COMPARATIVE ANALYSIS OF CNN ARCHITECTURES USING BLACK BOX ATTACK DATASET

Rebecca Liu[1], Claudia Pacori[1], Solenne Wolfe[1]

[1]Thayer School of Engineering

DARTMOUTH ENGINEERING

INTRODUCTION

Power systems face growing challenges from cyber threats as they become increasingly digitized and interconnected. Phasor Measurement Units (PMUs) provide critical synchronized data for Wide Area Monitoring, Protection, and Control (WAMPAC) systems but introduce potential security vulnerabilities.

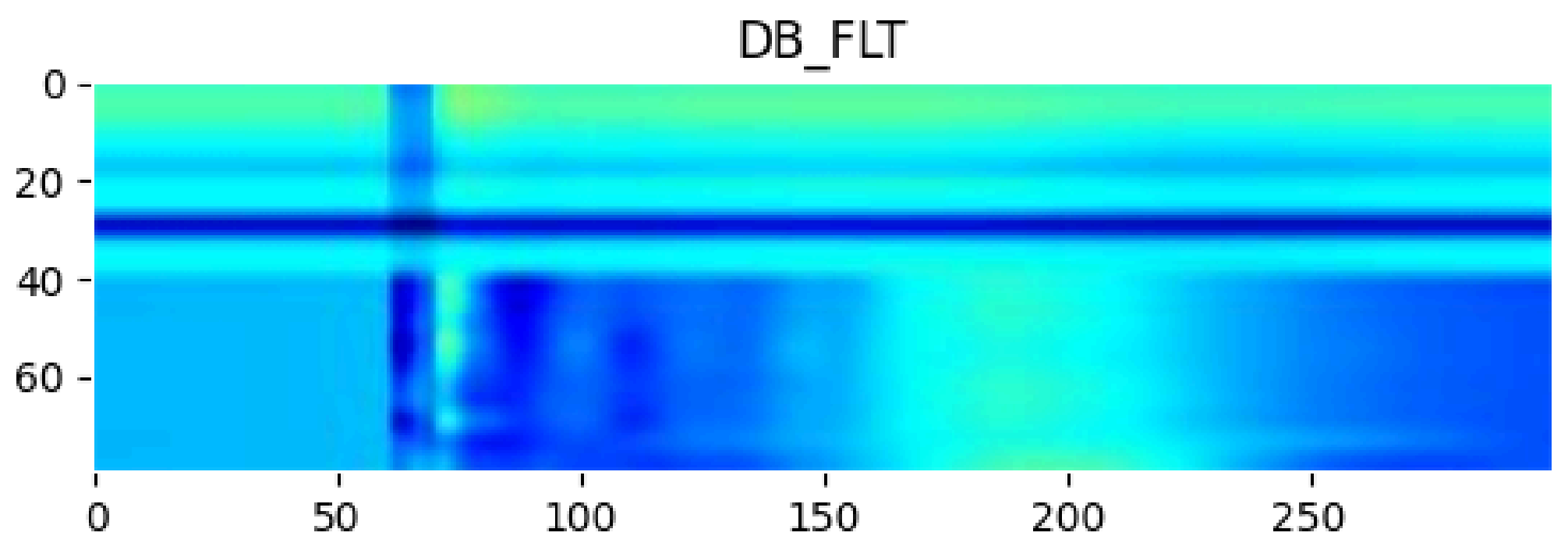
Modern grids must contend with multiple challenges including renewable energy integration, decentralized generation, and increasing demand, all while maintaining security against sophisticated attacks. Black box attacks represent a particularly concerning threat vector, where adversaries are able to manipulate system data without detailed knowledge of the underlying mechanisms.

