DogVCat: Create an algorithm to distinguish dogs from cats

Presenter: Bang Liu, Yan Liu, Kai Zhou
Instructor: Dr Russ Greiner
Co-coach: Junfeng Wen
Our Task

❖ **Basic task**
  ➢ To Create an algorithm to classify whether an image contains a **dog** or a **cat**
  ➢ Kaggle Competition

❖ **Dataset: the **ASIRRA** dataset** provided by Microsoft Research
  ➢ Training Dataset: 543.76MB, 25000 images(12500 dogs, 12500 cats)
  ➢ Testing Dataset: 271.15MB, 12500 images
Our Task

❖ **Input & Output**
  ➢ Input: images of *dogs* and *cats*
  ➢ Output: classification accuracy on testing dataset

❖ **Relevance to Machine Learning**
  ➢ **Learning task**
    ■ For *training dataset*, to learn a classification model to determine the decision boundary.
  ➢ **Performance task**
    ■ For *testing dataset*, to make classification for each image based on the learned model, and get the accuracy.
Outlines

- Our Task
- Our Solutions
  - Traditional Model
  - Deep Learning Model
- Our Results
- Future Work
Our Solutions

- **Traditional Model** for pattern recognition
  - Fixed features + Trainable classifiers

  - **Fixed features**
    - Low Level Features: *Color*
    - High Level Features: *Dense SIFT* (Scale-Invariant Feature Transform)
Our Solutions

❖ **Deep Learning Model**

➢ **Trainable features + Trainable classifiers**

➢ **Trainable features**
  ▪ from Deep Neural Networks

![Diagram](image.png)
Our Solutions

❖ Two trainable classifiers
  based on data property (high dimensionality) and previous work
  ➢ Support Vector Machines (SVMs)
  ➢ Deep Neural Networks
Outlines

- Our Task
- Our Solutions
  - Traditional Model
  - Deep Learning Model
- Our Results
- Future Work
Solution 1: Traditional Model

❖ Fixed Features
  ➢ High Level Features
    ■ Dense SIFT (Scale-Invariant Feature Transform)
  ➢ Low Level Features
    ■ Color
Solution 1: Traditional Model

**Feature Extraction - (Dense) SIFT feature**

- The SIFT features are local and based on the appearance of the object.
- Invariant to image scale and rotations

Image gradients → Keypoint descriptor

SIFT(ijcv04) 61.6%
Solution 1: Traditional Model

- Feature Extraction - \textbf{Color}
  - Use \textbf{HSV} (hue, saturation, value) model other than \textbf{RGB} (red, green, blue) model of color
    - Closer to human perception of color
    - Easier to interpret

61.6\% \rightarrow 71.5\%
Solution 1: Traditional Model

❖ Feature Representation - Bag of Words

http://cs.nyu.edu/~fergus/teaching/vision_2012/9_BoW.pdf
Solution 1: Traditional Model

Feature Extraction - Whole Process

1. feature detection
2. create dictionary
3. representation
4. training model
5. get decision boundary

http://cs.nyu.edu/~fergus/teaching/vision_2012/9_Bow.pdf
Solution 1: Traditional Model

❖ Preprocessing: **Image Segmentation**
  ➢ **Segmentation** of pets from backgrounds
    -- complicated and various backgrounds
  ➢ **Results** (good & bad):

  ➢ **Performance:** no improvement
Outlines

- Our Task
- Our Solutions
  - Traditional Model
  - Deep Learning Model
- Our Results
- Future Work
Solution 2: Deep Learning

- Representations are hierarchical and trained
Solution 2: Deep Learning

- Architecture of model: previous work

(Alex Krizhevsky, ImageNet 2012, 60 million parameters, 650,000 neurons, trained on 1.2 million training images)
Solution 2: Deep Learning

- Convolutional layers: feature maps

http://www1.i2r.a-star.edu.sg/~irkhan

(UFLDL)
Solution 2: Deep Learning

- Max pooling: sub-sampling

(Matthew D. Zeiler, 2013)
Solution 2: Deep Learning

- Fully connected layers: image representation
Solution 2: Deep Learning

- Our model: Learned feature + Classifier

![Diagram of a deep learning model]

Classifier (Neural Network, SVM) → Output (1: dog, 0: cat)

93.7%
Solution 2: Deep Learning

Top activation for Layer 1, 2

Layer 1:
edges, colors (low-level feature)

