NAZARBAYEV UNIVERSITY

Plant Disease Identification in Leaf Images using Multi-Scale High-Resolution Vision Transformers

Rakhat Yskak
Nazarbayev University

Motivation

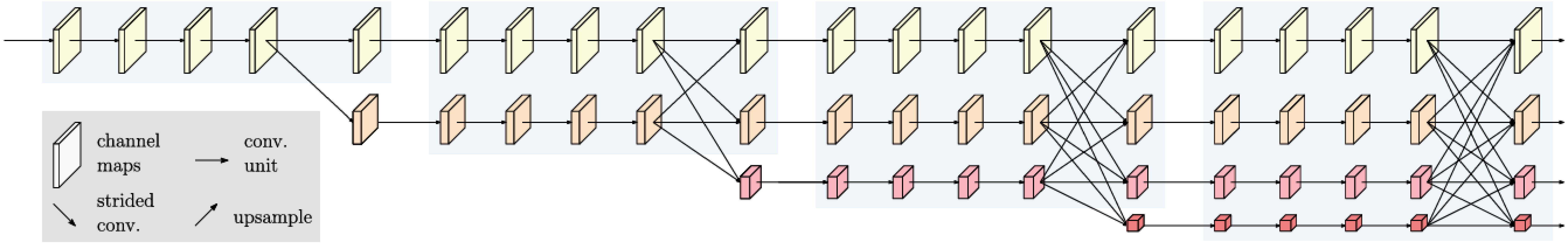
Reducing agricultural losses and guaranteeing food security depend on the precise and timely detection of plant diseases.

Conventional disease detection techniques frequently depend on labour-intensive, human-error-prone manual checks.

Develop an automated framework leveraging HRNet and Vision Transformers to enhance precision and efficiency in disease detection.

NAZARBAYEV
UNIVERSITY

What is HRNet?



