

Fire Sprinkler Classifier with Deep Learning

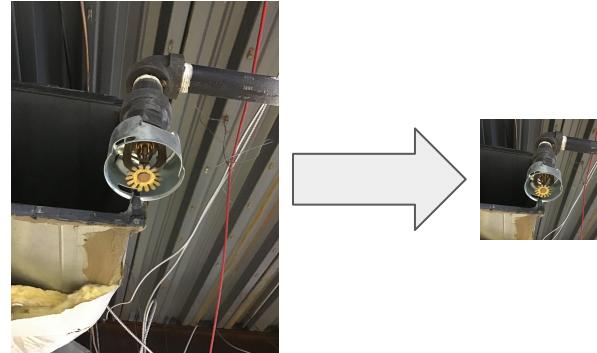
Justin Shin

Motivation

- Construction Site Surveyor for Fire Sprinkler Systems
 - Map out sprinkler system
 - Record elevations (pipes, ducts, obstructions, etc.)
 - Identify sprinkler head model (and take a picture)
- Convenient to classify with one picture
 - Out of reach
 - Dark environments
 - Obscure models
- Interest in Vision Transformers (ViT)
 - Interest in Transformer Architecture
 - Compare performance with Convolutional Neural Networks (CNN)

Method

1. Gather images of sprinkler heads (from 2020-2023 surveys)
2. Label with correct model number
3. Preprocess images
 - a. Resize
 - b. Crop
 - c. Force orientation
4. Experiment with models:
 - a. CNN (from scratch)
 - b. CNN (ResNet)
 - c. Vision Transformer



Dataset

- Gathered 2+ years worth of sprinkler images (~500)
- Hand-labelled each image
- Discard class if < 5 images
- Variations:
 - Blurry
 - Dark
 - Distant
 - Different colors
- **342 images; 12 classes**
 - 7 concealed, 4 recessed, 1 upright, 0 sidewalls

