VA Plant Image Classification

Introduction

- Virginians enjoy more than 3,000 square miles of waterways across the state

- Aquatic plants play a major role in their environmental health

* Photos used in this presentation were either taken by one of the group members or taken from our dataset; all images used in the dataset were from sources that allow use for educational purposes

Introduction

- Invasive species of aquatic plants threaten the waterway’s health, suffocating native plants, harming fish and aquatic organism populations, changing the water’s chemistry – and making it harder for humans to enjoy swimming, boating, and fishing

- Hydrilla is one of the most widespread invasive aquatic plant species in VA

- Government and community entities have to expend monetary and human resources to identify and eradicate it

Motivation

- Unfortunately, hydrilla can sometimes be hard to distinguish from other types of aquatic plants that are ”healthy” native plant species

- We built and developed 3 transfer learning/CNN models for image classification of 5 different types of aquatic plant species (hydrilla, arrowhead, duckweed, grassy mud plantain, and watercress)

- Image classification can reduce costs and increase efficiencies when identifying bodies of water where intervention to reduce invasive species is needed

Data Collection & Data

- Dataset
    - Image sources: invasive.org, Google, gbif.org, Shutterstock
    - 450 Images - 5 Aquatic Plants
        - Invasive
            - Hydrilla (101)
        - Non-invasive
            - Duckweed (98), Watercress (100), Arrowhead (76), Grassy Mud Plantain (75)
- Data Split
    - Train: 0.8 Validation: 0.1 Test: 0.1

Data Preprocessing

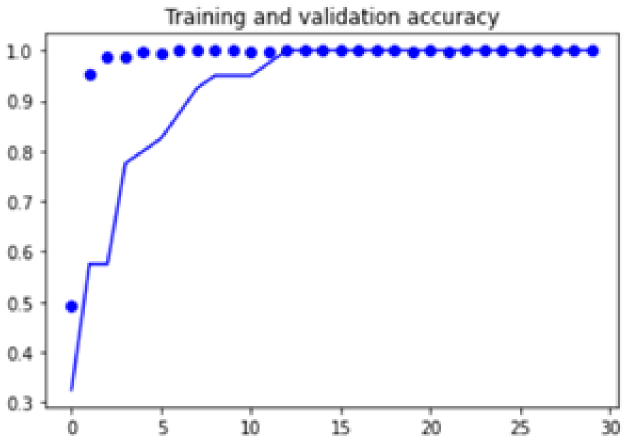
- Uploaded dataset to Google Drive and mounted to Google Colab
- Data Processing
  - Rescaled all images (224, 224)
  - Random Image Augmentation
    - Flip
    - Rotate
    - Contrast
    - Zoom

Initial Experiments

- Custom Neural Network
    - Custom CNN
- Transfer Learning
    - *Xception*
    - *DenseNet 121*
    - VGG19
    - *EfficientNet - B6*

Xception

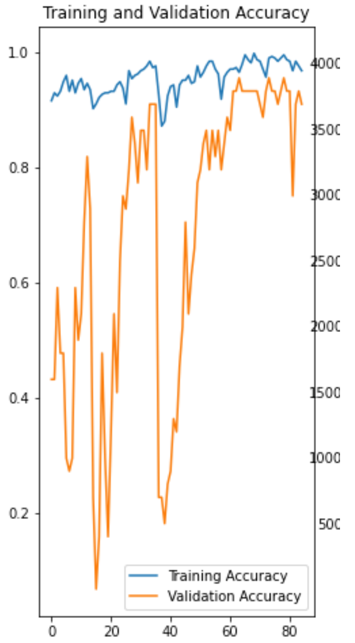
- Fewer parameters and computations than a regular convolution layer yet has better performance
- Added additional BatchNormalization layer
- **Optimizer**: SGD   **LR**: 0.01
- 5 prediction class



	Training	Validation	Testing
Accuracy	99.00%	99.00%	95.56%
Loss	0.0130	0.0119	0.0977

