DEEP LEARNING PYTORCH FINAL PROJECT
INFO 6147(01) – WINTER 2025 – LAM TRINH DINH

PROJECT REPORT

Topic: Plant Type Classification

1. Introduction

This project aims to develop a deep learning model for classifying different plant types using image data. The primary objective is to build and compare the performance of various convolutional neural network (CNN) architectures, including both custom-designed models and fine-tuned pretrained models.

2. Objectives

- To classify plant species using deep learning models.
- To compare the performance of traditional CNN models with fine-tuned CNN models incorporating various attention mechanisms.
- To evaluate the impact of attention mechanisms on model accuracy and efficiency.

3. Dataset

- The dataset used consists of images of 30 plant species, each category containing 1000 images.
- Data source: Plants Type Datasets - Kaggle
- The data is split into training, validation, and testing sets.

4. Methodology

4.1 Data Preprocessing

- Normalization, regularization, image transformation, and cropping are applied.
- Images are resized to 224x224 pixels for uniform input size.

4.2 Model Architectures

1. **Custom CNN Model:** A basic CNN architecture with two convolutional layers trained from scratch.

Architecture Description:

- Input: 224x224x3 (RGB image)
- Conv Layer 1: 32 filters of size 3x3, ReLU activation
- MaxPooling Layer: 2x2
- Conv Layer 2: 64 filters of size 3x3, ReLU activation
- MaxPooling Layer: 2x2

- Flatten layer
- Fully Connected Layer: 128 units, ReLU activation
- Output Layer: 30 classes (Softmax activation)

2. **Pretrained ResNet18 Model:** Utilizes the residual learning framework of ResNet18, which allows training of very deep networks by using identity mappings (skip connections) to address the vanishing gradient problem. In this project, the pretrained ResNet18 model was fine-tuned to adapt to the plant species classification task by replacing the final fully connected layer to match the number of plant classes (30). The model leverages transfer learning to utilize previously learned features from large-scale image datasets, improving convergence and accuracy for plant classification.
3. **SE-ResNet18 Model:** Incorporates Squeeze-and-Excitation (SE) blocks for channel-wise attention.

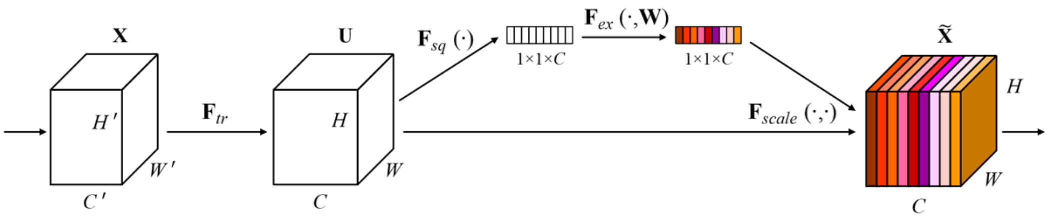
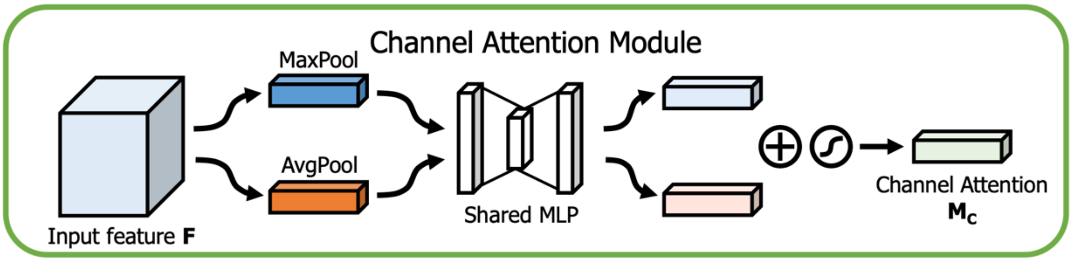