Dataset: 12 Classes



Central A



Central GB4



Central H



Reliable G



Reliable G4



Reliable RA1425



Tyco TY3231



Victaulic V2704



Victaulic V2708



Victaulic V3802

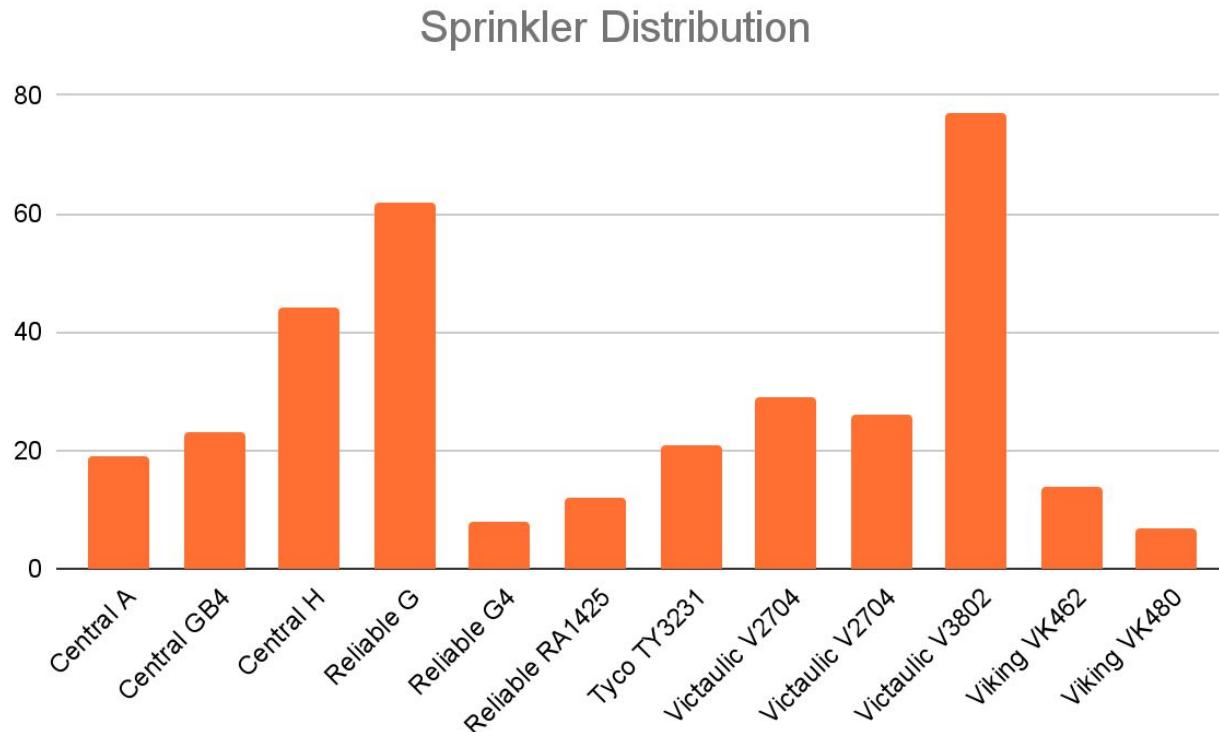


Viking VK462



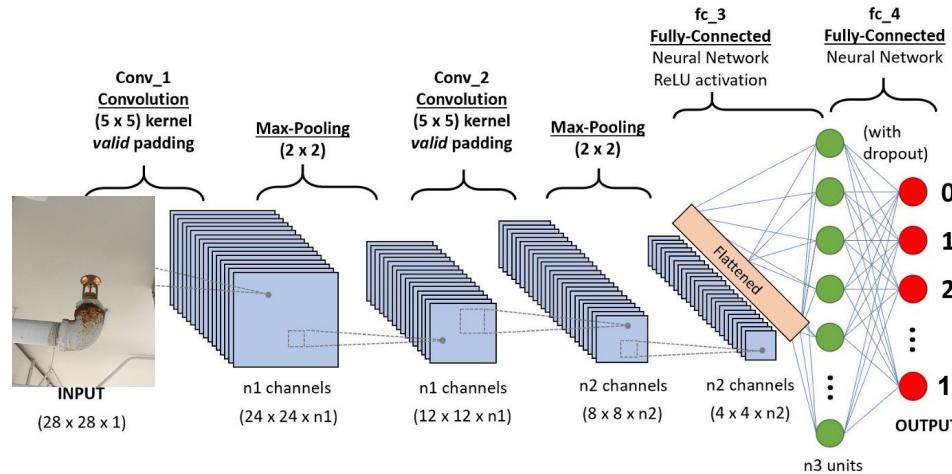
Viking VK480

Dataset: 12 Classes Distribution



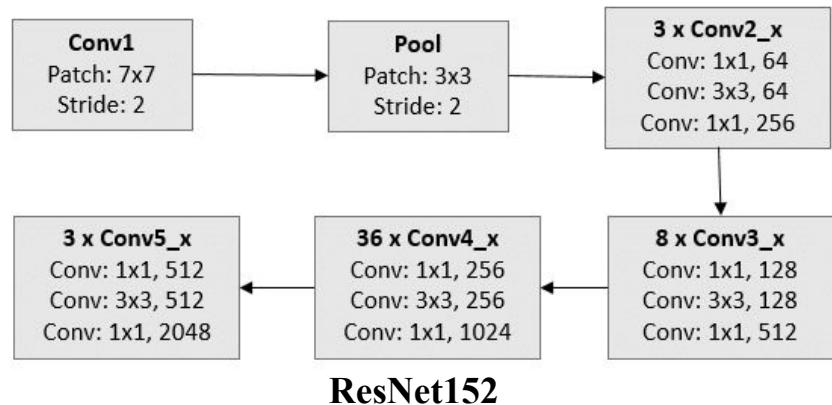
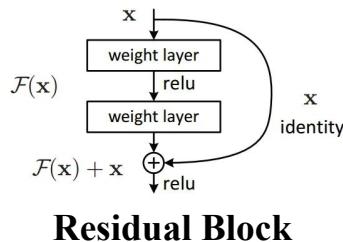
Models: CNN (Scratch)

- Trained from scratch
- Architecture (9 Layers):
 - 3 Convolutional Layers
 - 3 Max Pool Layers
 - 3 Fully-Connected Layers



