USING MOBILE NEURAL NETWORK FOR PETS CLASSIFICATION

Jie Wen 1155092211

Supervised by Prof. Michael R. LYU
INTRODUCTION

Development of AI
   AlphaGo - AlphaGo Master - AlphaGo Zero
   Number of applicants – Number of university programs

Popularity of Computer Vision
   University (CMU, CUHK)
   Company (Amazon Rekognition)
Facial analysis
Get a complete analysis of facial attributes, including confidence scores. (Your images aren't stored.)

Done with the demo?
Download SDKs

Faces | Confidence

- looks like a face: 99.8%
- appears to be female: 100%
- smiling: 99.4%
- appears to be happy: 93.2%
- wearing eyeglasses: 99.9%
- wearing sunglasses: 97.6%

Choose a sample Image

Use your own image

Upload or
Type or paste image URL Go

Show more
INTRODUCTION

Mobile Devices

- Face detection when taking photo
- Apple FaceID, Facebook, Alipay…

Compared to CV in the server end…
MOTIVATION

Why Mobile

- General public
- Number of User
- Cost of time/money
- Wide usage scenario
MOTIVATION

Mobile implementation

Need cloud server

Need Network
Mobile implementation

Inside Mobile Devices
What dataset to choose?

- BotanWiki -> AnimalWiki
  (Like ImageNet: dog, cat, cow, bird…)

- AnimalWiki -> PetWiki
  (Shiba, Husky, Scottish fold…)

DATASET
About 20000 images for 21(+1) species: Google, Bing, Baidu

17(+1) dogs and 4 cats

Large dog (7): German Shepherd, Greyhound, Saint Bernard, Tibetan Mastiff, Samoyed, Scotch Collie, Husky.

Mid-size dog (5+1): Shiba, Black Shiba, Border Collie, Dalmatian, Shar Pei, Pug.

Small dog (5): Bichon frise, Chihuahua, Corgi, Poodle, Schnauzer.

Cat (4): Bobcat, Persian Cat, Scottish Fold, Siamese Cat.
What model to choose?

Inception (V3)

MobileNet
Before Inception: Con Layer + Pooling Layer
Inception: Bottleneck Layer
Inception V2

BN Layer

5*5 Con Layer → 2 3*3 Layer
Inception V2
## MODEL

### Inception vs Inception V2

<table>
<thead>
<tr>
<th>Network</th>
<th>Top-1 Error</th>
<th>Top-5 Error</th>
<th>Cost Bn Ops</th>
</tr>
</thead>
<tbody>
<tr>
<td>GoogLeNet [20]</td>
<td>29%</td>
<td>9.2%</td>
<td>1.5</td>
</tr>
<tr>
<td>BN-GoogLeNet</td>
<td>26.8%</td>
<td>-</td>
<td>1.5</td>
</tr>
<tr>
<td>BN-Inception [7]</td>
<td>25.2%</td>
<td>7.8</td>
<td>2.0</td>
</tr>
<tr>
<td>Inception-v2</td>
<td>23.4%</td>
<td>-</td>
<td>3.8</td>
</tr>
<tr>
<td>Inception-v2 RMSProp</td>
<td>23.1%</td>
<td>6.3</td>
<td>3.8</td>
</tr>
<tr>
<td>Inception-v2 Label Smoothing</td>
<td>22.8%</td>
<td>6.1</td>
<td>3.8</td>
</tr>
<tr>
<td>Inception-v2 Factorized 7 × 7</td>
<td>21.6%</td>
<td>5.8</td>
<td>4.8</td>
</tr>
<tr>
<td>Inception-v2 BN-auxiliary</td>
<td>21.2%</td>
<td>5.6%</td>
<td>4.8</td>
</tr>
</tbody>
</table>
Inception V3
**MODEL**

Inception (V3)