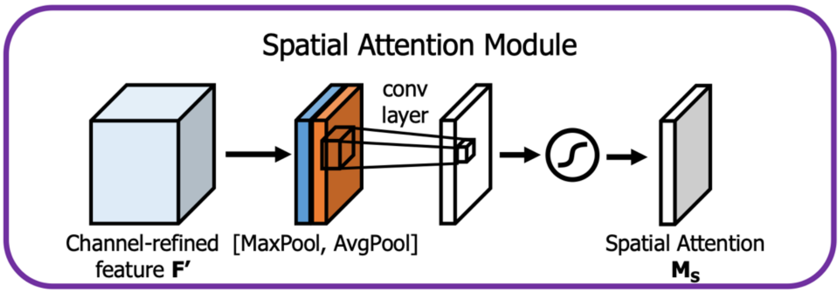
Fig. 1. A Squeeze-and-Excitation block.

- **Mechanism**: The Squeeze-and-Excitation (SE) block aims to enhance the representational power of a network by focusing on the most informative features. It works as follows:
  - **Squeeze:** Aggregates spatial information across each channel using global average pooling, resulting in a channel descriptor.
  - **Excitation:** Uses small neural network with fully connected layers to learn channel-wise dependencies and generate weights that indicate the importance of each channel.
  - **Recalibration:** The input feature map is scaled by the learned weights, emphasizing the most informative channels while suppressing less useful ones.

  This mechanism helps the network focus on the most critical parts of the image, improving accuracy in the task.

4. **CBAM-ResNet18 Model:** Uses Convolutional Block Attention Module (CBAM) for combined **channel** and **spatial attention.**

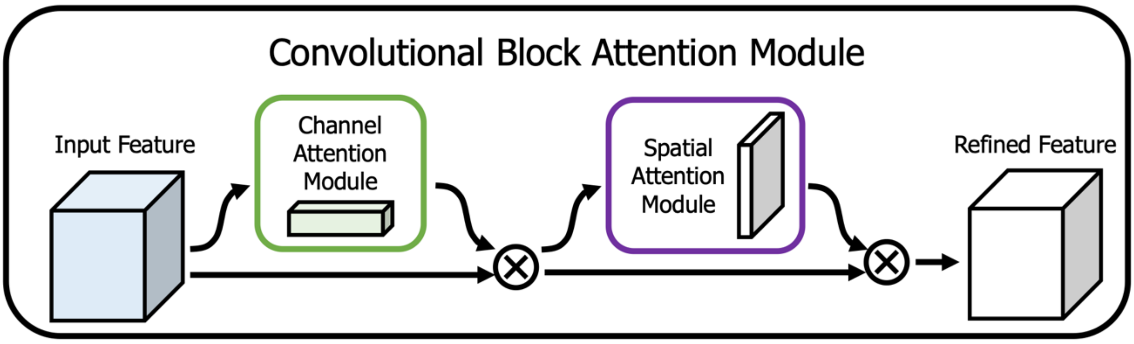




In the CBAM-ResNet18 model, we integrate the Convolutional Block Attention Module (CBAM) into the ResNet18 architecture to enhance its feature representation capabilities. CBAM sequentially applies two attention mechanisms—channel attention and spatial attention—to refine feature maps effectively.

