QUEST FOR THE BEST CAT PHOTO
FINAL PRESENTATION

Kyra Ballard
Andrew Fu
Shravan Nageswaran
Emily Xie

Harvard University
IACS Capstone, Spring 2020
ADDITONAL INFORMATION

To see a quick summary of this presentation, see the video below or this poster.
PROBLEM

- About 6.5M dogs and cats each year enter animal shelters, according to the ASPCA.
- But approximately 1.5M of these are put down.

That’s 1 in 4 pets.
The quality of the photo of a pet is correlated to its likelihood of being adopted.
GOALS

increase adoption rates

decrease euthanasia for shelter cats
Animal shelter based in Texas
No-kill shelter
Saved over 70,000 animal lives since 2008

SPONSOR: AUSTIN PETS ALIVE!

Austin Pets Alive! is not your average animal shelter. We pioneer innovative lifesaving programs designed to save the animals most at risk of euthanasia.

Adopt
Looking for a furry friend to add to the family? We have thousands of animals that would love to be part of your home.

Foster
Open your heart and home to a pet in need, and be the bridge to a dog or cat's forever home.

Volunteer
Our volunteers make lifesaving possible – become a volunteer today!
Software company

Primary goals

- Increase adoption
- Decrease euthanasia
- Increase shelter engagement

Algorithm that optimizes image taking

- For the best chance of adoption
Dog model process

- Takes in **video** of animal
- Selects **optimal shot**
- Automatically **edits** image
- Outputs **enhanced optimal** image
IMPACT

124% increase in adoption
41% reduction in euthanasia

27% increase in adoption
56% reduction in euthanasia
# SCOPE OF WORK

<table>
<thead>
<tr>
<th>In scope</th>
<th>Out of scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Model taking cat videos and outputting best frame</td>
<td>● Stylized front-end</td>
</tr>
<tr>
<td>○ Length: &lt;60s</td>
<td>● Measuring adoption rates</td>
</tr>
<tr>
<td>○ Unobstructed view of a single cat</td>
<td></td>
</tr>
<tr>
<td>● Functional web app for mobile devices</td>
<td></td>
</tr>
</tbody>
</table>
# CHALLENGES

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Video Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fur covering face</td>
<td>Unstable camera</td>
</tr>
<tr>
<td>Not facing camera</td>
<td>Camera quality (phone vs. laptop)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small number of cat videos</td>
<td>Environment</td>
</tr>
<tr>
<td>No labeled data</td>
<td>Equipment</td>
</tr>
</tbody>
</table>
THE DATA

The Good

- Full body visible
- Looking directly at camera
- Clear, high quality image
- Good lighting

The Bad

- Full body not visible
- Can’t distinguish facial features
- Looking away from camera
- Blurry image
- Darker area
HEURISTICS

- Detection of Cat Features
- Image Sharpness
- Relative Size of Cat Features
EDA: GENERAL DATA SENSE

- Initial data cleaning yielded 79 videos
- Duration: Avg: 23 seconds. Min 2 seconds. Max 55 seconds
LITERATURE REVIEW: VIOLA-JONES

- Rapid Object Detection using a Boosted Cascade of Simple Features
- Haar-like Features
  - Pre-Compute Integral Image
- AdaBoost on Decision Stumps
- Cascade
- Sliding Windows

High-level view of cascade approach
EDA: CAT FACE DETECTION

- Ran cat face detection using Haar Cascade
  - Low Recall Rate
- Subsample of 10 videos
- Every 10th frame per video
EDA: MEASURES OF SHARPNESS
EDA: HEAD SIZE RATIOS
<table>
<thead>
<tr>
<th>baseline</th>
<th>initial</th>
</tr>
</thead>
<tbody>
<tr>
<td>random image selected from set of frames with cat head detected</td>
<td>image selected from set of frames with cat head detected, with highest combined scores of sharpness and best cat head size</td>
</tr>
</tbody>
</table>
TESTING THE MODELS

Implementation

Blind A/B testing: baseline vs. developed output

Results

64% of the time developed model produced “better” image
LITERATURE REVIEW: YOLO

- You Only Look Once: Unified, Real-Time Object Detection
- Simultaneous Box and Class Proposal
- Simplicity: CNN
- Optimized for Speed

Each grid cell is responsible for producing exactly $B=2$ bounding boxes representing existence of any object with center in the cell.
LITERATURE REVIEW: YOLO

- You Only Look Once: Unified, Real-Time Object Detection
- Simultaneous Box and Class Proposal
- Simplicity: CNN
- Optimized for Speed