01 **Parallel Multi-Resolution convolution**

Different-sized magnifying glasses lined up

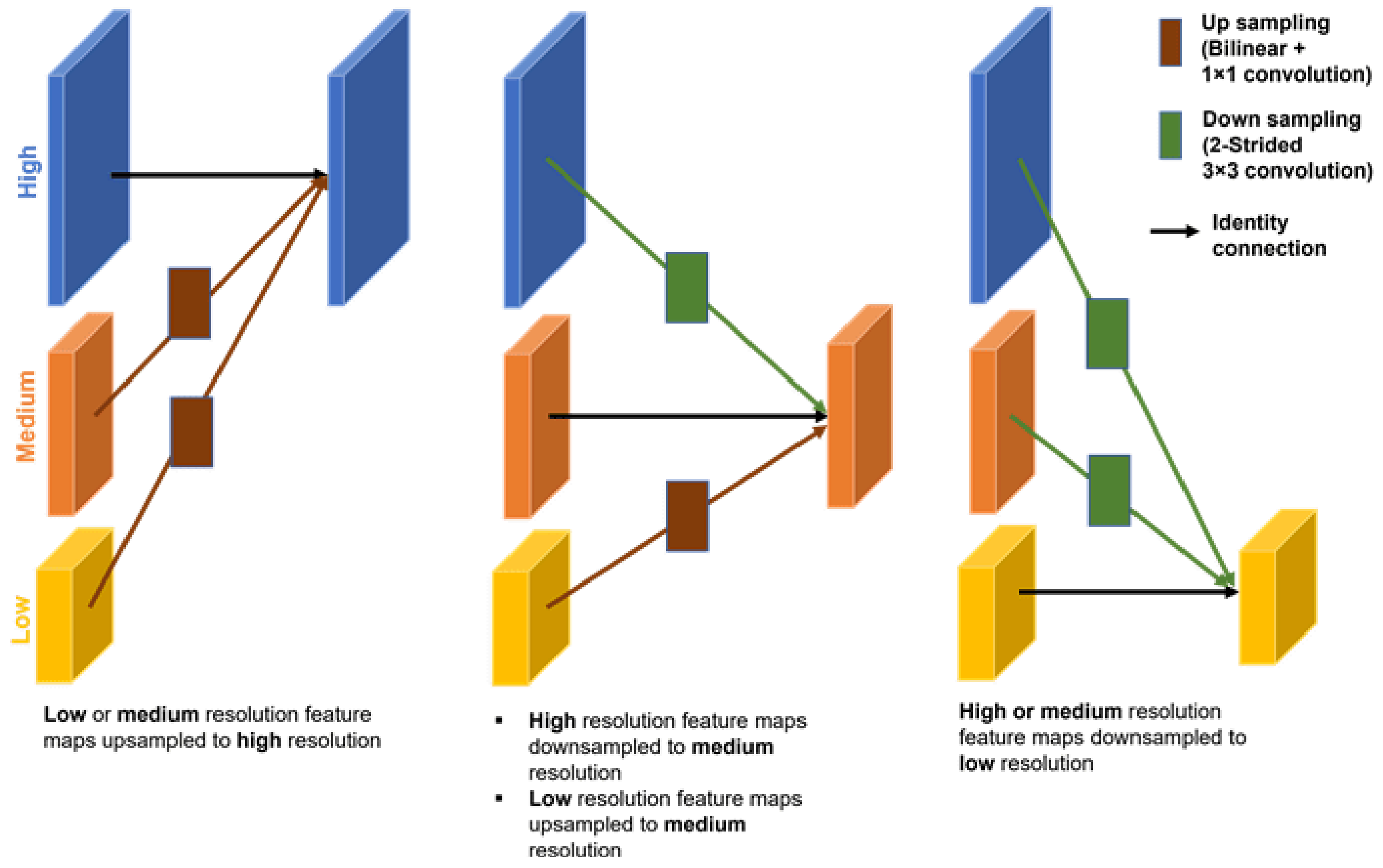
02 **Repeated Multi-Resolution Fusion**

The network frequently merges or fuses the information from all those parallel streams of different resolutions.

This continuous exchange ensures that all parts of the network have the best possible information.

03 **Representation Head**

The part responsible for delivering the final output



NAZARBAYEV UNIVERSITY

ViTaL

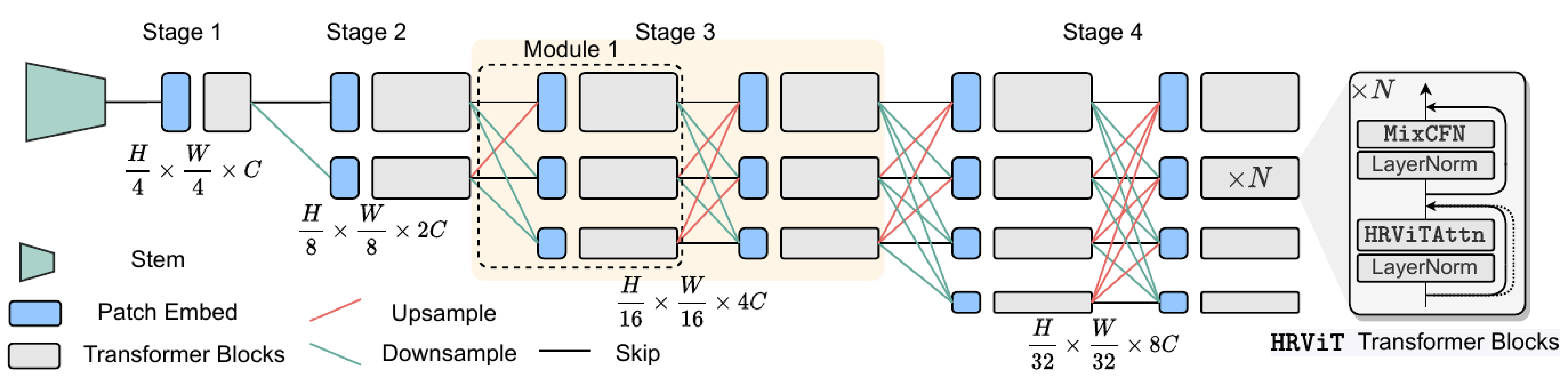
Uses self-attention for capturing global dependencies, crucial for disease identification in leaves.

Positional Embedding with Linear Projection for Feature Reduciton

HRViT

Maintains high-resolution feature streams throughout, allowing precise spatial feature representation.