> 45 layers
## Inception vs Inception V3

<table>
<thead>
<tr>
<th>Network</th>
<th>Crops Evaluated</th>
<th>Top-5 Error</th>
<th>Top-1 Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>GoogLeNet [20]</td>
<td>10</td>
<td>-</td>
<td>9.15%</td>
</tr>
<tr>
<td>GoogLeNet [20]</td>
<td>144</td>
<td>-</td>
<td>7.89%</td>
</tr>
<tr>
<td>VGG [18]</td>
<td>-</td>
<td>24.4%</td>
<td>6.8%</td>
</tr>
<tr>
<td>BN-Inception [7]</td>
<td>144</td>
<td>22%</td>
<td>5.82%</td>
</tr>
<tr>
<td>PReLU [6]</td>
<td>10</td>
<td>24.27%</td>
<td>7.38%</td>
</tr>
<tr>
<td>PReLU [6]</td>
<td>-</td>
<td>21.59%</td>
<td>5.71%</td>
</tr>
<tr>
<td>Inception-v3</td>
<td>12</td>
<td>19.47%</td>
<td>4.48%</td>
</tr>
<tr>
<td>Inception-v3</td>
<td>144</td>
<td><strong>18.77%</strong></td>
<td><strong>4.2%</strong></td>
</tr>
</tbody>
</table>
### Training Accuracy & Model Size

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inception v3</td>
<td>0.98</td>
<td>88M</td>
</tr>
<tr>
<td>MobileNet 100, 224</td>
<td>0.94</td>
<td>10M</td>
</tr>
<tr>
<td>MobileNet 050, 224</td>
<td>0.92</td>
<td>3M</td>
</tr>
<tr>
<td>MobileNet 050, 128</td>
<td>0.91</td>
<td>3M</td>
</tr>
<tr>
<td>MobileNet 035, 224</td>
<td>0.94</td>
<td>2M</td>
</tr>
</tbody>
</table>
Application

1. Take/Choose photo with Inception model
2. Real-time Classify with Inception model
3. Real-time Classify with MobileNet model
Take/Choose photo with Inception model

1. Take a photo in APP
Take/Choose photo with Inception model

1. Take a photo in APP

2. Choose a photo from album: Crop & Not Crop
DEMO

Real-time Classify with Inception model

3. Difficult/Incomplete/Blur pictures/memes
DEM0

Real-time Classify
with Inception model

3. Difficult/Incomplete/Blur pictures/memes

4. Different Angles
Real-time Classify with Inception model

3. Difficult/Incomplete/Blur pictures/memes

4. Different Angles

5. Other Species
Real-time Classify
with Inception model

3. Difficult/Incomplete/Blur pictures/memes

4. Different Angles

5. Other Species

6. Similar Species
   (Shiba – Black Shiba – Husky)
<table>
<thead>
<tr>
<th></th>
<th>SIBERIAN HUSKY</th>
<th>ALASKAN MALAMUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Picture</strong></td>
<td><img src="Image" alt="Siberian Husky" /></td>
<td><img src="Image" alt="Alaskan Malamute" /></td>
</tr>
<tr>
<td><strong>Origin</strong></td>
<td>Siberia</td>
<td>Alaska</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>51 - 60 cm</td>
<td>58 - 71 cm</td>
</tr>
<tr>
<td><strong>Weight</strong></td>
<td>16 - 34 kg</td>
<td>39 - 57 kg</td>
</tr>
<tr>
<td><strong>Function</strong></td>
<td>To carry a light load at moderate speed over great distances</td>
<td>To carry a heavy load</td>
</tr>
<tr>
<td><strong>Eyes</strong></td>
<td>Blue or Brown</td>
<td>Only Brown</td>
</tr>
<tr>
<td><strong>Ears</strong></td>
<td>Set high on the head</td>
<td>Set wide apart on the head</td>
</tr>
<tr>
<td><strong>Tail</strong></td>
<td>Fox brush carried in a sickle</td>
<td>A waving plume</td>
</tr>
<tr>
<td><strong>Personality Traits</strong></td>
<td>Highly Active &amp; Vocal</td>
<td>Laid Back</td>
</tr>
<tr>
<td></td>
<td>Friendly towards other dogs</td>
<td>Gender aggressive towards dogs of the same sex</td>
</tr>
<tr>
<td></td>
<td>No loyalty to one person - they love everyone &amp; everything</td>
<td>Family orientated - Babysat the Mahlemut children in the tribe</td>
</tr>
</tbody>
</table>
Real-time Classify with Inception model
Real-time Classify with MobileNet model
DEMO

MobileNet vs Inception

Speed

Stability

“Unknown” (Human face)

-> Sure about usage: MobileNet, Inception
CONCLUSION

Shortcomings & Future Work

1. Classify “unknown” images: threshold, build an unknown class
   -> detect adversarial samples?
2. Result of MobileNet varies rapidly in some cases.
3. Accuracy from different dataset
4. Evaluate our model
5. UI
6. More similar species (Husky vs Alaskan)
CONCLUSION

Term Review