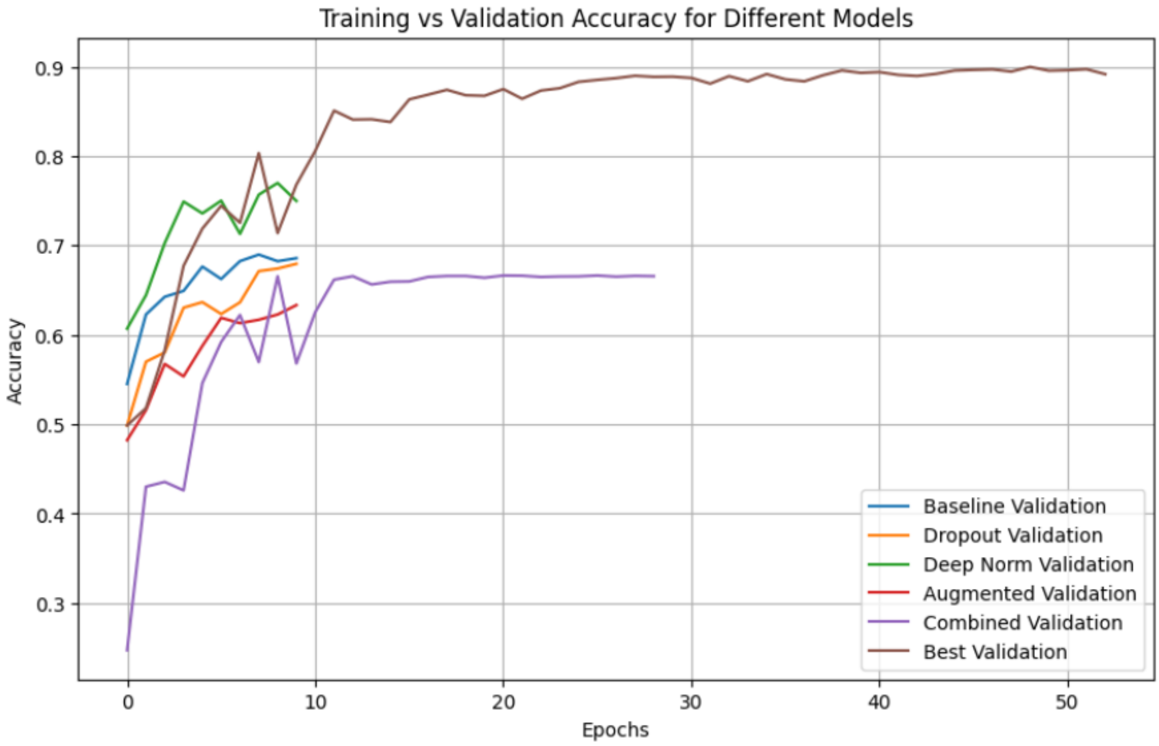
1. Overfitting:
    - Early models showed overfitting, with training accuracy much higher than validation accuracy.
    - Solution: Applied dropout, batch normalization, and data augmentation to improve generalization.
2. Training Stability:
    - Some models had unstable validation loss.
    - Solution: Used learning rate scheduling (Cosine Decay, ReduceLROnPlateau) to stabilize training.
3. Computational Cost:
    - Deeper models required more computation and memory.
    - Solution: Used mini-batch training and early stopping to optimize training time.

Visualizations

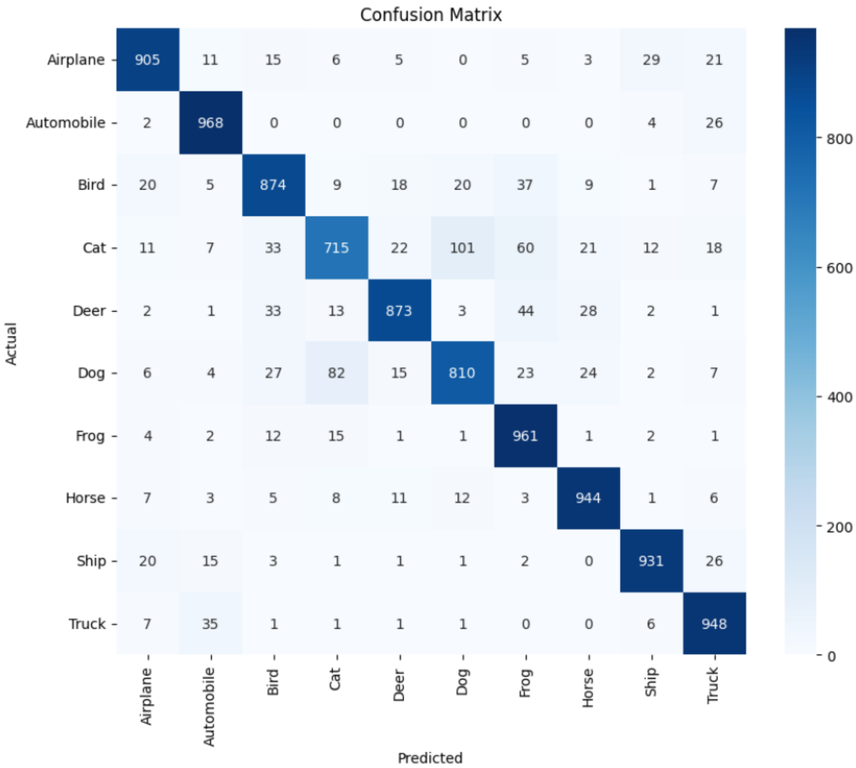
- Accuracy/Loss Curves: Plotted training and validation accuracy/loss for all models.
- Confusion Matrices: Analyzed misclassification patterns in the best model.
- Sample Predictions: Compared correct and incorrect predictions to identify common failure cases.