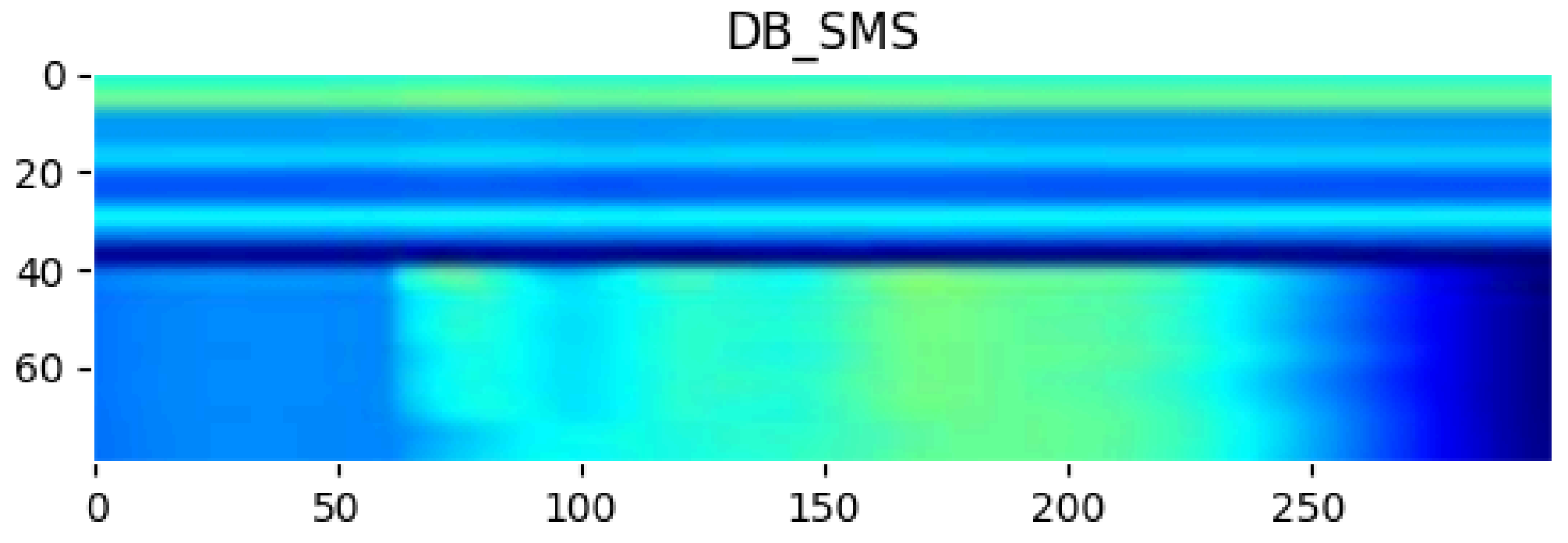
Our research evaluates the comparative performance of three convolutional neural network (CNN) architectures—the "base CNN", ResNet50 v2, and VGG16—against black box attacks targeting PMU data. CNNs can detect anomalies that might evade traditional monitoring systems. With that, we aim to identify optimal approaches for enhancing power system security against these threats.