Layer 2:
corners, edge/color conjunctions (low-level feature)

(Matthew D. Zeiler, 2013)
Solution 2: Deep Learning

❖ Top activation for Layer 3

textures (mid-level feature)

(Matthew D. Zeiler, 2013)
Solution 2: Deep Learning

- Top activation for Layer 4

part of objects (high-level feature)

(Matthew D. Zeiler, 2013)
Solution 2: Deep Learning

- Top activation for Layer 5 (objects)

entire objects (high-level feature)

(Matthew D. Zeiler, 2013)
Outlines

- Our Task
- Our Solutions
  - Traditional Model
  - Deep Learning Model
- Our Results
- Future Work
Our Results

❖ **Performance** on Fixed Features

➢ **Fixed features + SVMs**
  - Dense SIFT features
    - 0.61804 → 0.65067 → 0.67600
  - Dense SIFT + Color features
    - 0.54960 → 0.55867 → 0.71467
  - Dense SIFT features (with segmentation)
    - 0.49600 → 0.50133
Our Results

❖ Performance on Trainable Features

➢ Trainable Features + Deep Neural Network
  ■ Original neural network (Alex 2012)
    0.83200
  ■ Decaf features + BP Neural Network
    0.93013
  ■ Decaf features + SVM (RBF Kernel)
    0.93787
Our Results

❖ Increasing Performance
### Our Results

- **Leaderboard on Kaggle** (ranked 9th in 71 teams)

<table>
<thead>
<tr>
<th>#</th>
<th>Team Name</th>
<th>Score</th>
<th>Entries</th>
<th>Last Submission UTC (Best - Last Submission)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DaggerFS</td>
<td>0.97040</td>
<td>4</td>
<td>Mon, 18 Nov 2013 17:54:09 (-3d)</td>
</tr>
<tr>
<td>2</td>
<td>Charlie</td>
<td>0.96987</td>
<td>5</td>
<td>Sat, 16 Nov 2013 17:30:20 (-30.6d)</td>
</tr>
<tr>
<td>3</td>
<td>Jeff</td>
<td>0.96773</td>
<td>2</td>
<td>Thu, 26 Sep 2013 13:54:31</td>
</tr>
<tr>
<td>4</td>
<td>wqren</td>
<td>0.96667</td>
<td>4</td>
<td>Wed, 23 Oct 2013 02:12:19 (-1.1h)</td>
</tr>
<tr>
<td>5</td>
<td>Kyle Kastner</td>
<td>0.96667</td>
<td>10</td>
<td>Mon, 25 Nov 2013 15:17:11 (-4.1d)</td>
</tr>
<tr>
<td>6</td>
<td>Daniel Nouri</td>
<td>0.96587</td>
<td>8</td>
<td>Fri, 29 Nov 2013 19:11:26 (-10.9d)</td>
</tr>
<tr>
<td>7</td>
<td>hungry red panda</td>
<td>0.96427</td>
<td>2</td>
<td>Wed, 27 Nov 2013 09:17:39</td>
</tr>
<tr>
<td>8</td>
<td>nayoji</td>
<td>0.94933</td>
<td>10</td>
<td>Thu, 24 Oct 2013 19:00:57 (-21h)</td>
</tr>
<tr>
<td>9</td>
<td><strong>UA_551</strong></td>
<td>0.93787</td>
<td>18</td>
<td>Tue, 03 Dec 2013 01:55:13 (-0.7h)</td>
</tr>
<tr>
<td>10</td>
<td>Poly</td>
<td>0.92000</td>
<td>4</td>
<td>Sun, 17 Nov 2013 09:19:38 (-4d)</td>
</tr>
</tbody>
</table>
Outlines

- Our Task
- Our Solutions
  - Traditional Model
  - Deep Learning Model
- Our Results
- Future Work
Future Work

❖ What we are going to do...
➢ to achieve **higher performance**

- different parameter settings for SVMs and Deep Neural Networks
- Features combination
  thanks to Dr Russ Greiner
- Face detection
- multi classifiers for different types of images
  thanks to Dr Mohamed Elgendi
Done

- Works:
  - Combined SIFT with Color features
  - Tried different types of Decaf features
  - SVMs on Decaf features

- Doesn’t work:
  - Image Segmentation

ToDo

- Combine Decaf features with other features
- Face Detection


