Tomato Disease Classifier

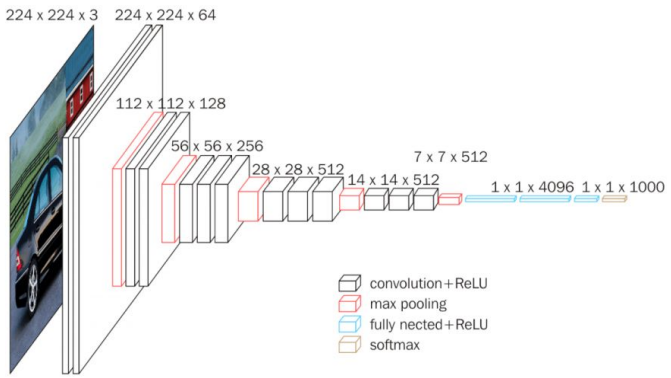
Evie Klaassen
Ka Yam

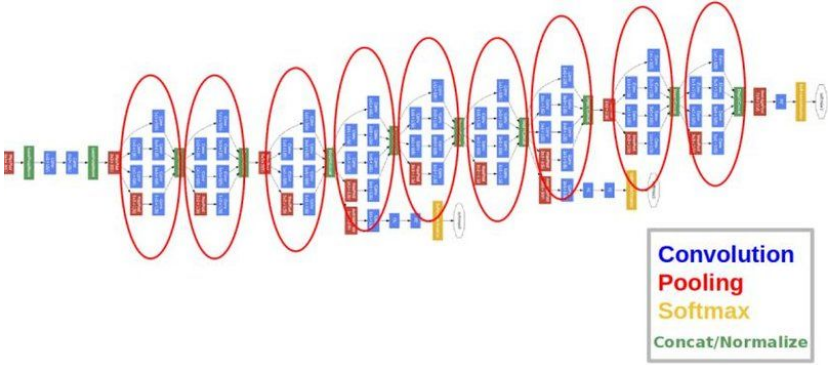
Data Sources:

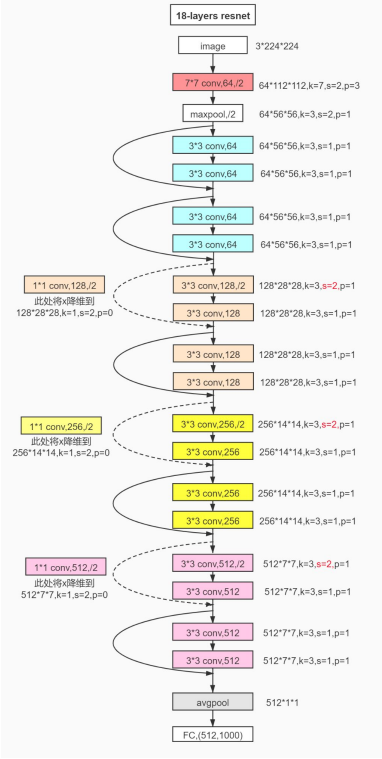
- Mendeley Data – Tomato Leaf Image Dataset
  - 306 images
  - 3 labels (healthy, mosaic virus, yellow leaf curl virus)
- Kaggle – PlantVillage Dataset
  - Color, greyscale, and segmented images (~18k images in each category)
  - 10 labels (i.e. healthy, bacterial spot, early blight, mold, etc.)

Preprocessing

- Resizing to fit model input needs
- Albumentations:
  - Spatial transformations
  - Pixel-level transformations
    - No augmentation, random brightness contrast, blur, Gaussian noise
  - Normalization







