Photo by Patrick Rex

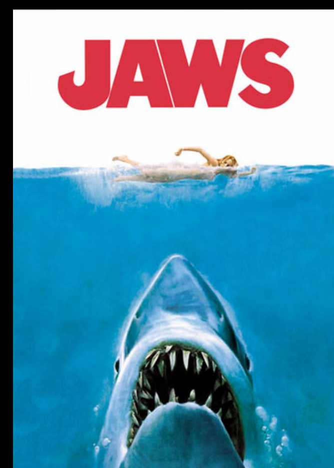
Identifying Humans in Drone Footage from Local Beaches

Echelle Burns – Feb 2020

Introduction: The Problem

Shark Attacks are glorified in the media

- Shark attacks do not occur as frequently as portrayed by pop culture



- No current research on the spatio-temporal overlap between White Sharks (*Carcharodon carcharias)* and beach recreationalists

Introduction: The Solution

Drone Technology in Marine Biology:

- Increasingly popular
- Allows observation of animals without influencing their behavior

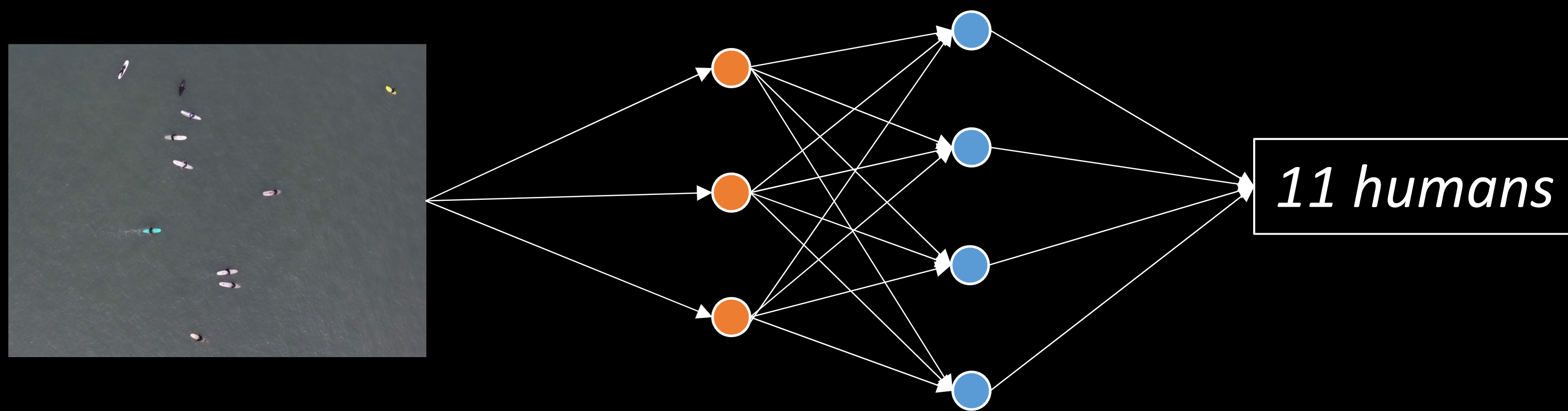
Caveats:

- Amount of footage increases substantially with each drone flight
- Need a fast way to identify whether humans are present in a image

Introduction: Project Objectives

Objectives:

- Use a Deep Neural Network to identify how many people are present in each drone still image



Audience:

- Researchers with similar problems or who are trying to answer similar research questions

Data Sources

- 2,465 drone images (3840 x 2160 px) from CSU Long Beach Shark Lab
- Surveys along beaches in Southern California
- Include footage of:
  - Beach Recreationalists:
    - Walking
    - Wading
    - Swimming
    - Paddle boarding
    - Surfing
    - Kayaking
  - White Sharks
  - Other marine animals

Source: Scenic Sotheby's International Realty

Methods: Image Labeling

- Images resized (960 x 540 px) to make labeling easier
- Manual labeling by placing dots on locations of humans