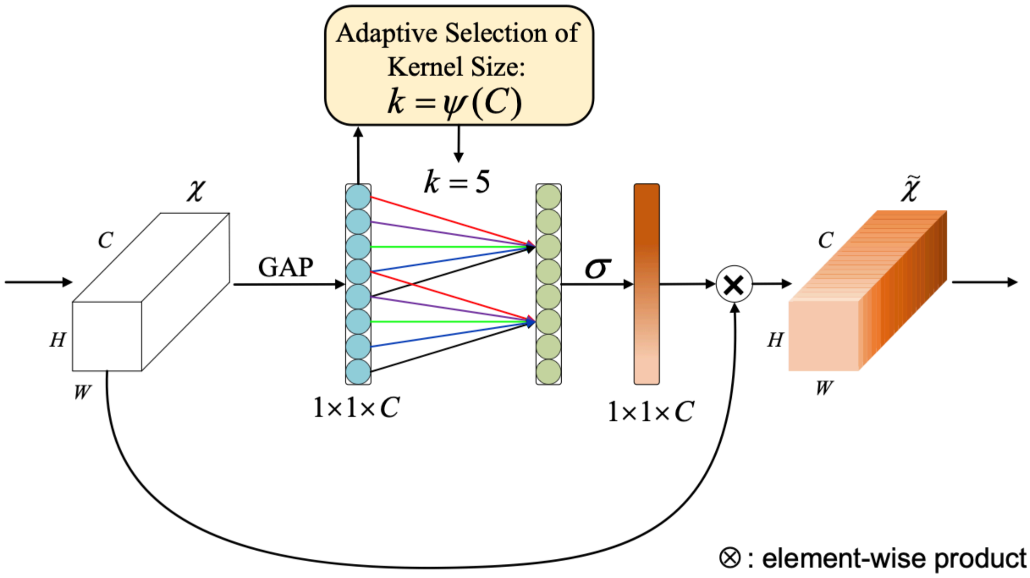
Channel Attention: This mechanism focuses on 'what' is meaningful by emphasizing important feature channels. It utilizes both average and max pooling operations across the spatial dimensions to gather channel-wise information, which is then processed through fully connected layers to generate attention weights. These weights are applied to the feature maps to highlight significant channels.

Spatial Attention: Following channel refinement, spatial attention concentrates on 'where' the important information is located within the feature maps. It pools the feature maps along the channel dimension using average and max pooling, concatenates the results, and applies a convolutional layer to produce a spatial attention map. This map helps in focusing on crucial spatial regions.



By embedding CBAM into ResNet18, the model gains the ability to adaptively prioritize informative features both channel-wise and spatially, leading to improved performance in tasks such as plant species classification.

5. **ECA-ResNet18 Model:** Integrates Efficient Channel Attention (ECA) to improve channel-wise feature refinement.