How to integrate ViT in HRNet

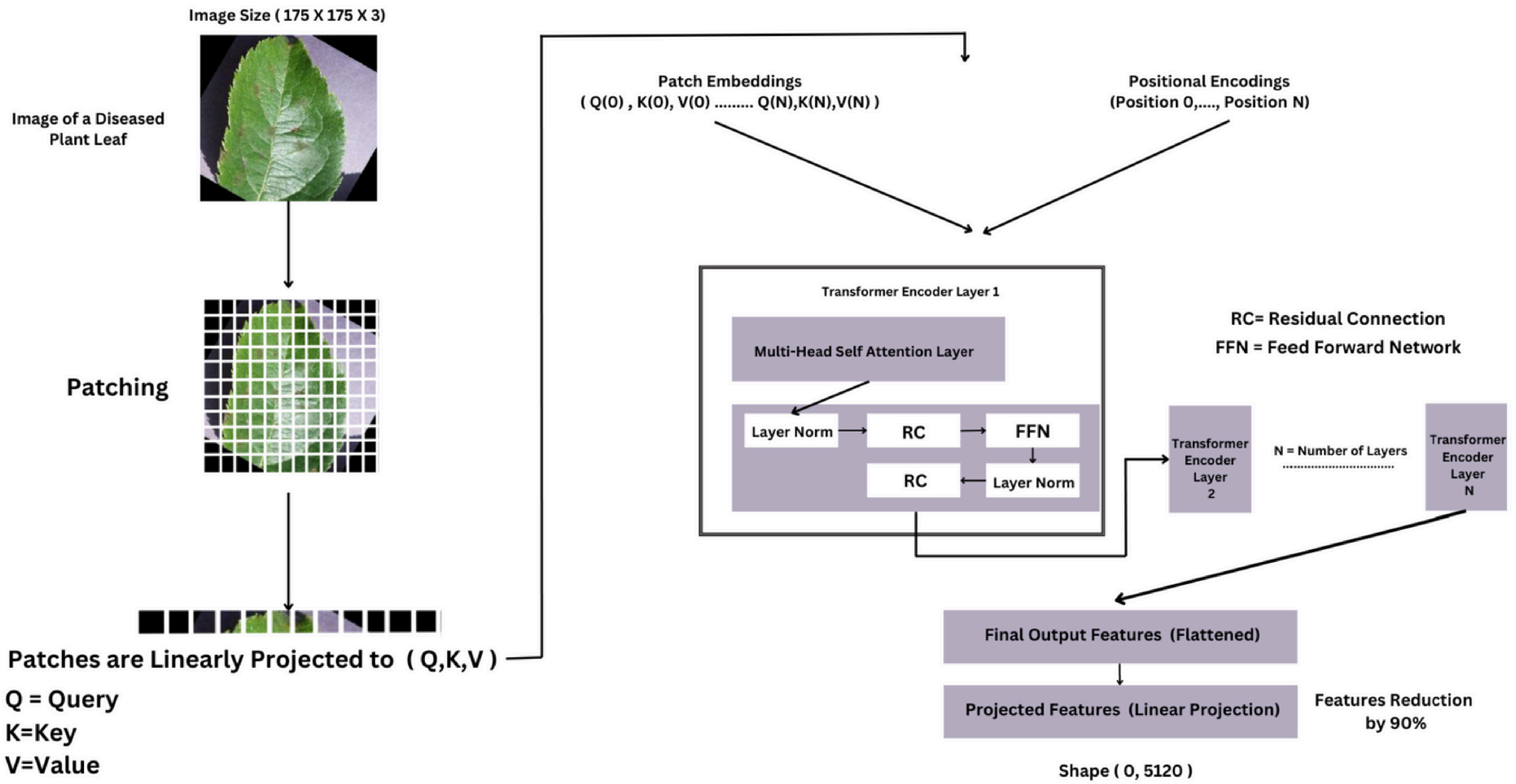


NAZARBAYEV UNIVERSITY

Positional Embedding with Linear Projection for Feature Reduciton

Vision Transformers treat image patches as tokens, similar to words in NLP.

Positional embeddings maintain spatial relationships, essential for detecting diseases that affect specific leaf areas.



Data and Preprocessing

01 **Dataset**

A customized subset of the PlantVillage dataset, containing 11 classes with 275 images per class.