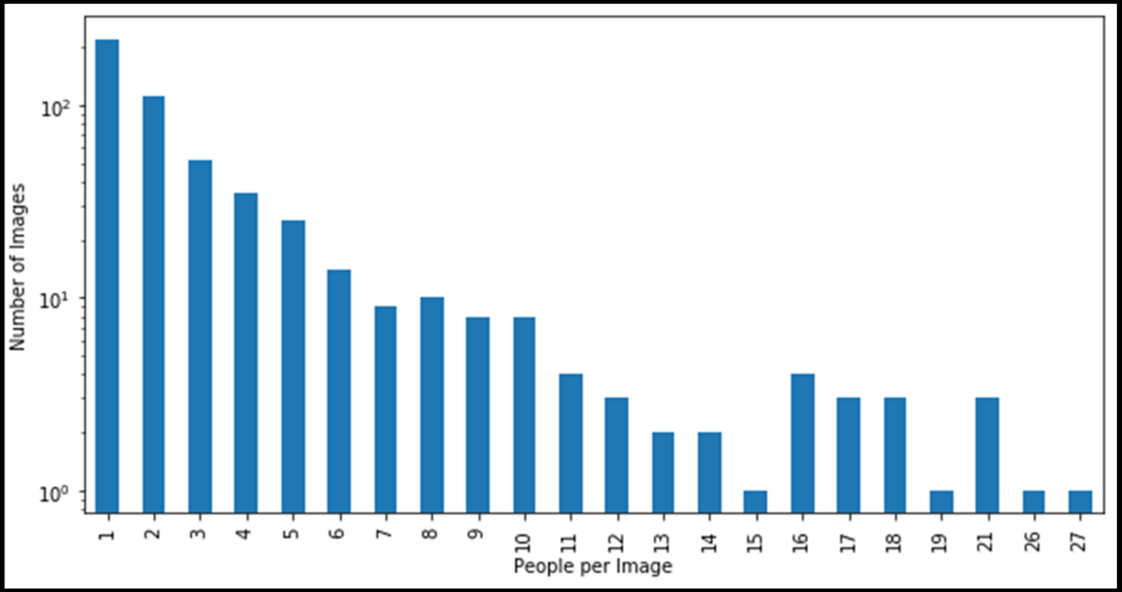
Original Manual Labeling Labeled Image

- Labeled images saved in black and white
- Number of humans generated via *skimage measure.label*

Methods: Image Slicing and Contrast Editing

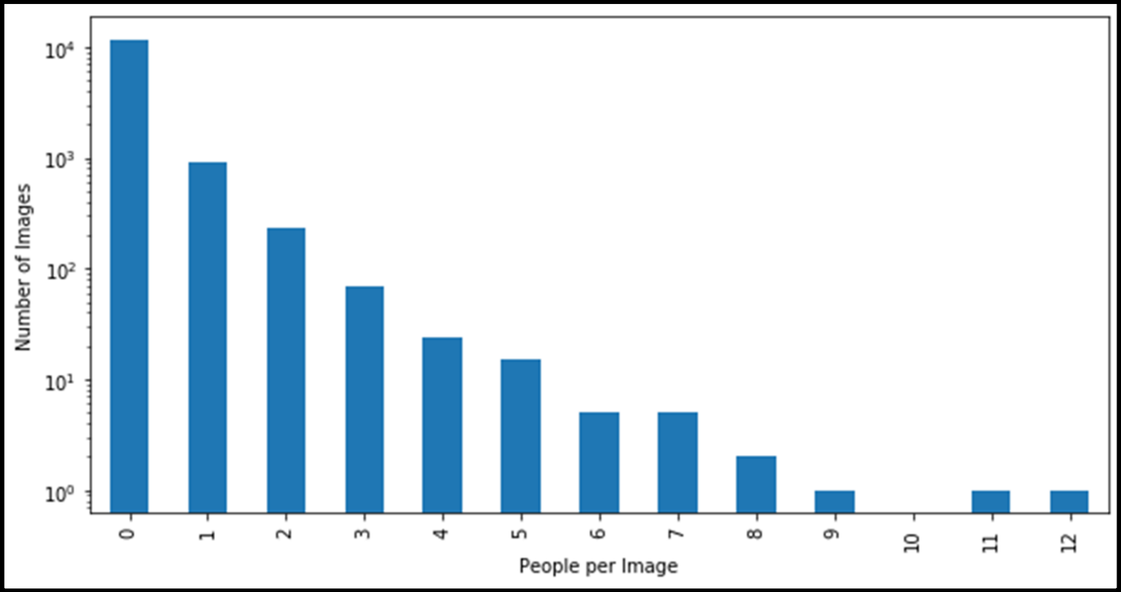
- Resized Images cut into 25 smaller images (192 x 108 px)
  - Increase model efficiency
- Smaller images changed to grayscale/HSV to increase color contrast