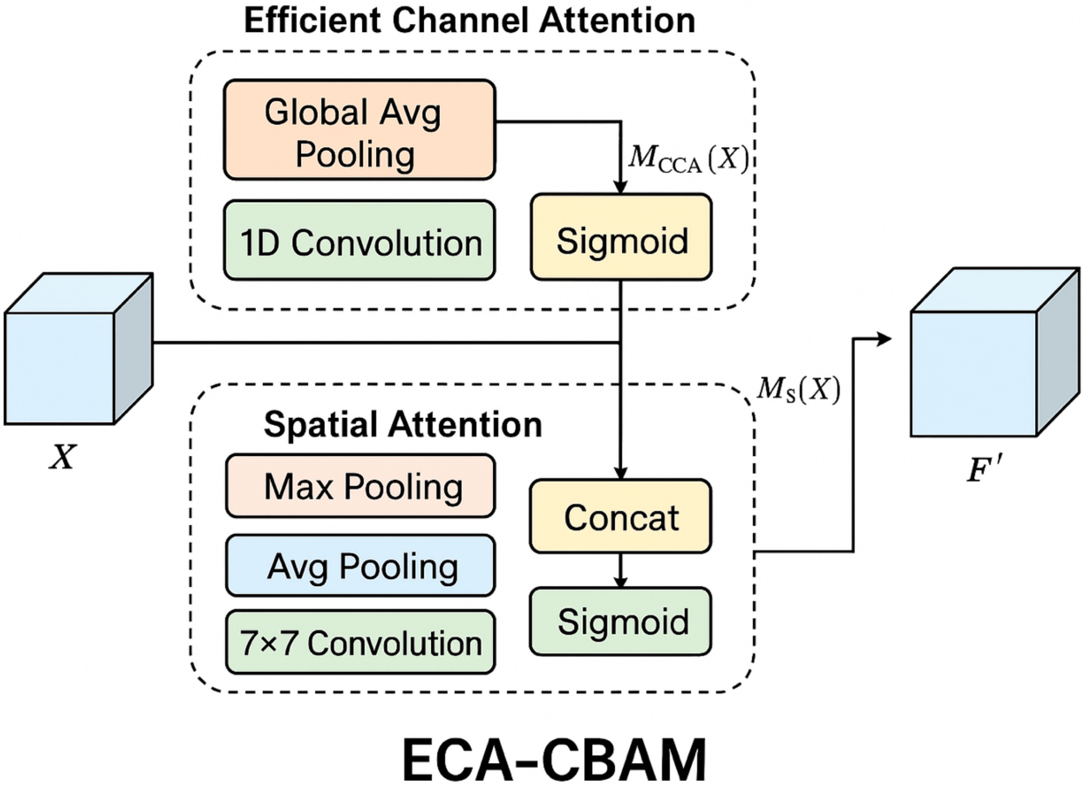


The Efficient Channel Attention (ECA) module enhances the representational power of CNNs by allowing each channel to interact with its neighbors without reducing dimensionality. Unlike traditional channel attention mechanisms that use fully connected layers, ECA employs a 1D convolution with an adaptive kernel size to capture cross-channel interactions efficiently.

In our project, we integrate ECA into the ResNet18 model to improve classification accuracy while maintaining computational efficiency. This approach proved effective, achieving a test accuracy of 97.06%, highlighting the balance between model performance and computational cost.

6. **ECA+Spatial Attention Hybrid ResNet18 Model:** Combines Efficient Channel Attention (ECA) with spatial attention to enhance both channel-wise and spatial

feature refinement.



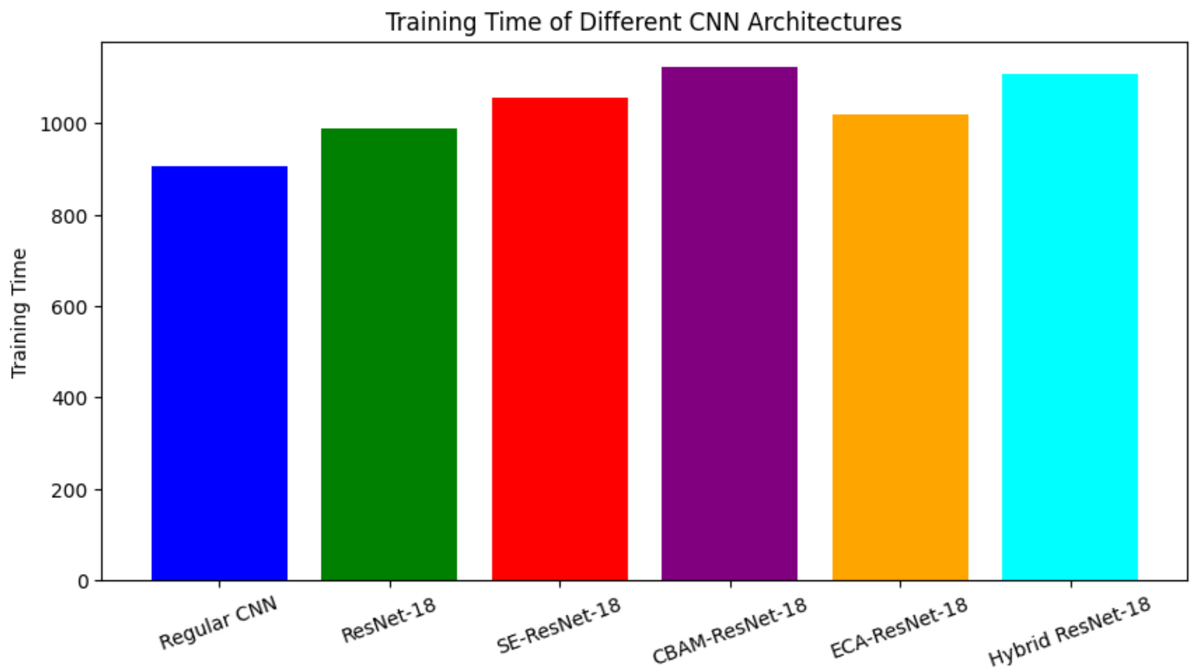
The ECA+Spatial Attention Hybrid model aims to leverage the strengths of both channel and spatial attention mechanisms. ECA efficiently captures channel dependencies through a lightweight 1D convolution with an adaptive kernel size, while spatial attention refines the spatial relevance of feature maps by focusing on the most informative regions.

