HOGgles

Visualizing Object Detection Features

S.Bak

http://web.mit.edu/vondrick/ihog/

HOGgles: Visualizing Object Detection Features

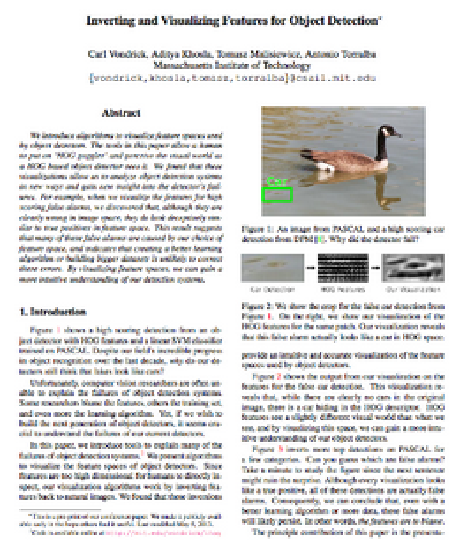
Carl Vondrick Aditya Khosla Tomasz Malisiewicz Antonio Torralba

Massachusetts Institute of Technology

Oral presentation at ICCV 2013

We introduce algorithms to visualize feature spaces used by object detectors. The tools in this paper allow a human to put on "HOG goggles" and perceive the visual world as a HOG based object detector sees it.

Check out this page for a few of our experiments, and read our paper for full details. Code is available to make your own visualizations.


Inverting and Visualizing Features for Object Detection

Carl Vondrick, Aditya Khosla, Tomasz Malisiewicz, Antonio Torralba

Massachusetts Institute of Technology

Abstract

1. Introduction

Download Paper

Read about it in the MIT news!

Download slides or watch

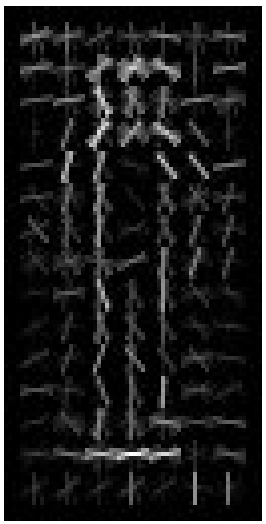
Quick Jump:

1. Code
2. Overview
3. Why did my detector fail?
4. Visualizing Top Detections
5. What does HOG see?
6. Eye Glass
7. Visualizing Learned Models
8. Recovering Color
9. Videos
10. HOGgles

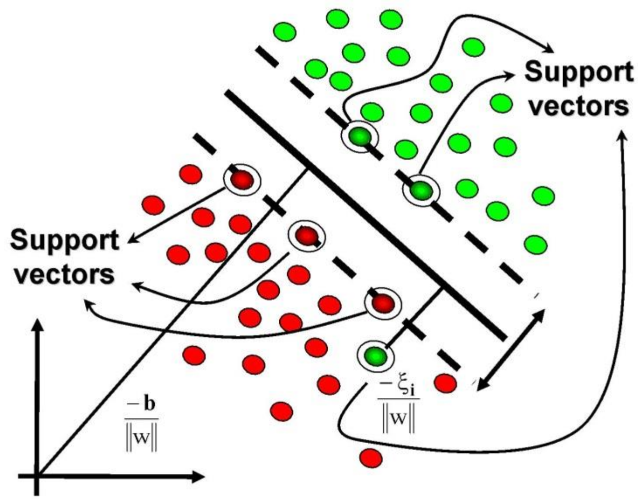

Chair
Car
Aeroplane
Aeroplane
Person

Common detectors…

Feature (e.g. HOG)