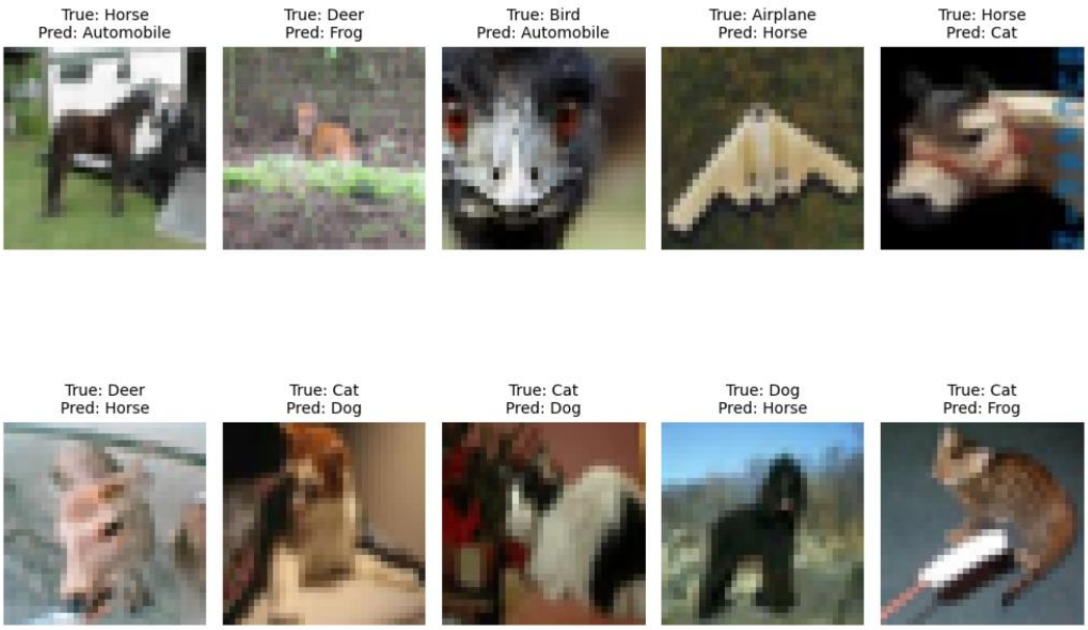
confusion matrices, sample predictions and curves visualizations :

all models :