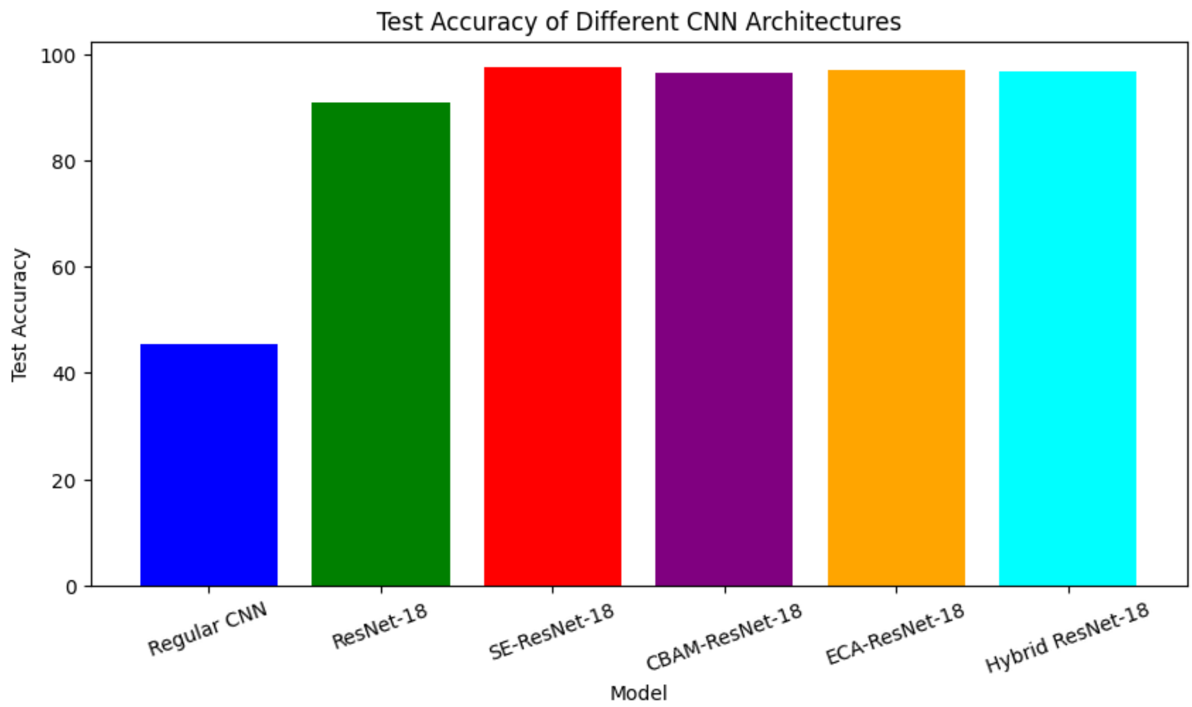
In our project, this hybrid model showed competitive performance with a test accuracy of 96.80%, indicating that combining ECA and spatial attention contributes to robust feature extraction and accurate classification, although with a slightly increased computational cost compared to ECA-ResNet18 alone.

5. Results and Analysis

5.1 Model Performance Comparison

Model	Test Accuracy	Training Time (s)
Regular CNN	45.50%	905
ResNet18	90.79%	989
SE-ResNet18	97.46%	1033
CBAM-ResNet18	96.36%	1190
ECA-ResNet18	97.06%	1104
ECA+Spatial-ResNet18	96.80%	1189