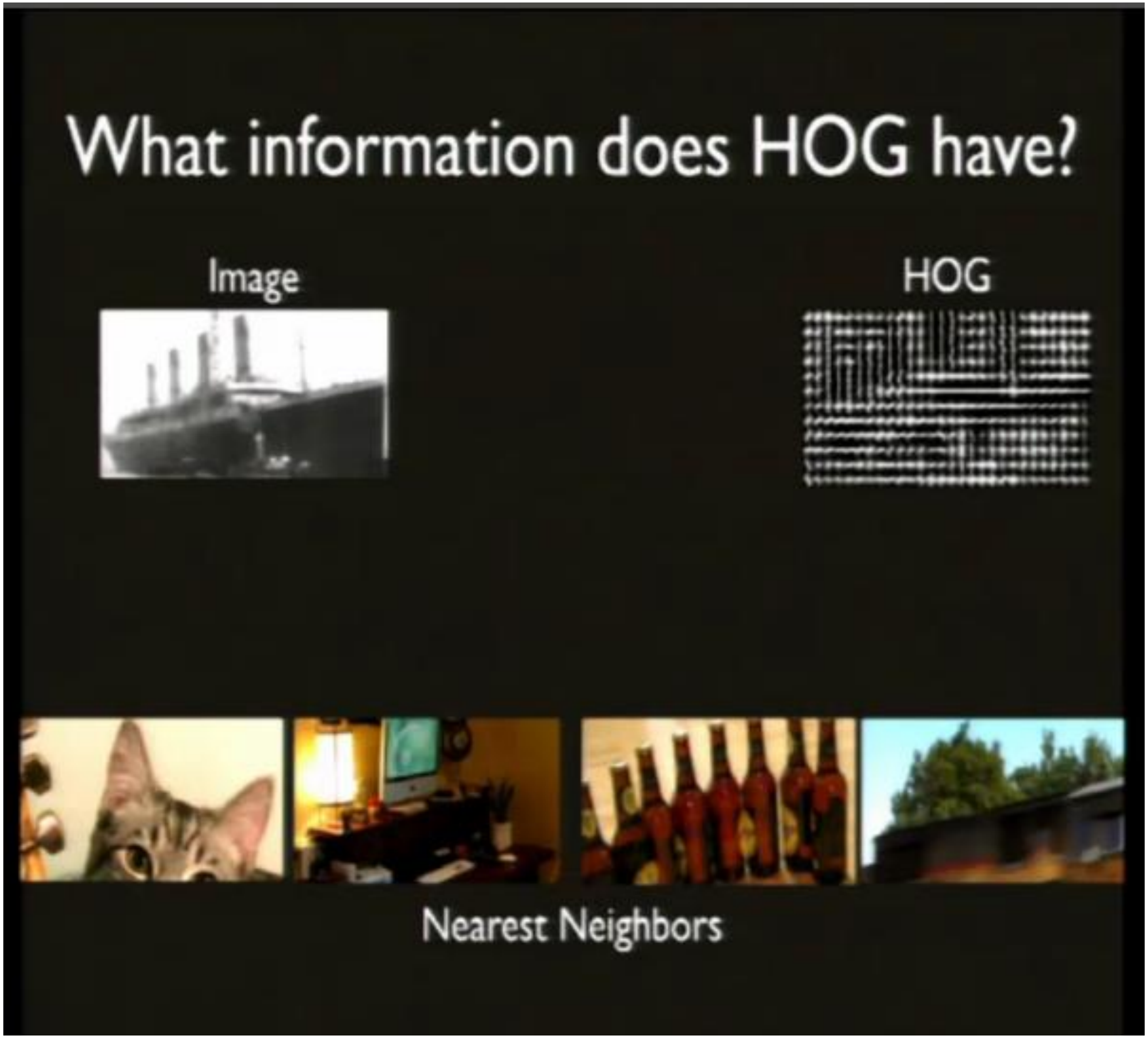
Classifier (e.g. SVM)



Training data

DETECTOR

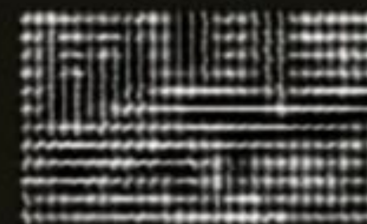
“Some researchers blame the **features**, others the **training set**, and even more the **learning algorithm**.”


What information does HOG have?
Image
HOG
Nearest Neighbors

What information does HOG have?