YOLO architecture; note only convolutional and fully connected layers
TRAINING YOLO

- YOLO vs Haar Cascade
- 4 Features
  - Eyes, Nose, Ears, Head
- 100 Training Examples
- AWS EC2 g3s.xlarge
  - NVIDIA Tesla M60 GPU
LITERATURE REVIEW: CPBD

- A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD)
- Probabilistic model for sharpness
- Percentage of detected edges where blur is not detected
- \(0 \leq \text{cpbd} \leq 1\)
REGRESSION: LABELING

- Data-driven approach to weighting features
- Likert scale
  - 5 classes
REGRESSION: EDA ON PREDICTORS

Confidence of Ear Detection by Class

Confidence of Nose Detection by Class
REGRESSION: EDA ON PREDICTORS

**Sharpness (CPBD) of Image by Class**

**Ratio of Eye Sizes by Class**
## Logistic Regression using `sklearn`

- **L1 Regularization**
- **most salient features:** confidence of object detection
- trained on dataset with **any** features detected
- Select for highest probability of either class 4 or 5

### REGRESSION RESULTS: INITIAL

<table>
<thead>
<tr>
<th>methodology</th>
<th>results</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Logistic Regression using <code>sklearn</code></td>
<td><strong>User selected</strong></td>
</tr>
<tr>
<td>- L1 Regularization</td>
<td><strong>Percent of time</strong></td>
</tr>
<tr>
<td>- most salient features: confidence of object detection</td>
<td>Frame from model</td>
</tr>
<tr>
<td></td>
<td><strong>77.2%</strong></td>
</tr>
<tr>
<td>- trained on dataset with <strong>any</strong> features detected</td>
<td>Frame from baseline</td>
</tr>
<tr>
<td></td>
<td><strong>8.9%</strong></td>
</tr>
<tr>
<td>- Select for highest probability of either class 4 or 5</td>
<td>neither</td>
</tr>
<tr>
<td></td>
<td><strong>13.9%</strong></td>
</tr>
<tr>
<td></td>
<td>Selected either frame from model or neither</td>
</tr>
</tbody>
</table>
**REGRESSION RESULTS: ITERATION 2**

**methodology**

- Only examine frames in which **all** features detected
  - Inspired by Decision Trees
- Select for highest probability of either class 4 or 5

<table>
<thead>
<tr>
<th>method</th>
<th>results</th>
</tr>
</thead>
<tbody>
<tr>
<td>User selected</td>
<td>Percent of time</td>
</tr>
<tr>
<td>Frame from model</td>
<td>74.6%</td>
</tr>
<tr>
<td>Frame from baseline</td>
<td>3.8%</td>
</tr>
<tr>
<td>neither</td>
<td>21.5%</td>
</tr>
<tr>
<td>Selected either frame from model or neither</td>
<td><strong>96.2%</strong></td>
</tr>
</tbody>
</table>
## MODEL COMPARISONS

<table>
<thead>
<tr>
<th>baseline</th>
<th>initial</th>
<th>final (iter #2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>random image selected from set of frames with cat head detected</td>
<td>image selected from set of frames with cat head detected, with highest combined scores of sharpness and best cat head ratio</td>
<td>maximum probability of class 4 or 5 produced from logistic regression trained on selected features</td>
</tr>
</tbody>
</table>
## MODEL COMPARISONS

<table>
<thead>
<tr>
<th>baseline</th>
<th>initial</th>
<th>final (iter #2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>random image selected from set of frames with cat head detected</td>
<td>image selected from set of frames with cat head detected, with highest combined scores of sharpness and best cat head ratio</td>
<td>maximum probability of class 4 or 5 produced from logistic regression trained on selected features</td>
</tr>
</tbody>
</table>

Performance against baseline:

- **64%**
- **96%**
OTHER MODELS EXPLORED

**Linear Regression**
unexplainable results; didn’t fit our data well enough

**Convolutional Neural Nets**
30 - 40% classification accuracy depending on parameter tuning

**Transfer Learning**
Trained on InceptionV3 Neural Network pretrained on ImageNet
(60% binary accuracy, 24% 5-class accuracy)
WEB INTERFACE

- Simple web application connected to our model
  - Takes in cat video
  - Executes model
  - Returns optimal frame produced by model
WEB INTERFACE: DEMO

We turned this... ... into this!
EXTENSIONS: BG SUBTRACTION
EXTENSIONS

● Parallelize web app
● Make app front-end prettier
● Measure downstream impact: adoption rates
● Even more model refinements
THANK YOU
Questions?