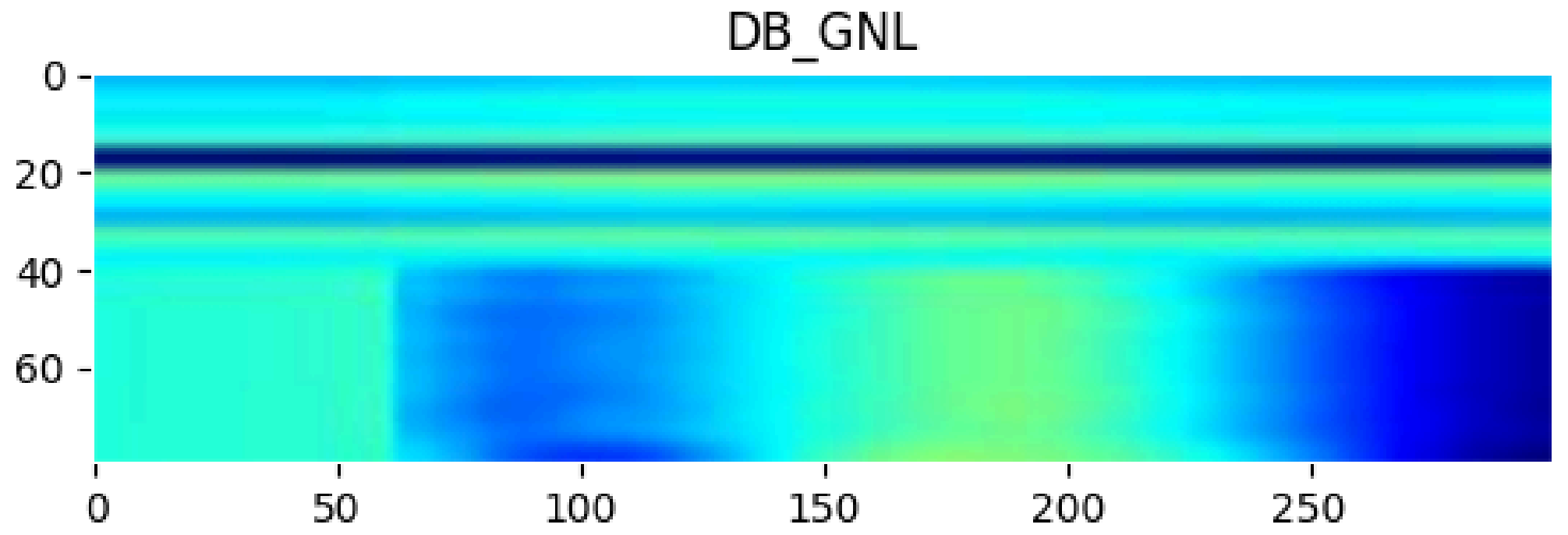
DATASET

This dataset captures power grid disturbance patterns represented as 300×20×3 images, showing 5 seconds of synchronized measurements from 10 PMUs, capturing voltage and frequency changes and using 3 color channels to encode measurement intensity. The dataset is composed of:

- 344 fault instances (DB_FLT)
- 140 generation loss events (DB_GNL)
- 21 motor switching events (DB_SMS)