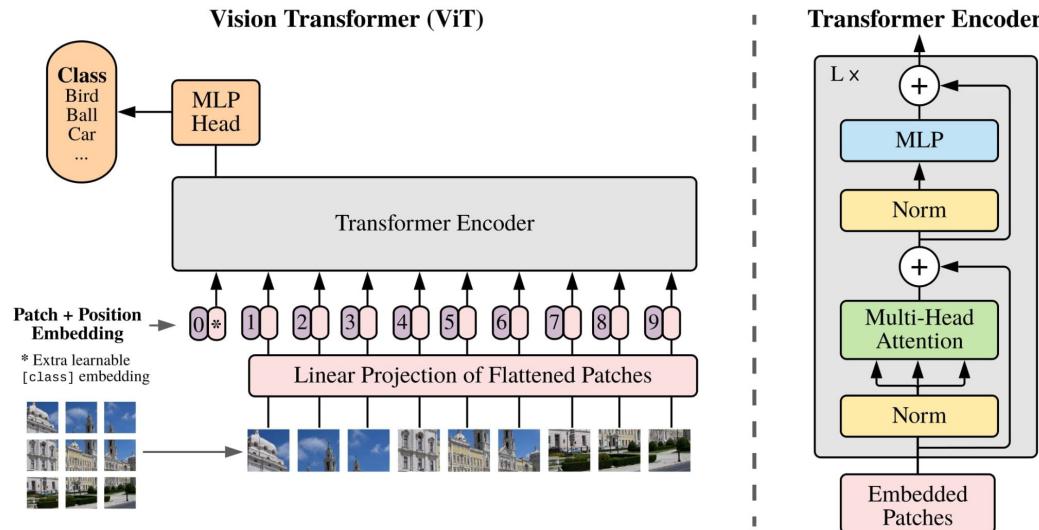
Models: CNN (ResNet)

- He et al (Microsoft), "Deep Residual Learning for Image Recognition", 2017
 - Residual Blocks that passes output to layer deeper in network
 - Allows for deeper network architectures (increase number of layers)
 - Reduce training error with deeper networks (vanishing gradient)
- Pretrained ResNet152 on ImageNet-1k (1M images, 1k classes)
- Architecture (152 Layers):
 - 151 Convolution Layers
 - 1 Fully-Connected Layer



Models: Vision Transformer (ViT)

- Dosovitskiy et al (Google), "An image is worth 16x16 words", 2021
 - Hacking an NLP transformer to process image pixels instead of word tokens
- Pretrained ViT on ImageNet-21k (14M images, 21k classes)



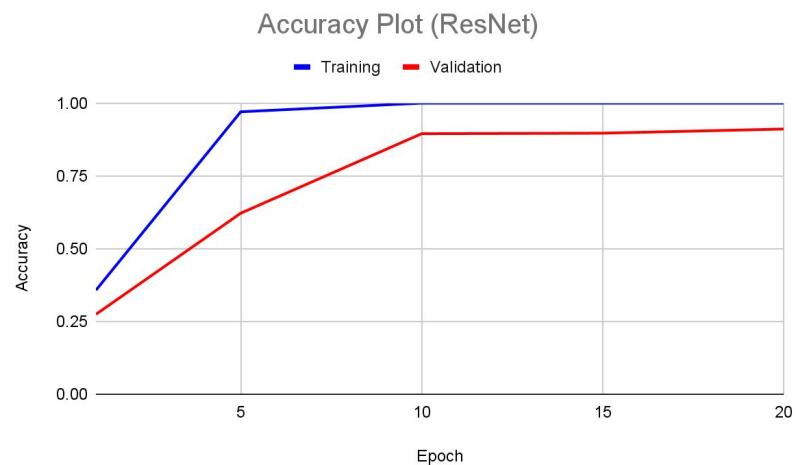
Models: Hyperparameters

Model	Learning Rate	Optimizer	Epochs
CNN (Scratch)	1e-3	Adam	20
CNN (ResNet)	1e-2	SGD	20
Vision Transformer	2e-4	AdamW	10