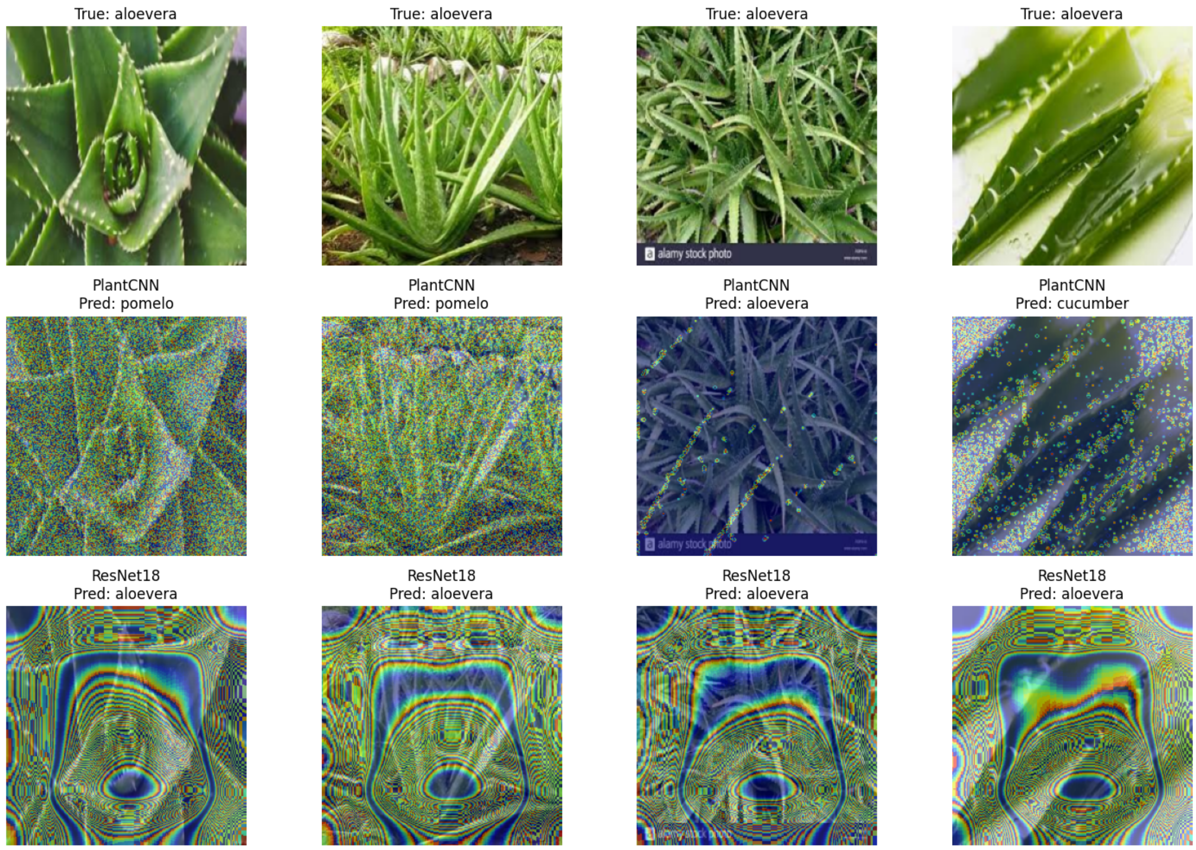
5.2 Analysis

- The Custom CNN model shows relatively low accuracy compared to the fine-tuned ResNet variants, indicating the benefit of leveraging pretrained models.
- SE-ResNet18 outperforms other architectures, achieving the highest accuracy of 97.46%.
- Attention mechanisms (SE, CBAM, ECA) significantly improve the model performance compared to the vanilla ResNet18.