DenseNet 121

- 121 layers within four “dense blocks”
- Experimented with early stopping and drop-out; did not improve performance
- **Optimizer**: Adam **LR**: 0.001
- Epochs: 85



	Training	Validation	Testing
Accuracy	96.37%	90.91%	93.33%
Loss	0.2803	0.9246	0.7544

EfficientNet-B6

- EfficientNet is more accurate and efficient than past CNNs under resource constraints
- Added additional layers
    - GlobalAveragePooling2D
    - BatchNormalization
    - Dropout
- **Optimizer**: SGD **LR:** 0.01
- **Callbacks**: EarlyStopping, ReduceLROnPlateau

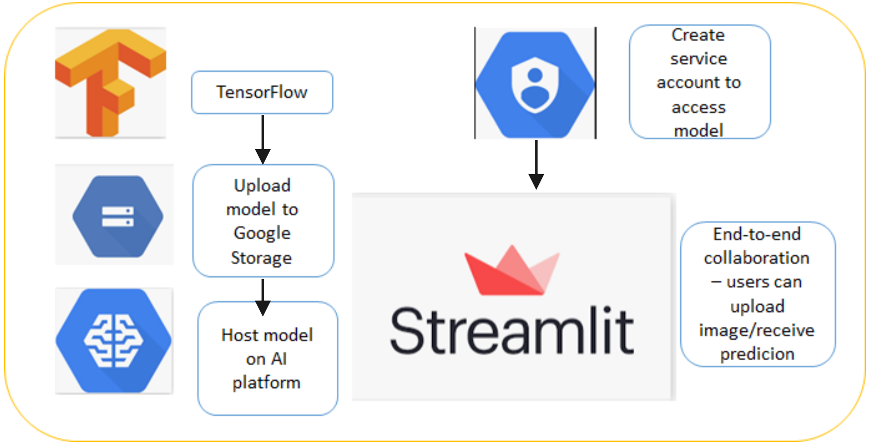
	Training	**Validation**	**Testing**
Accuracy	98.44%	100.00%	97.78%
Loss	0.0562	0.0067	0.1106

Comparison

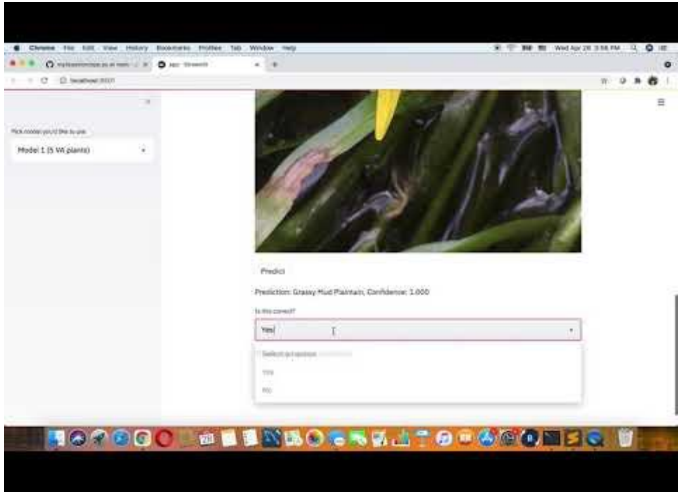
Model	Testing Accuracy
Xception	95.56%
DenseNet 121	93.33%
EfficientNet - B6	97.78%

Model Deployment

- Deployed all models to Google Cloud Platform



Xception Streamlit App Demo



Conclusions

- Our models are able to distinguish 5 classes of plants, including Hydrilla
- Model accuracy indicates sufficient reliability (more limited risk in incorrectly identifying a plant as invasive, resulting in a "healthy" plant being eradicated from a waterway)
- Models provide a platform to more efficiently identify where invasive species are located in waterways through crowdsourcing of photos, rather than limited in-person inspections
- Efficiencies on identification allows for more resources to be allocated directly to intervention
- Future work can expand the training set to improve accuracy and generalizability, include more types of aquatic plant species, and create an app for wider citizen participation