02 **Preprocessing Techniques**

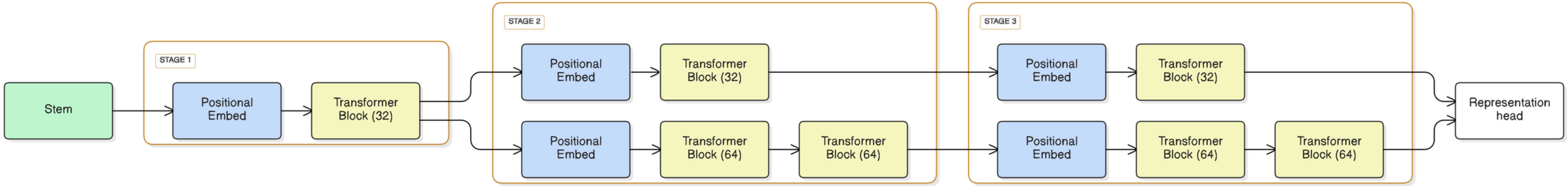
Resizing: Bilinear interpolation for size standardization.

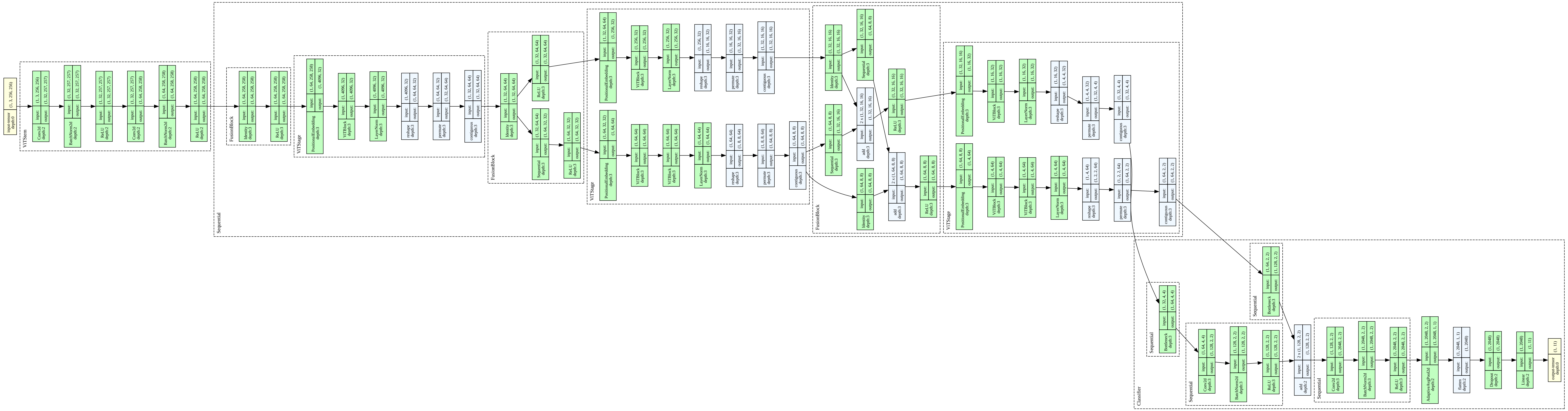
Normalization: Min-max normalization for consistent pixel intensity distribution.

03 **Data Augmentation**

Random horizontal flips, rotations, and color jitter to improve model generalization(report).