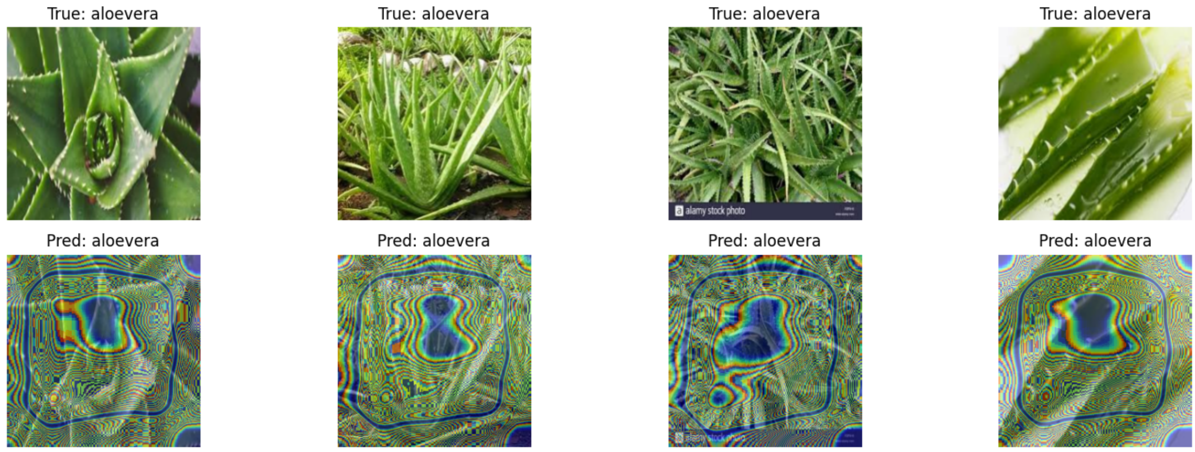
- Although CBAM-ResNet18 and ECA+Spatial-ResNet18 models demonstrate slightly lower accuracy than SE-ResNet18, they still show considerable improvements over the baseline ResNet18.
- Training time increases with the complexity of the model, with CBAM-ResNet18 taking the longest to train.

6. Gradcam Visualization

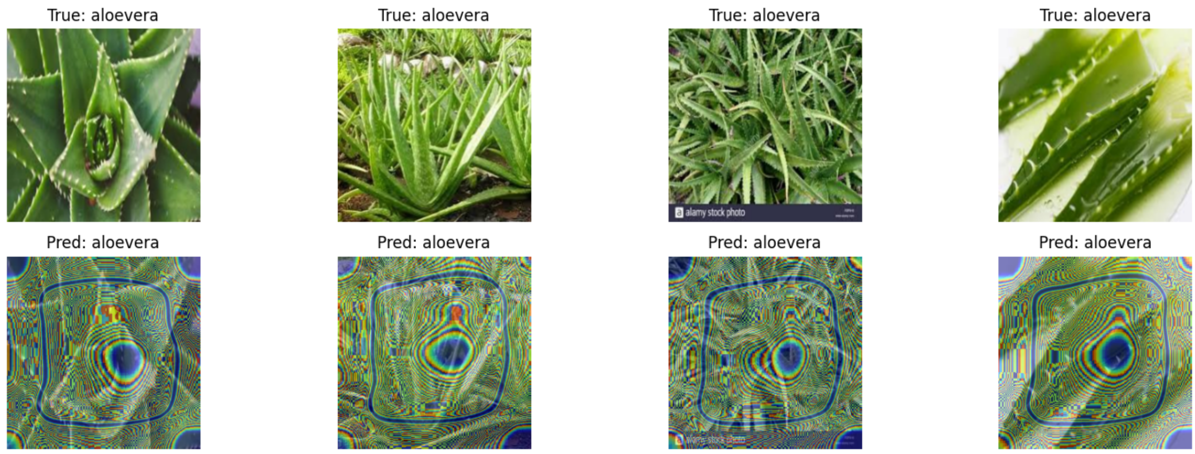
Here, we present the Grad-CAM visualizations of the convolutional layer (layer 4) for each model to compare how they focus on different parts of the plant images. These visualizations provide insights into how attention mechanisms and model architectures impact feature importance, allowing us to interpret model decisions more effectively.