Methods/Results: Exploratory Data Analysis



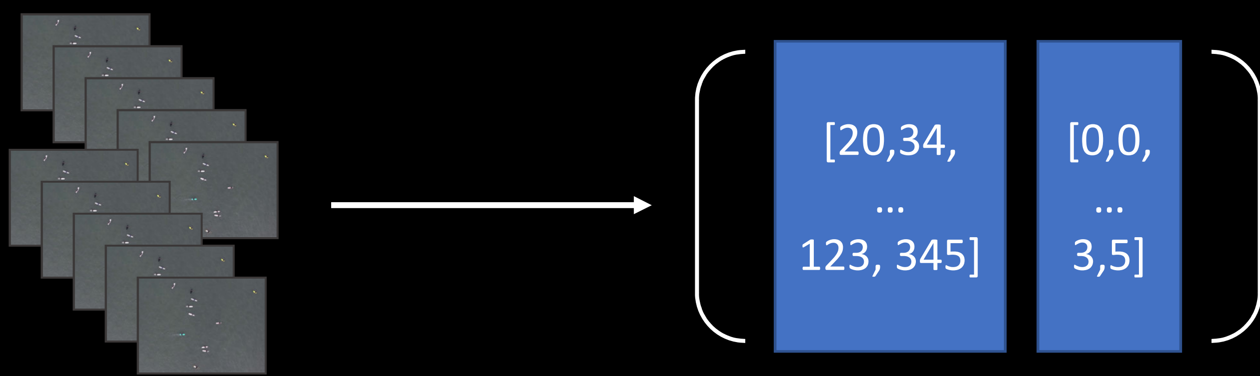
- 521 images with humans present
  - Most with ≤ 7 humans

- 521 large images yielded 13,025 smaller images
  - ~90% had 0 people
  - ~7% had one person



Methods: Convolutional Neural Networks

- Split into training (0.8) and testing data (0.2)
  - Training data then split into training (0.8) and validation (0.2)
- Generator: Grabs groups of images
  - Batch Size = 10 (default)
  - Yields: (pixel values, number of people)



- Model Selection: Brute Force Method
  - Tried many models and chose best performing
    - Best = lowest mean squared error for training and testing

Methods: Convolutional Neural Networks

- Model Selection: Brute Force Method
  - Tried many models and chose best performing
  - General Format:

Convolution Layer
Max Pooling Layer

x4

Flattening Layer

Dense Layer
Dropout Layer

x2

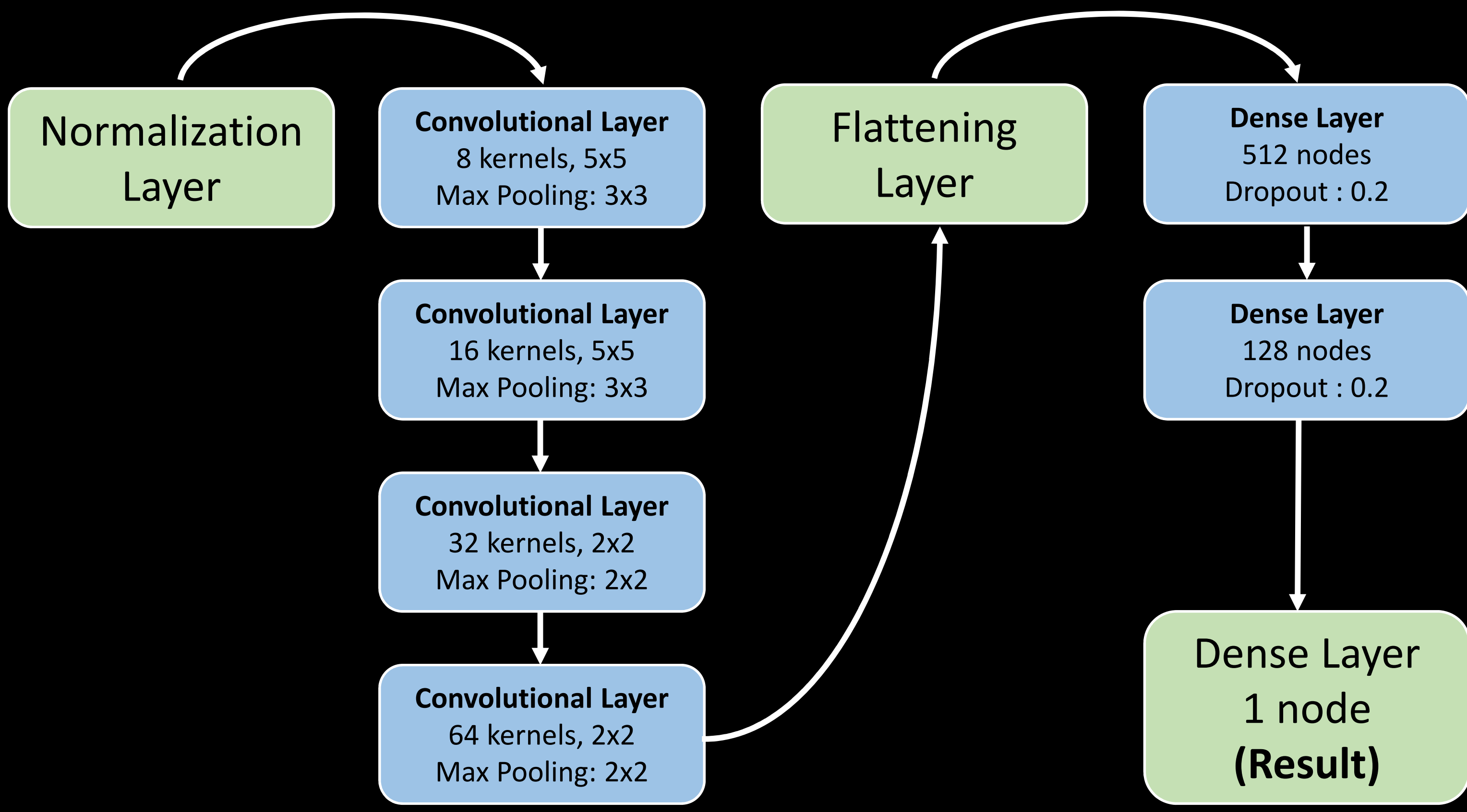
Dense Layer → # of people

Methods: Convolutional Neural Networks

- Model Selection: Brute Force Method
  - Tried many models and chose best performing
  - Altered:
    - Presence of a normalization layer
    - # of kernels in convolutional layers
    - Size of kernels in convolutional layers
    - # nodes in dense layers
    - Using color, grayscale, HSV images
    - Different HSV channels
    - Using full images vs splits
    - Using different activation functions
      - ReLU, sigmoid, linear, tanh, hard sigmoid, PReLU, swish