Model Architecture





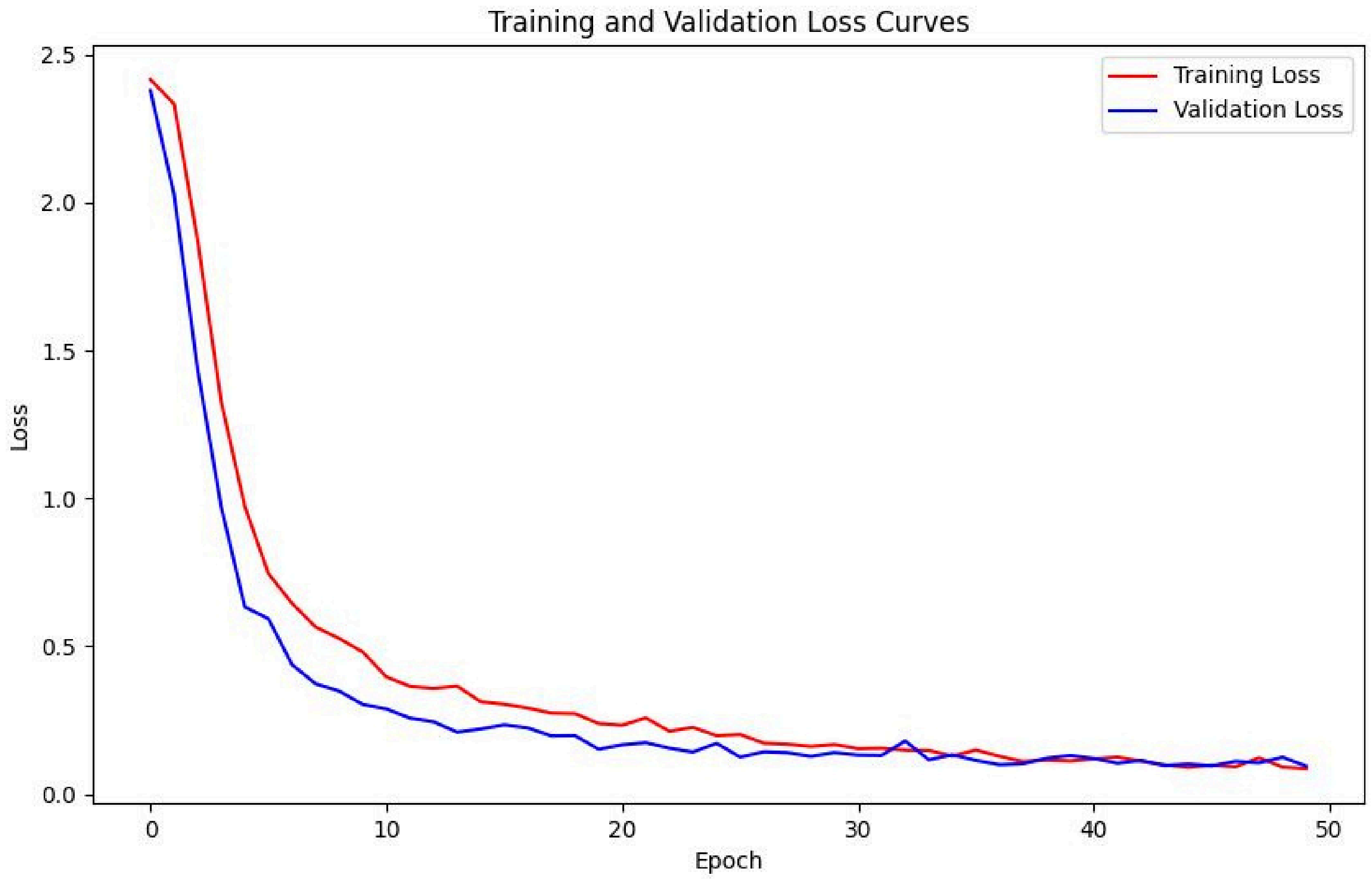
NAZARBAYEV UNIVERSITY

Model Training

- **AdamW** Optimizer
- **Cross-entropy** Loss Function
- **50** Epochs
- Batch Size **32**
- Learning Rate **0.0001**
- Weight Decay **0.00001**

Implemented **early stopping with patience of 10** to prevent overfitting if validation loss stagnates.