Models

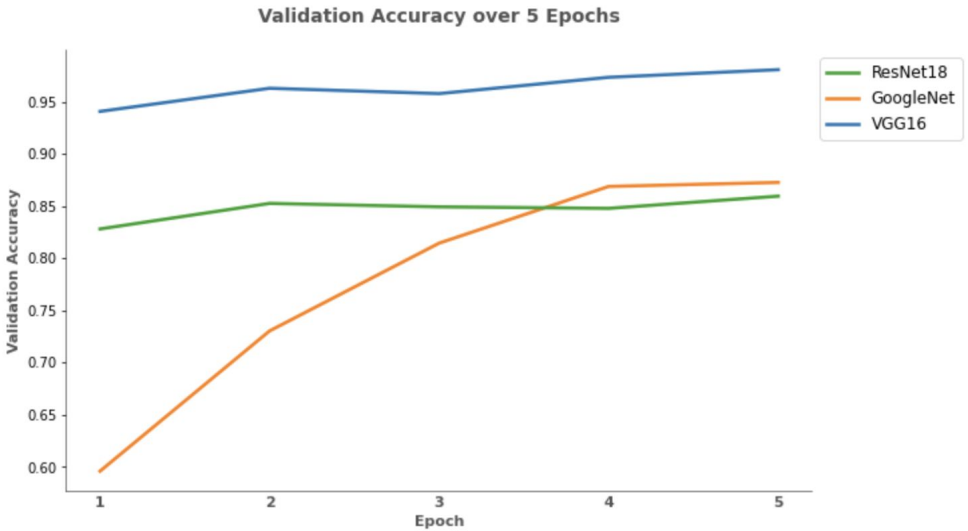
- Pretrained Models:
  - Fine-Tuned VGG16
  - GoogleNet
  - Partially Frozen ResNet18

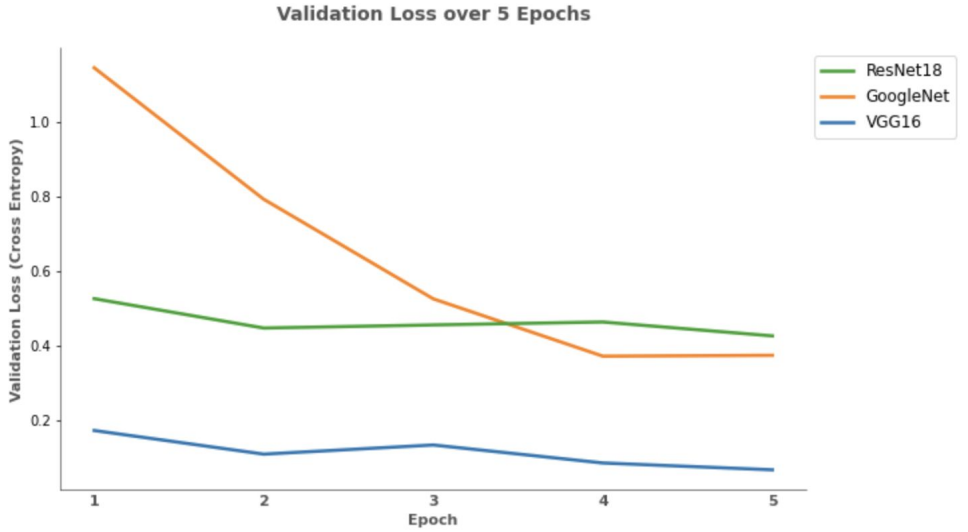


Results - Candidate Models*

Model	Validation Loss	Validation Accuracy
Fine-tuned VGG16	0.086	0.97
GoogleNet	0.0998	0.96
Partially Frozen ResNet18	0.433	0.86

*After training for 5 epochs


Validation Accuracy over 5 Epochs
ResNet18
GoogleNet
VGG16
Validation Accuracy
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
1
2
3
4
5
Epoch


Validation Loss over 5 Epochs
ResNet18
GoogleNet
VGG16
Validation Loss (Cross Entropy)
1.0
0.8
0.6
0.4
0.2
1
2
3
4
5
Epoch



Final Model Performance

- Fine-Tuned VGG16
- Training dataset: 43,826 images (80%)
- Test dataset: 10,957 images (20%)
- Results:
  - **Test Loss: 0.057**
  - **Test Accuracy: 0.98**



Future Directions

- Figure out what classes are frequently confused with others due to high leaf pattern variation
- Integrate the model with built in cell phone cameras to provide real time tomato leaf disease detection and identification

Thank you!