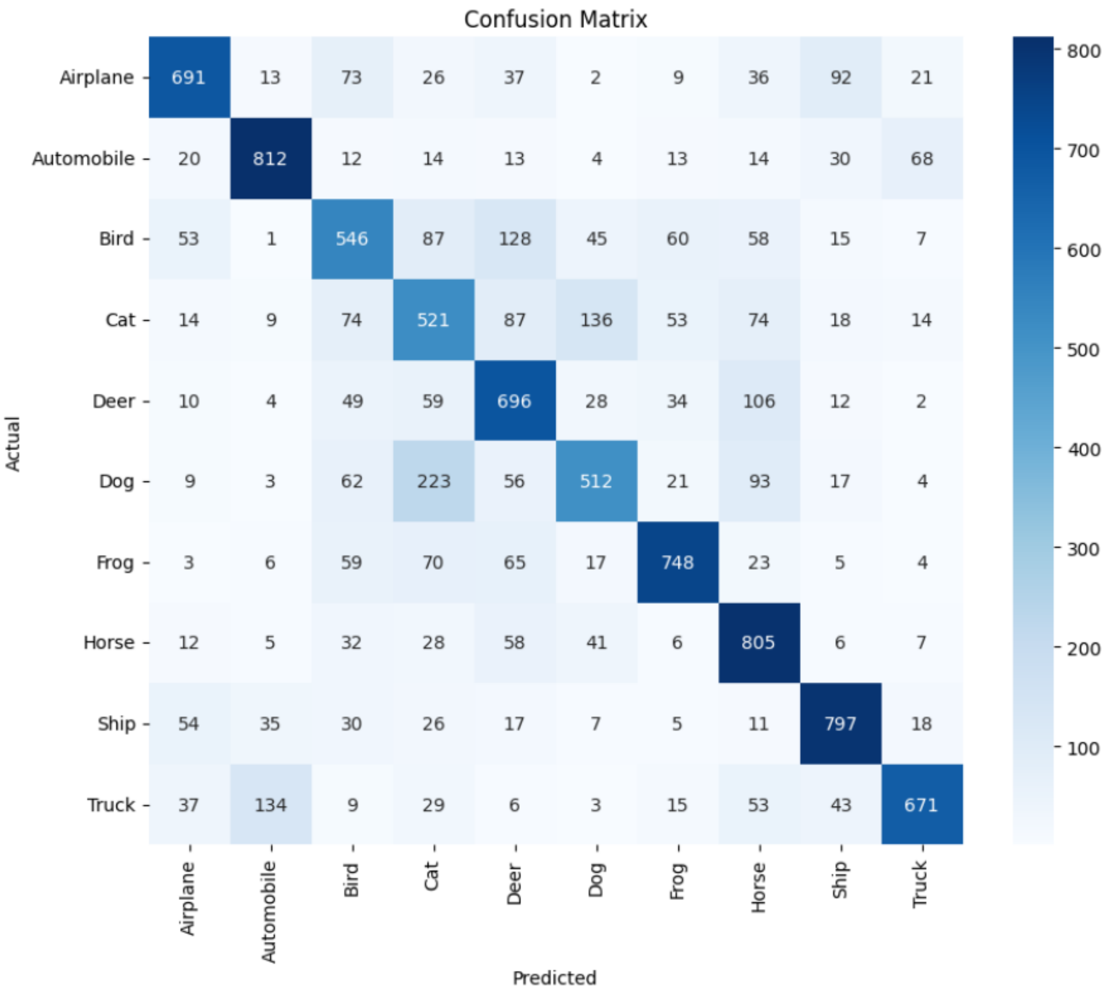


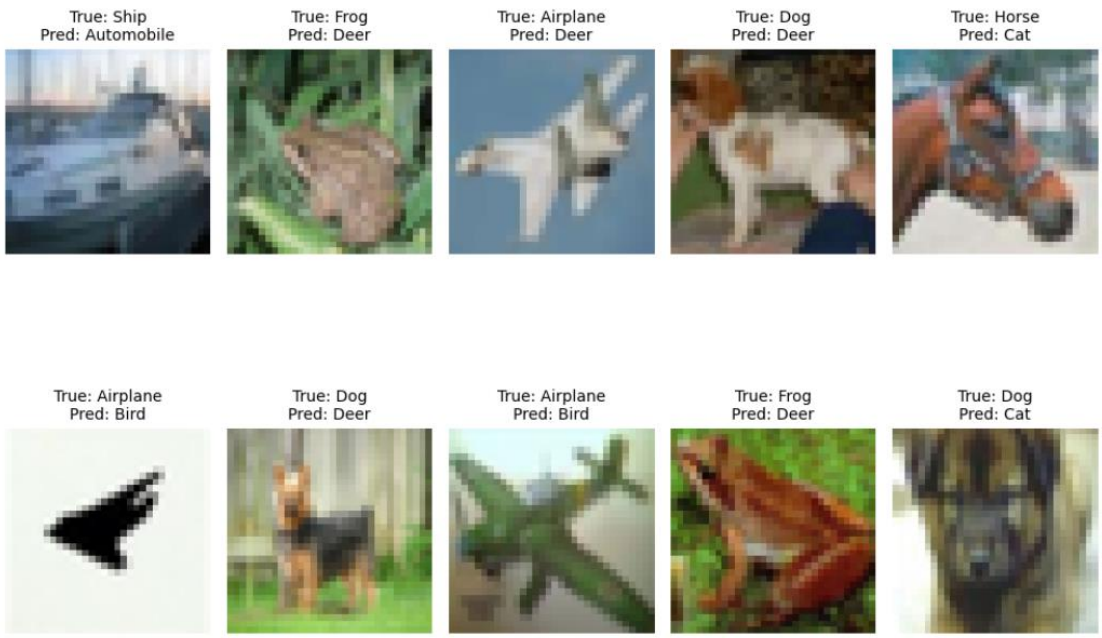
best model :


Confusion Matrix
Actual
Predicted
Airplane
Automobile
Bird
Cat
Deer
Dog
Frog
Horse
Ship
Truck
905 11 15 6 5 0 5 3 29 21
2 968 0 0 0 0 0 0 4 26
20 5 874 9 18 20 37 9 1 7
11 7 33 715 22 101 60 21 12 18
2 1 33 13 873 3 44 28 2 1
6 4 27 82 15 810 23 24 2 7
4 2 12 15 1 1 961 1 2 1
7 3 5 8 11 12 3 944 1 6
20 15 3 1 1 1 2 0 931 26
7 35 1 1 1 1 0 0 6 948
800
600
400
200
0


True: Horse
Pred: Automobile
True: Deer
Pred: Frog
True: Bird
Pred: Automobile
True: Airplane
Pred: Horse
True: Horse
Pred: Cat
True: Deer
Pred: Horse
True: Cat
Pred: Dog
True: Cat
Pred: Dog
True: Dog
Pred: Horse
True: Cat
Pred: Frog

baseline model :





Conclusion

This study demonstrated that increasing model depth, adding batch normalization, using dropout, and applying data augmentation improved CIFAR-10 classification. The best-performing model achieved 89.29% test accuracy. Future improvements could involve fine-tuning architectures like ResNet or EfficientNet for even better performance.