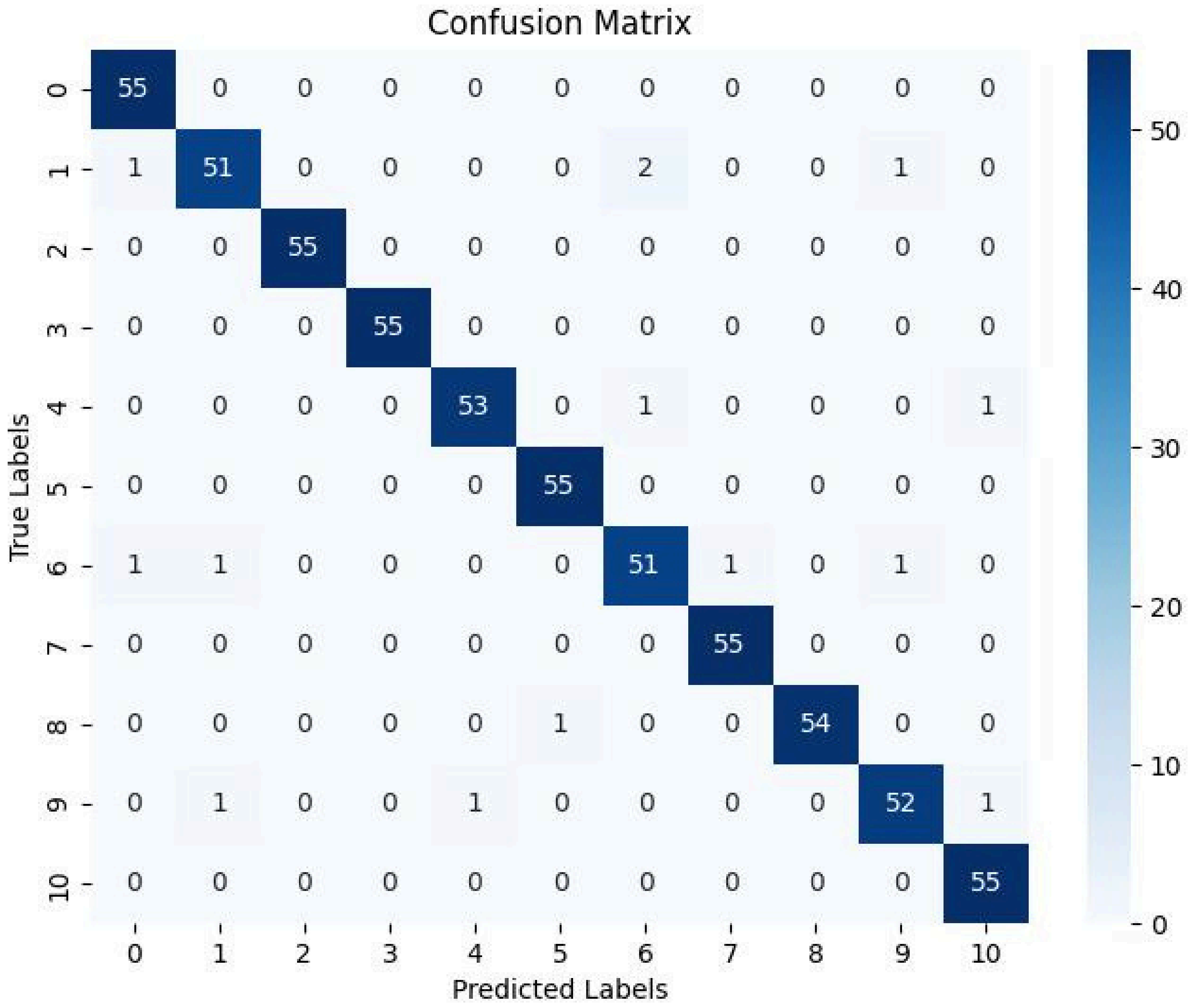
Loss Curves



Results

Test Accuracy: **97 %**
Test Loss: **0.08**

	Precision	Recall	F1-score
Micro-Averaged	0.98	0.98	0.98
Macro-Averaged	0.977	0.977	0.977
Hamming Loss	0.023		



Discussion & Future Work

Misclassifications in visually similar classes suggest a need for a more diverse dataset.

Since the model is too powerful, need for a larger high-resolution plant dataset arise

	Precision	Recall	F1-score	Support
Apple Healthy	0.96	0.93	0.94	55
Cedar Apple Rust	0.96	1.00	0.98	55
Maize Healthy	1.00	1.00	1.00	55
Maize Common Rust	1.00	1.00	1.00	55
Grape Healthy	0.98	1.00	0.99	55
Grape Esca (Black Measles)	0.98	0.96	0.97	55
Peach Healthy	0.98	1.00	0.99	55
Peach Bacterial Spot	0.94	0.93	0.94	55
Raspberry Healthy	1.00	0.98	0.99	55
Tomato Healthy	0.96	1.00	0.98	55
Tomato Yellow Leaf Curl Virus	0.96	0.95	0.95	55
Accuracy			**0.98**	**605**
Macro avg	**0.98**	**0.98**	**0.98**	**605**
Weighted avg	**0.98**	**0.98**	**0.98**	**605**

NAZARBAYEV
UNIVERSITY

Q&A

Thank You