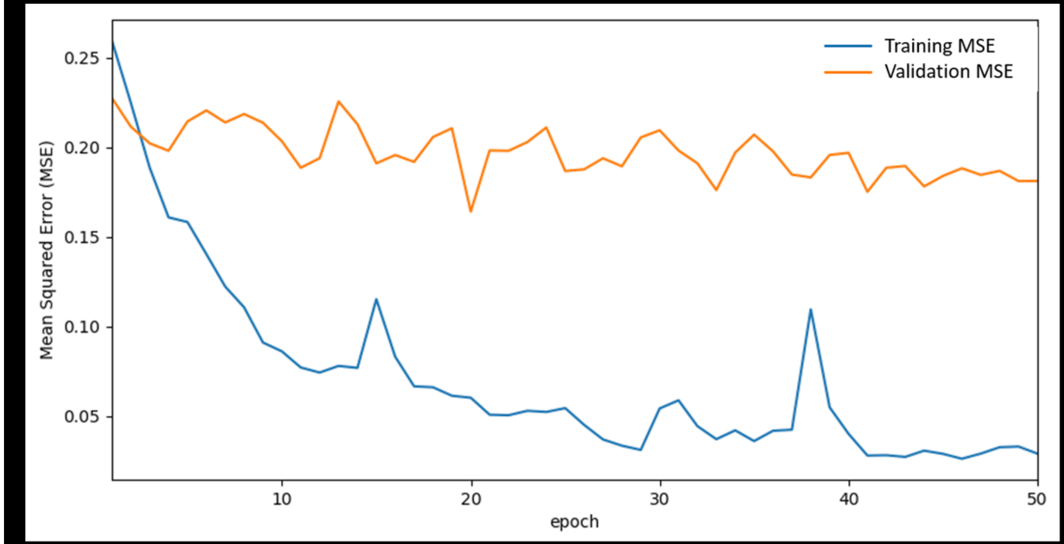
Results: Convolutional Neural Networks

- Model Selection: Brute Force Method
  - Best Model:
    - Split images, HSV channel 2



Results: Convolutional Neural Networks

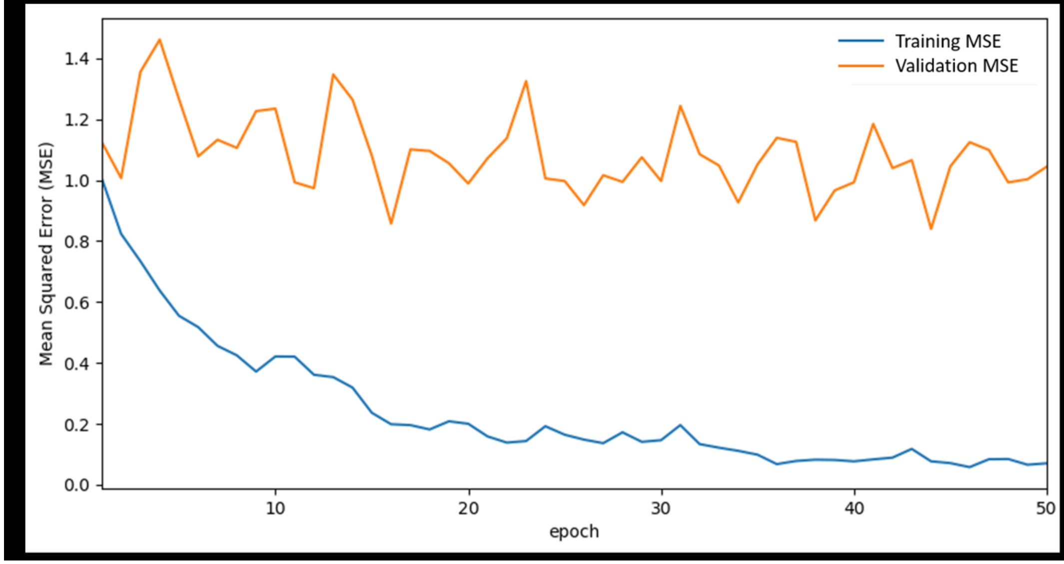
Trained Best Model w/ unbalanced and balanced datasets



Unbalanced Model

Images w/ 0 people >> images w/ 1+ person

Convergence at ~ 40 epochs
Training MSE ~ 0.02 at convergence
Validation MSE consistently ~ 0.2



Balanced Model

Images w/ 0 people == images w/ 1+ person

Convergence at ~ 35 epochs
Training MSE ~ 0.1 at convergence
Validation MSE consistently ~ 1.2