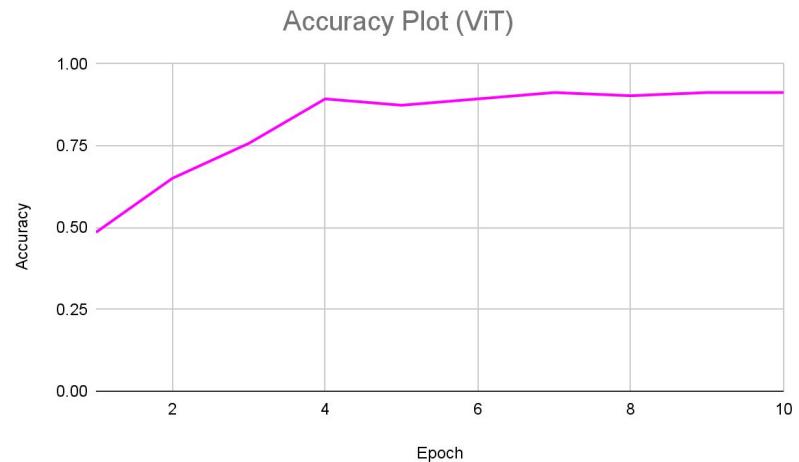
Results: CNN (Scratch)



Results: CNN (ResNet)

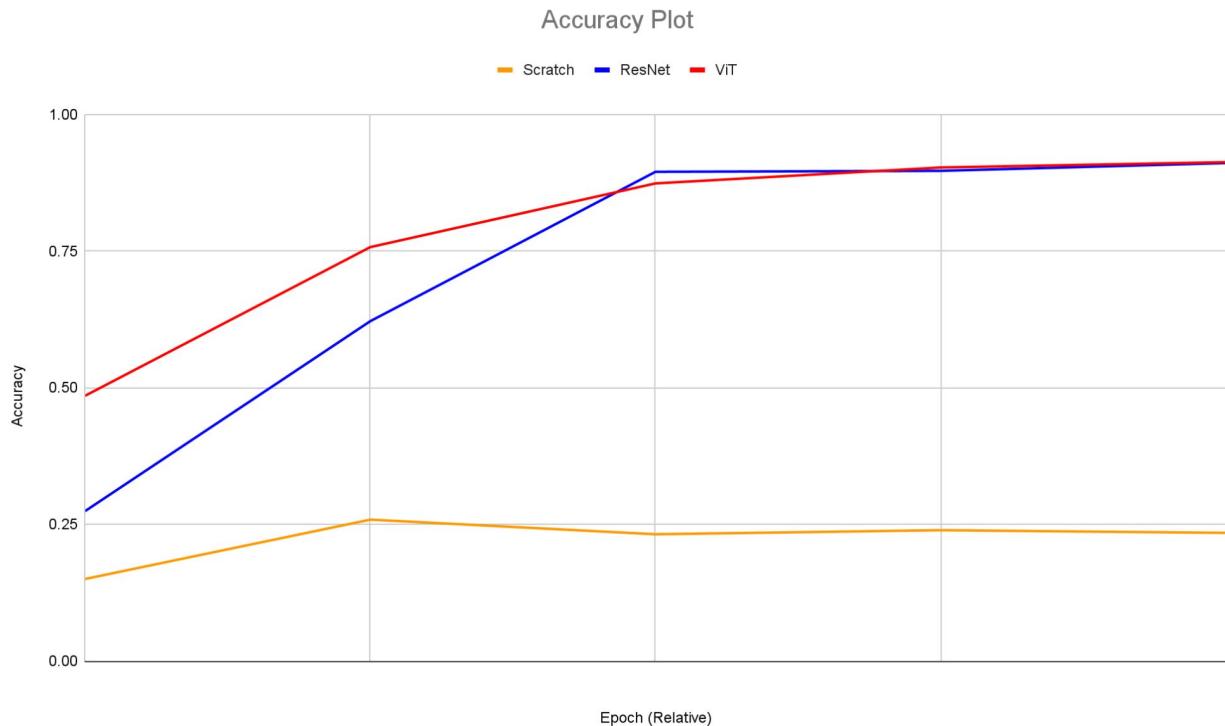


Results: Vision Transformer



Results: Comparison

Model	Accuracy
Scratch	23.47%
ResNet	91.09%
ViT	91.26%



Future Work

- Increase dataset size
 - More low light images
 - More blurry images
- Support more sprinkler models
 - More uprights
 - Sidewalls
- Web application



Thank You