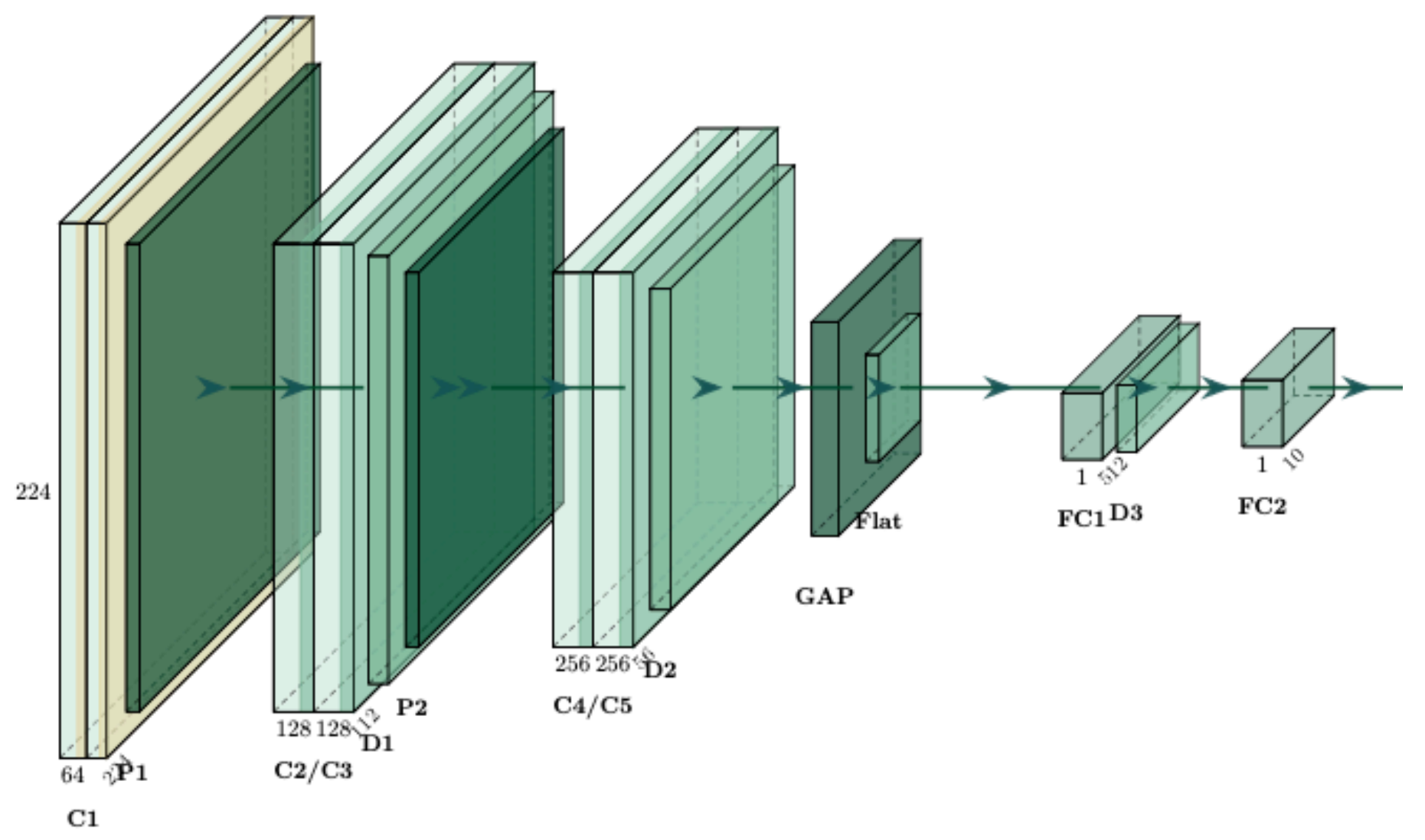




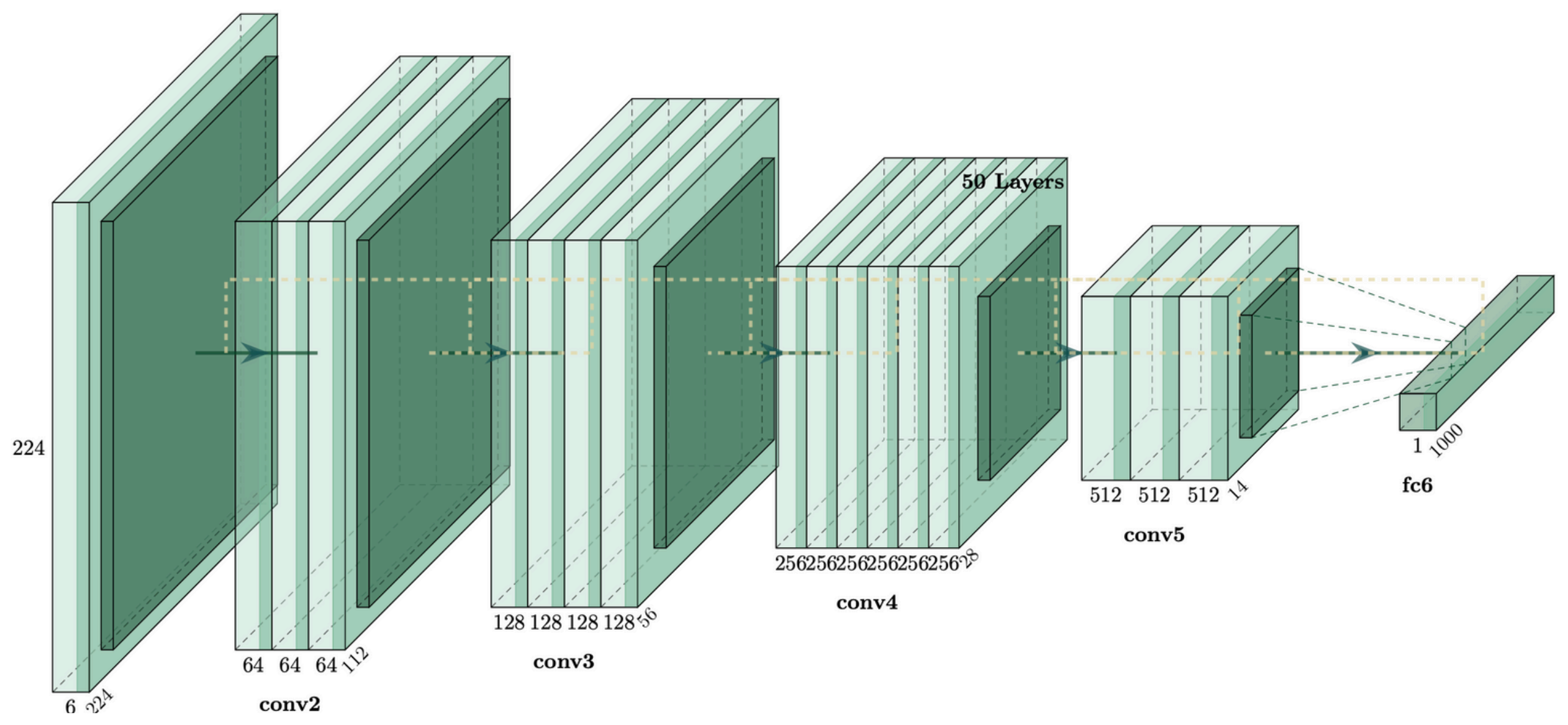
CNN ARCHITECTURES

Selected models offer a balanced comparison: **Base CNN** as a benchmark [1], **ResNet50 v2** for deep learning optimization with residual connections [4] , and **VGG16** for strong feature extraction [3].