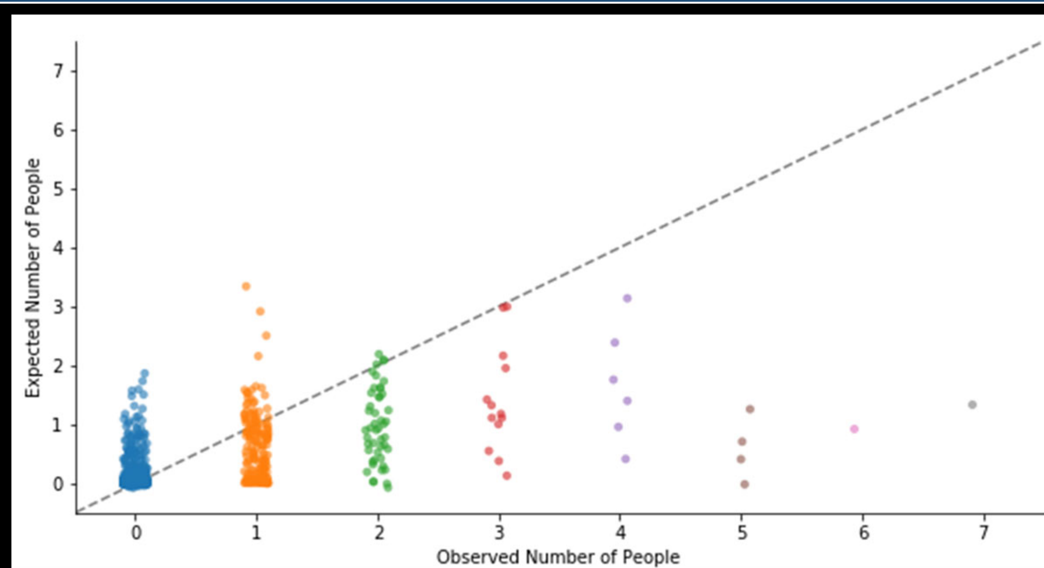
Results: Model Performance

Plotted Residuals for each model using testing data

Unbalanced Model

Images w/ 0 people >> images w/ 1+ person

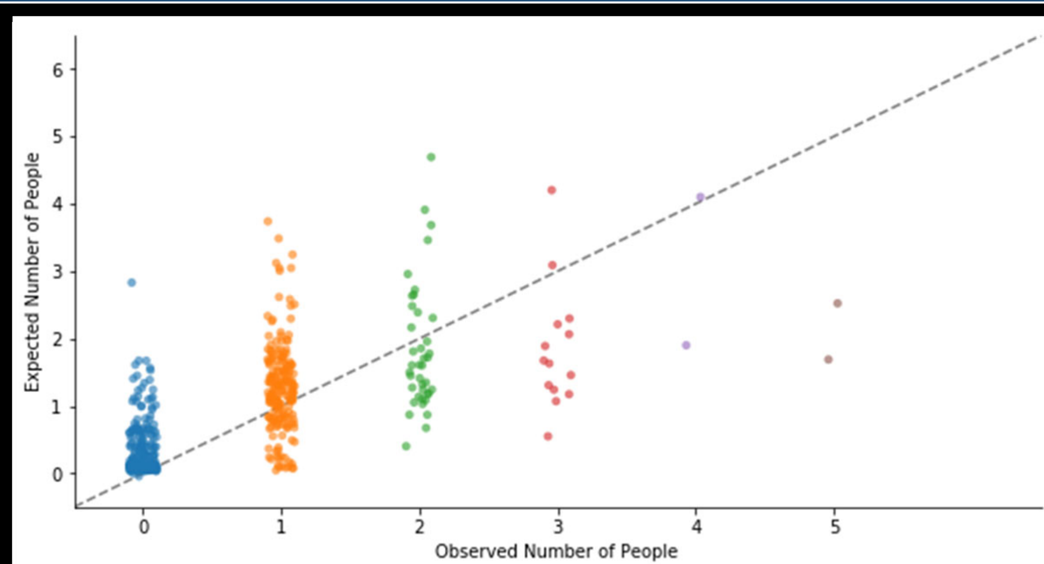
Over-estimates when # people == 0
Under-estimates when # people > 0
Lower variance along the y axis