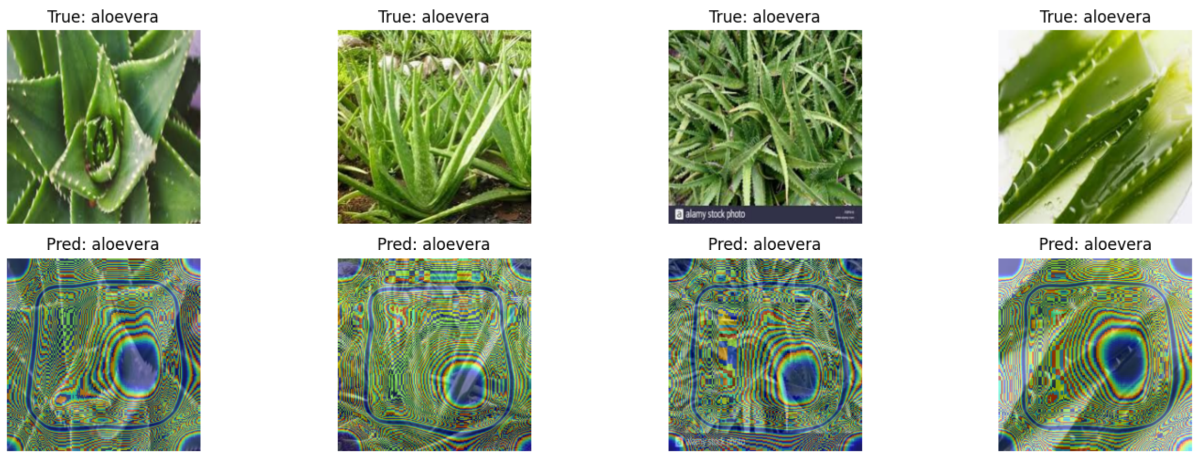
Result of Custom CNN vs Resnet18



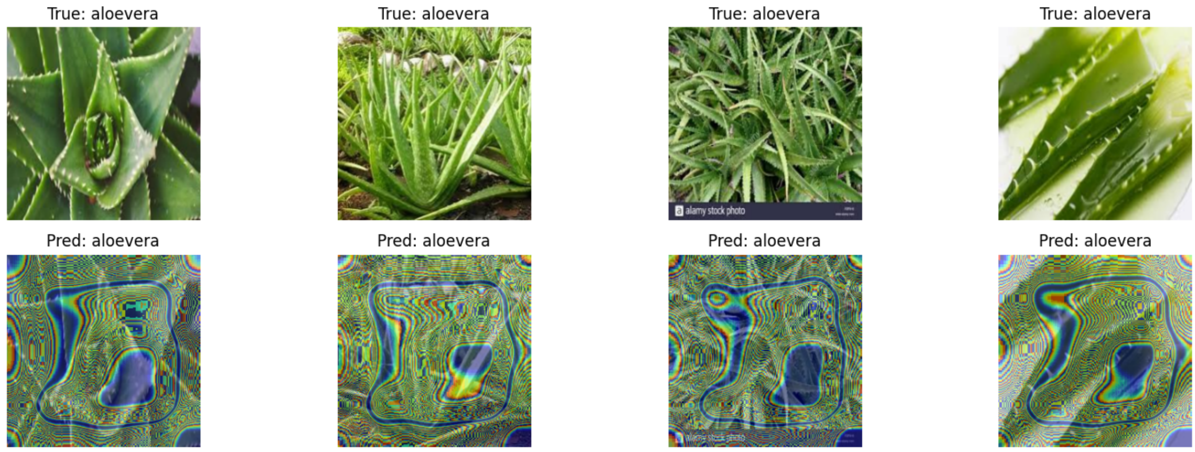
Result of Squeeze and Excitation Resnet18



Result of embedded Convolutional Block Attention Module on Resnet18



Result of embedded Efficient Convolutional Attention on Resnet18



Result of Efficient Convolutional vs Spatial Attention Hybrid Resnet18

7.Reference:

- Data source: https://www.kaggle.com/datasets/yudhaislamisulistya/plants-type-datasets

- Convolutional Block Attention Module: https://arxiv.org/pdf/1807.06521

- Efficient Convolutional Attention: https://arxiv.org/pdf/1910.03151

- Squeeze and Excitation Network: https://arxiv.org/pdf/1709.01507