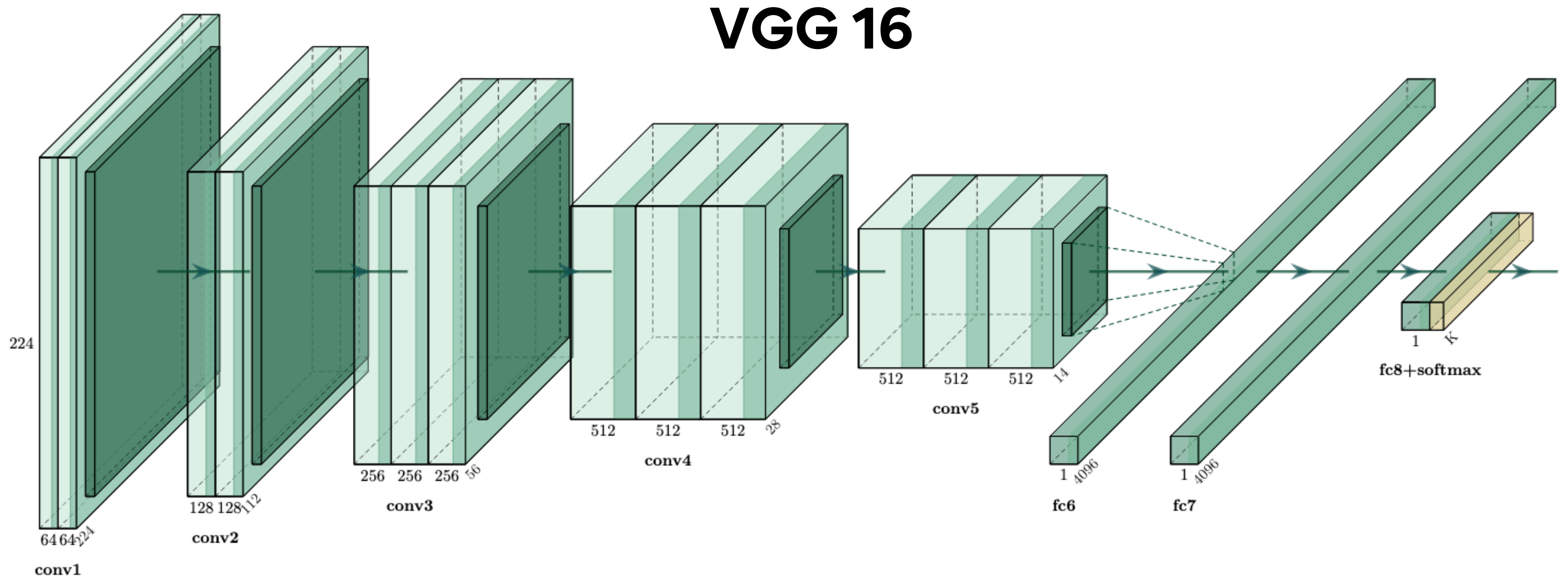
BASE CNN



RESNET 50 v2



VGG 16

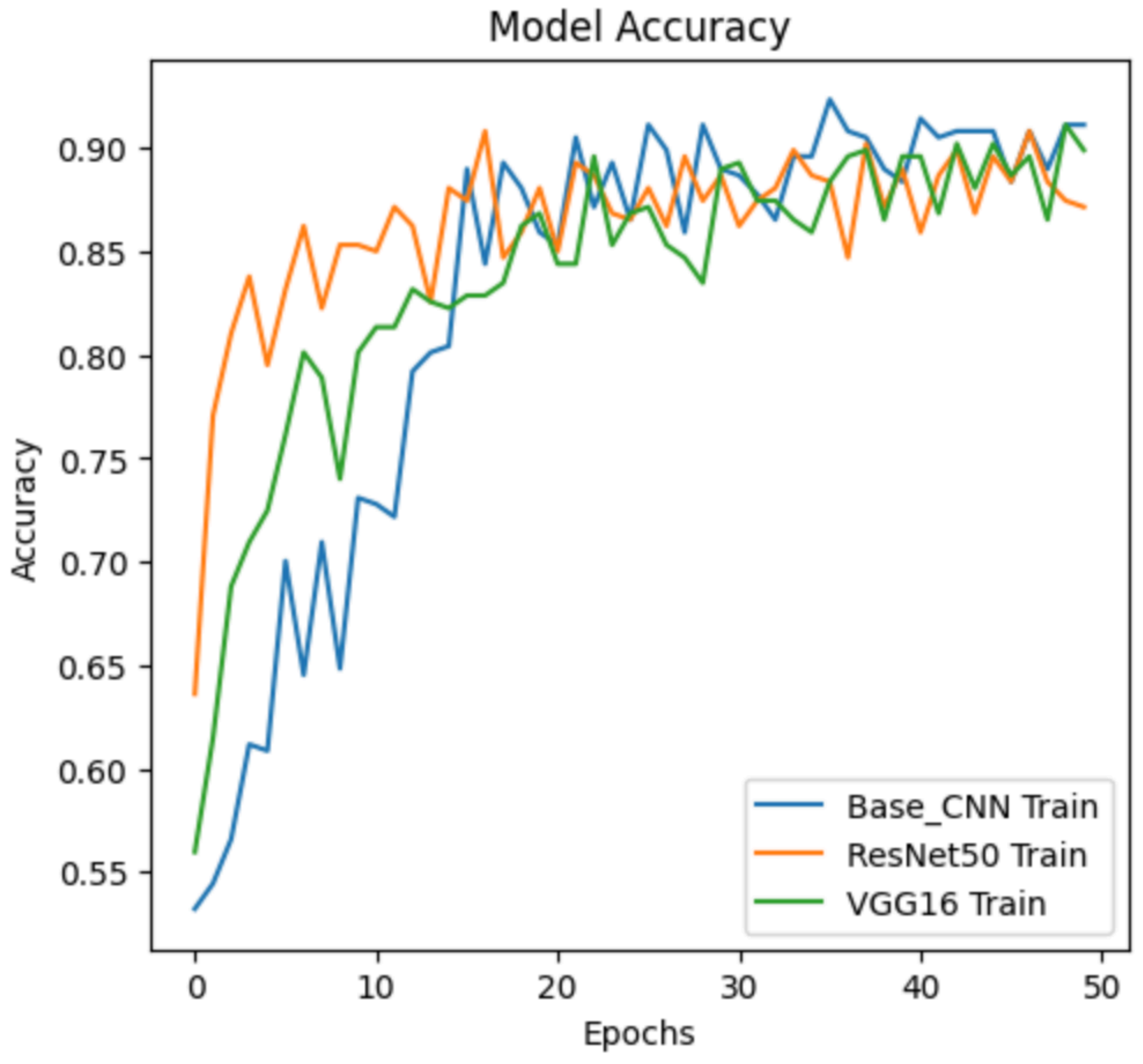


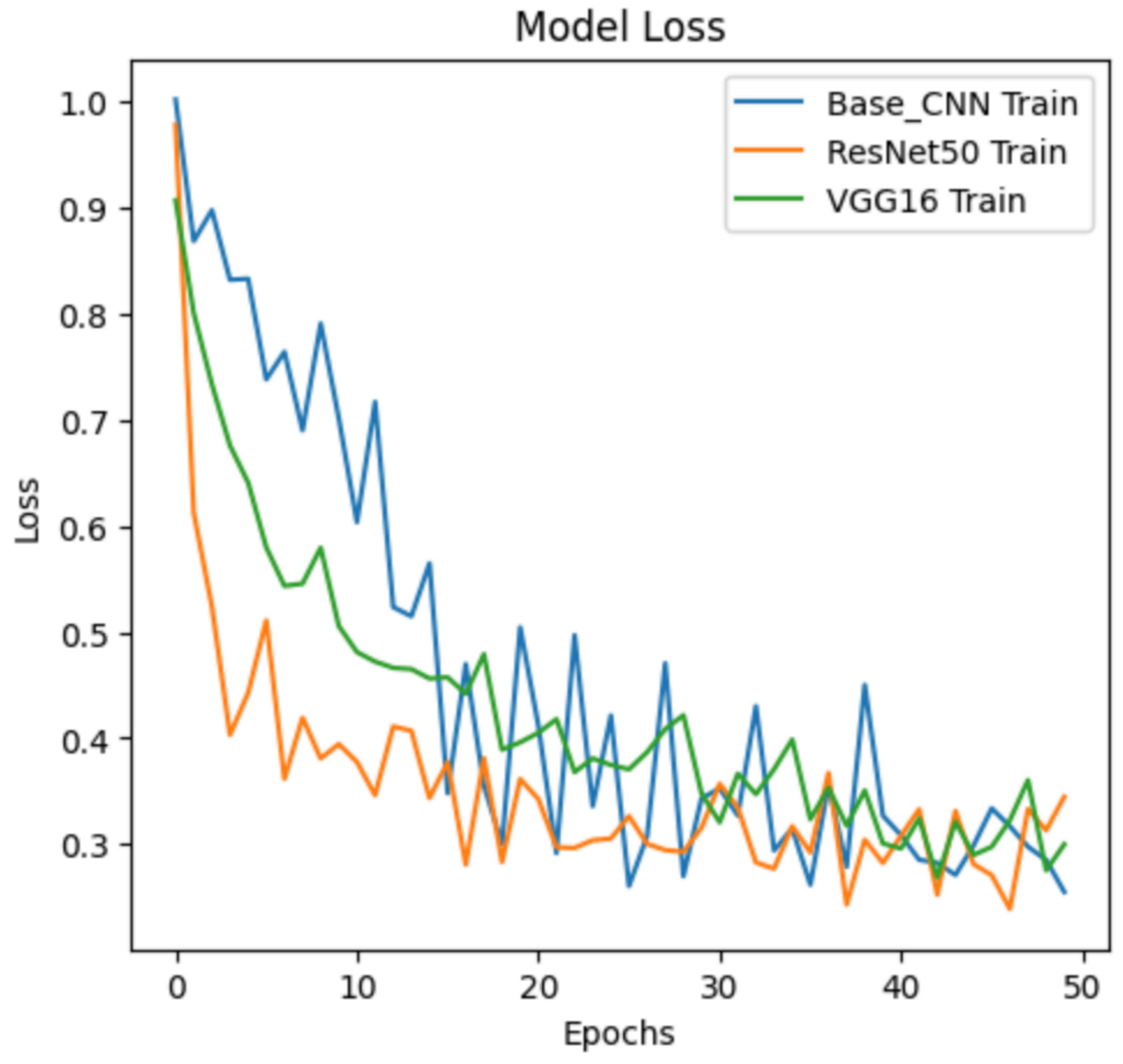
Legend:

C1-C5	Convolution layers
BN	Batch Normalization
P1-P2	Max Pooling
D1-D3	Dropout
GAP	Global Average Pooling
Flat	Flatten
FC1-FC2	Fully Connected

RESULTS

The graphs below illustrate the **accuracy** and **loss** evolution during the training phase for each model over 50 epochs:





DISCUSSION

Each model presented distinct strengths and trade-offs in accuracy, stability, and generalization:

- **Base CNN** shows solid performance and achieves the highest accuracy, but with higher variability, indicating limited feature extraction capabilities. It required fewer computational resources, making it efficient but potentially less robust for complex patterns.