Balanced Model

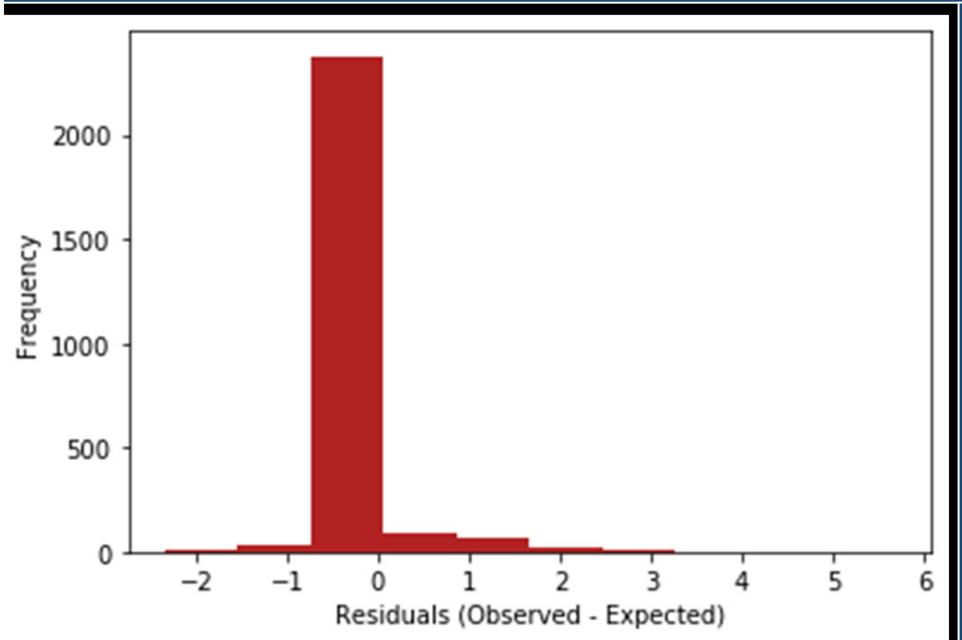
Images w/ 0 people == images w/ 1+ person

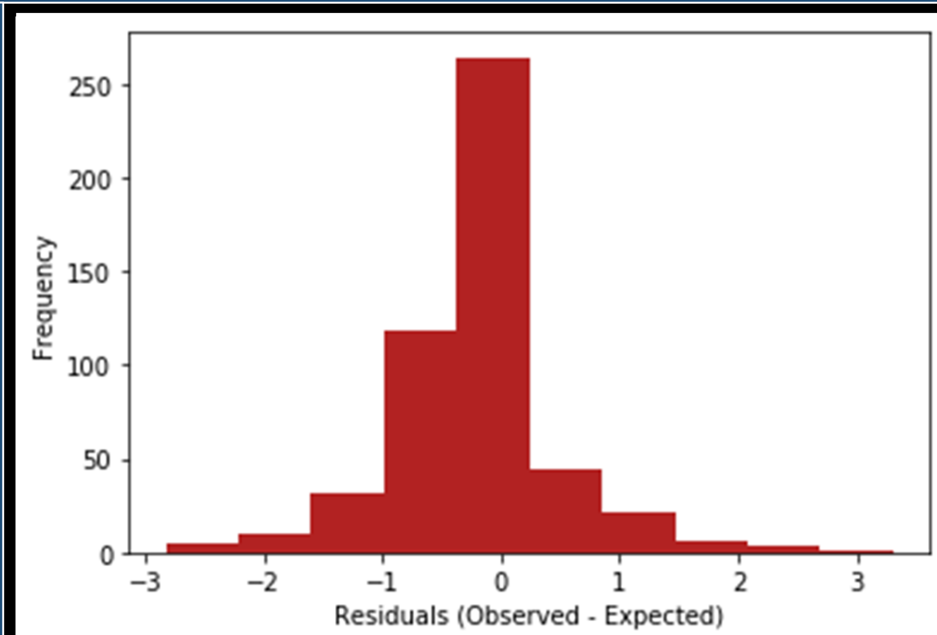
Over-estimates when # people == 0
Under-estimates when # people > 0
Higher variance along the y axis



