HOGgles: Visualizing Object Detection Features

Carl Vondrick   Aditya Khosla   Tomasz Malisiewicz   Antonio Torralba

Massachusetts Institute of Technology
Computer Science and Artificial Intelligence Laboratory
Aeroplane
Aeroplane
Car
Car
What information does HOG have?
What information does HOG have?
What information does HOG have?
What information does HOG have?

Image

HOG

Nearest Neighbors
What information does HOG have?

Image

HOG

Nearest Neighbors
What information does HOG have?

Image

HOG

Nearest Neighbors
What information does HOG have?

Image

HOG

Nearest Neighbors
What information does HOG have?

Image

HOG

Nearest Neighbors
What information is lost?
What information is lost?
What information is lost?

$$\min_{x \in \mathbb{R}^d} \| \phi(x) - y \|_2^2$$
What information is lost?
Method: Paired Dictionary
Method: Paired Dictionary

\[ = \alpha_1 + \alpha_2 + \ldots + \alpha_k \]
Method: Paired Dictionary

\[ = \alpha_1 + \alpha_2 + \ldots + \alpha_k \]
Method: Paired Dictionary

\[ \alpha_1 + \alpha_2 + \ldots + \alpha_k = \]
Method: Paired Dictionary

\[ = \alpha_1 + \alpha_2 + \ldots + \alpha_k \]

\[ \alpha_1 \rightarrow \]  
\[ \alpha_2 \rightarrow \]  
\[ \ldots \rightarrow \]  
\[ \alpha_k \rightarrow \]  

= Image of a person
Method: Paired Dictionary

\[ \hat{y} = f(x) = V\hat{\alpha} \]

where \( \hat{\alpha} = \arg \min_{\alpha} \| x - U\alpha \|_2^2 \) s.t. \( \| \alpha \|_1 \leq \lambda \)
Paired Dictionary Learning

\[
\arg \min_{U,V,\alpha} \sum_{i=1}^{N} \|x_i - U\alpha_i\|_2^2 + \|y_i - V\alpha_i\|_2^2
\]

s.t. \(\|\alpha_i\|_1 \leq \lambda, \|U\|_2^2 \leq \gamma_1, \|V\|_2^2 \leq \gamma_2\)

Just sparse coding!
(Optimize using off the shelf solvers)
A microscope to view HOG
A microscope to view HOG

2x more intuitive
Human Vision

vs

HOG Vision
Human Vision vs HOG Vision
The HOGgles Challenge
The HOGgles Challenge

Clap your hands when you see a person
The HOGgles Challenge
The HOGgles Challenge
The HOGgles Challenge

Humans detect & DPMs detect
The HOGgles Challenge

Humans miss & DPM miss
Chair Detections
Chair Detections
Car Detections
Car Detections
HOG+Human Detector

![Precision vs Recall Graph]

Chair
HOG+Human Detector

![Graph showing precision-recall curve for Chair detection with HOG+DPM]
HOG+Human Detector

![Precision vs Recall for Chair detection with HOG+DPM and RGB+Human](image)
HOG+Human Detector

![Precision-Recall Curve for Chair Detection](image)

- **HOG+DPM**
- **RGB+Human**
- **HOG+Human**
HOG+Human Detector
HOG+Human Detector

Poor despite perfect learning
HOG+Human Detector

![Graph showing precision-recall curves for different methods.]

- HOG+DPM
- RGB+Human
- HOG+Human

0.5 correlation coefficient
HOG+Human Detector

Amount of “juice” that HOG lost
HOG+Human Detector

Chair

Recall

Precision

HOG+Human AP = 0.63
RGB+Human AP = 0.96
HOG+DPM AP = 0.51

Cat

Recall

Precision

HOG+Human AP = 0.78
HOG+DPM AP = 0.58

Car

Recall

Precision

HOG+Human AP = 0.83
HOG+DPM AP = 0.87

Person

Recall

Precision

HOG+Human AP = 0.69
HOG+DPM AP = 0.79
Comparing Features

Grandma's Girls
Comparing Features

Grandma's Girls

HOG
Comparing Features

HSC

HOG
Comparing Features

HSC

HOG
Comparing Features

What HOG sees, but HSC does not
Inverting Deep Learning
Inverting Deep Learning

Last convolutional layer of Krizhevsky et al.
Code from DeCAF by Donahue et al.
Extracting Color from HOG

HOG | Image
Extracting Color from HOG

HOG

Image
Visualizing Learned Models
Visualizing Learned Models

Car
Visualizing Learned Models

Car

Person
Visualizing Learned Models

Car  Person  Bottle
Visualizing Learned Models

Car
Person
Bottle
Bicycle
Visualizing Learned Models

Car  Person  Bottle  Bicycle

Motorbike  Chair  TV  Horse
Why did the detector fail?
Why did the detector fail?
Why did the detector fail?
Code Available

ihog = invertHOG(feat);

Download at mit.edu/hoggles
Thanks!