- **ResNet50** achieves high accuracy early with good stability, benefiting from residual connections that mitigate vanishing gradients. It demonstrated strong generalization but demanded significant computational power while running.
- **VGG16** delivers a competitive accuracy compared to the other architectures. Its deep yet simple architecture required more parameters, increasing computational cost while maintaining strong feature extraction.

FUTURE WORK

- Optimize each model individually on the black box attack dataset using data augmentation, fine-tuning, and hyperparameter tuning.
- Enhance the CNN model by experimenting with deeper architectures, more filters, and adjusted learning rates.
- Fine-tune pre-trained models by unfreezing earlier layers and training with a lower learning rate for better adaptation to the dataset.
- Implement and compare other pre-trained models (e.g., EfficientNet, InceptionV3, DenseNet) and evaluate their performance on the dataset.

REFERENCES

[1] M. Biswal, S. Misra, and A. S. Tayeen. "Black Box Attack on Machine Learning Assisted Wide Area Monitoring and Protection Systems," *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT).*
[2] S. Vahidi, M. Ghafouri, M. Au, M. Kassouf, A. Mohammadi and M. Debbabi, "Security of Wide-Area Monitoring, Protection, and Control (WAMPAC) Systems of the Smart Grid: A Survey on Challenges and Opportunities," in *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1294-1335.
[3] PlotNeuralNet. *Neural network architectures visualization. HarisIqbal88.* GitHub. VGG16 architecture.
[4] Soufiane Hamida et al. "Handwritten computer science words vocabulary recognition using concatenated convolutional neural networks", *Multimedia Tools and Applications 2022.*