Results: Model Performance

Plotted Residuals for each model using testing data





Unbalanced Model

Images w/ 0 people >> images w/ 1+ person

Highly Skewed towards -1

Balanced Model

Images w/ 0 people == images w/ 1+ person

Slightly skewed towards -1

Discussion

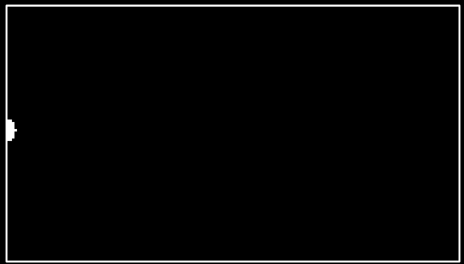
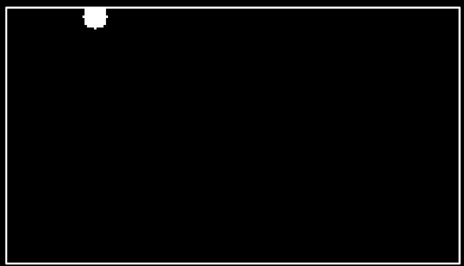
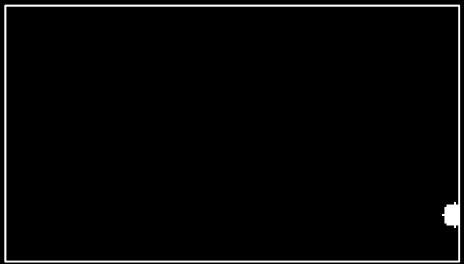
- Sources of Error
  - Different recreational activities yield different abstract idea of humans

≠

  - Solution: Train separate models for separate recreational types
    - Re-label images focusing on each recreational type independently
    - Use present model for transfer-learning
    - Must have many training images with various orientations

Discussion

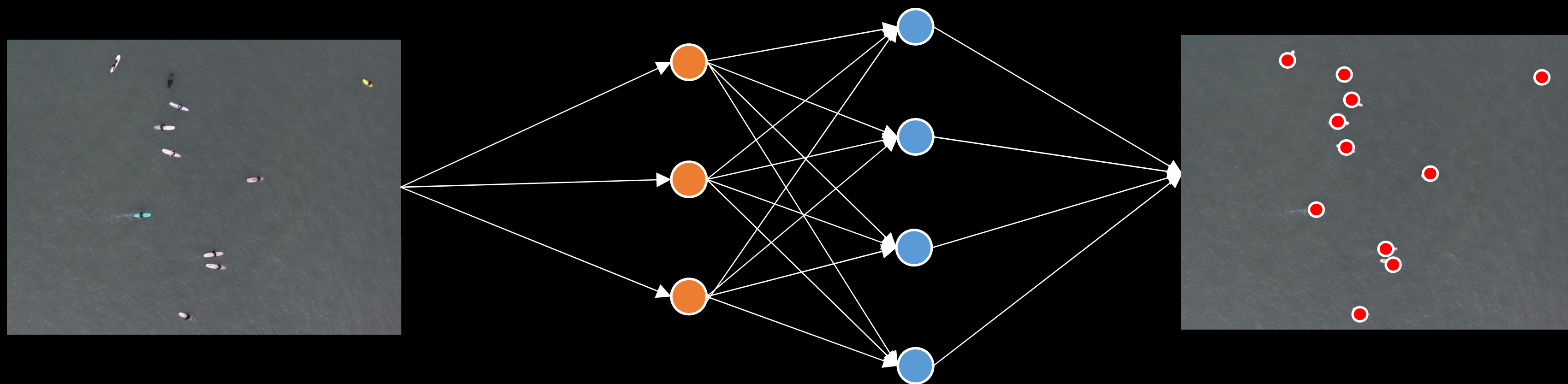
- Sources of Error
  - Splitting images led to instances of only parts of humans being present in the dataset
    - Part of a person in a labeled image still represented 1 person in the model

  - Solution: Consider using a more conservative algorithm to generate the # of humans from the labeled images
    - Of the first 200 images w/ people, 18% had partial humans
    - But most partial humans were in images w/ only 1 human

Discussion

- Future Projects
  - Identify *where* humans are in images



Result → Center coordinates for human locations

- Allows:
  - Calculation of distance between subjects
  - Categorization of ‘interaction’ or ‘solitary’ behavior
  - Examine specific behavioral relationships between subjects

Discussion

Which Model?
Unbalanced or Balanced?

Unbalanced Model

Lower MSE

Residuals skewed towards -1

Less variation in predicted values

Balanced Model

Higher MSE

Residuals more towards 0

More variation in predicted values

After more training: Balanced model will likely produce more better results

Conclusion

Summary

Models performed relatively well on identifying the number of humans in an image.

Purpose

Less time labeling and sorting images with humans

More time running more sophisticated analyses

Audience

Direct benefits for students in the CSULB Shark Lab

Indirect benefits for researchers who have similar data and are trying to answer similar questions