Code + demos + more: 
mit.edu/hoggles
## Recovering Pixels

<table>
<thead>
<tr>
<th>Category</th>
<th>ELDA</th>
<th>Ridge</th>
<th>Direct</th>
<th>PairDict</th>
</tr>
</thead>
<tbody>
<tr>
<td>bicycle</td>
<td>0.452</td>
<td>0.577</td>
<td>0.513</td>
<td>0.561</td>
</tr>
<tr>
<td>bottle</td>
<td>0.697</td>
<td>0.683</td>
<td>0.666</td>
<td>0.671</td>
</tr>
<tr>
<td>car</td>
<td>0.668</td>
<td>0.677</td>
<td>0.652</td>
<td>0.639</td>
</tr>
<tr>
<td>cat</td>
<td>0.749</td>
<td>0.712</td>
<td>0.687</td>
<td>0.705</td>
</tr>
<tr>
<td>chair</td>
<td>0.666</td>
<td>0.621</td>
<td>0.604</td>
<td>0.617</td>
</tr>
<tr>
<td>table</td>
<td>0.656</td>
<td>0.617</td>
<td>0.582</td>
<td>0.614</td>
</tr>
<tr>
<td>horse</td>
<td>0.686</td>
<td>0.633</td>
<td>0.586</td>
<td>0.635</td>
</tr>
<tr>
<td>motorbike</td>
<td>0.573</td>
<td>0.617</td>
<td>0.549</td>
<td>0.592</td>
</tr>
<tr>
<td>person</td>
<td>0.696</td>
<td>0.667</td>
<td>0.646</td>
<td>0.646</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>0.671</strong></td>
<td><strong>0.656</strong></td>
<td><strong>0.620</strong></td>
<td><strong>0.637</strong></td>
</tr>
</tbody>
</table>
HOGgles on MTurk

Image Classification

In each image below, indicate whether you think it is a Chair or not. We know the answers to some of the easy images, and we will use this to verify your work.  

Showing image 1 of 10

Is this image a Chair?

○ Image is NOT a Chair (press n)  ○ Image is a Chair (press y)

Previous Image  Next Image

Submit HIT
## Recovering Semantics

<table>
<thead>
<tr>
<th>Category</th>
<th>ELDA</th>
<th>Ridge</th>
<th>Direct</th>
<th>PairDic</th>
<th>Glyph</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>bicycle</td>
<td>0.327</td>
<td>0.127</td>
<td>0.362</td>
<td>0.307</td>
<td>0.405</td>
<td>0.438</td>
</tr>
<tr>
<td>bottle</td>
<td>0.269</td>
<td>0.282</td>
<td>0.283</td>
<td>0.446</td>
<td>0.312</td>
<td>0.222</td>
</tr>
<tr>
<td>car</td>
<td>0.397</td>
<td>0.457</td>
<td>0.617</td>
<td>0.585</td>
<td>0.359</td>
<td>0.389</td>
</tr>
<tr>
<td>cat</td>
<td>0.219</td>
<td>0.178</td>
<td>0.381</td>
<td>0.199</td>
<td>0.139</td>
<td>0.286</td>
</tr>
<tr>
<td>chair</td>
<td>0.099</td>
<td>0.239</td>
<td>0.223</td>
<td>0.386</td>
<td>0.119</td>
<td>0.167</td>
</tr>
<tr>
<td>table</td>
<td>0.152</td>
<td>0.064</td>
<td>0.162</td>
<td>0.237</td>
<td>0.071</td>
<td>0.125</td>
</tr>
<tr>
<td>horse</td>
<td>0.260</td>
<td>0.290</td>
<td>0.354</td>
<td>0.446</td>
<td>0.144</td>
<td>0.15</td>
</tr>
<tr>
<td>motorbike</td>
<td>0.221</td>
<td>0.232</td>
<td>0.396</td>
<td>0.224</td>
<td>0.298</td>
<td>0.35</td>
</tr>
<tr>
<td>person</td>
<td>0.458</td>
<td>0.546</td>
<td>0.502</td>
<td>0.676</td>
<td>0.301</td>
<td>0.375</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>0.282</strong></td>
<td><strong>0.258</strong></td>
<td><strong>0.355</strong></td>
<td><strong>0.383</strong></td>
<td><strong>0.191</strong></td>
<td><strong>0.223</strong></td>
</tr>
</tbody>
</table>