Image

HOG

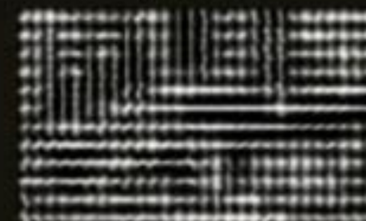


Nearest Neighbors

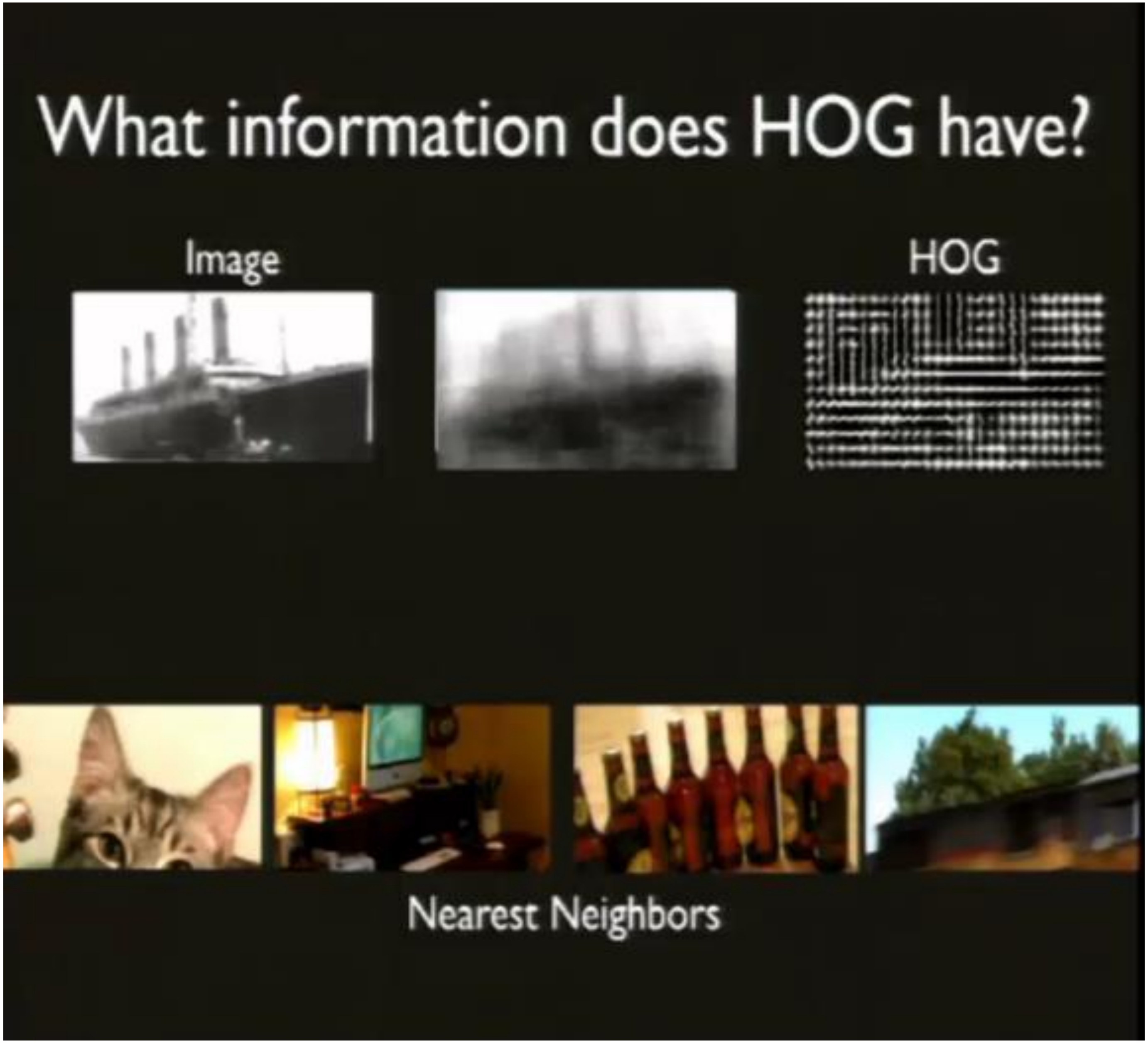
What information does HOG have?

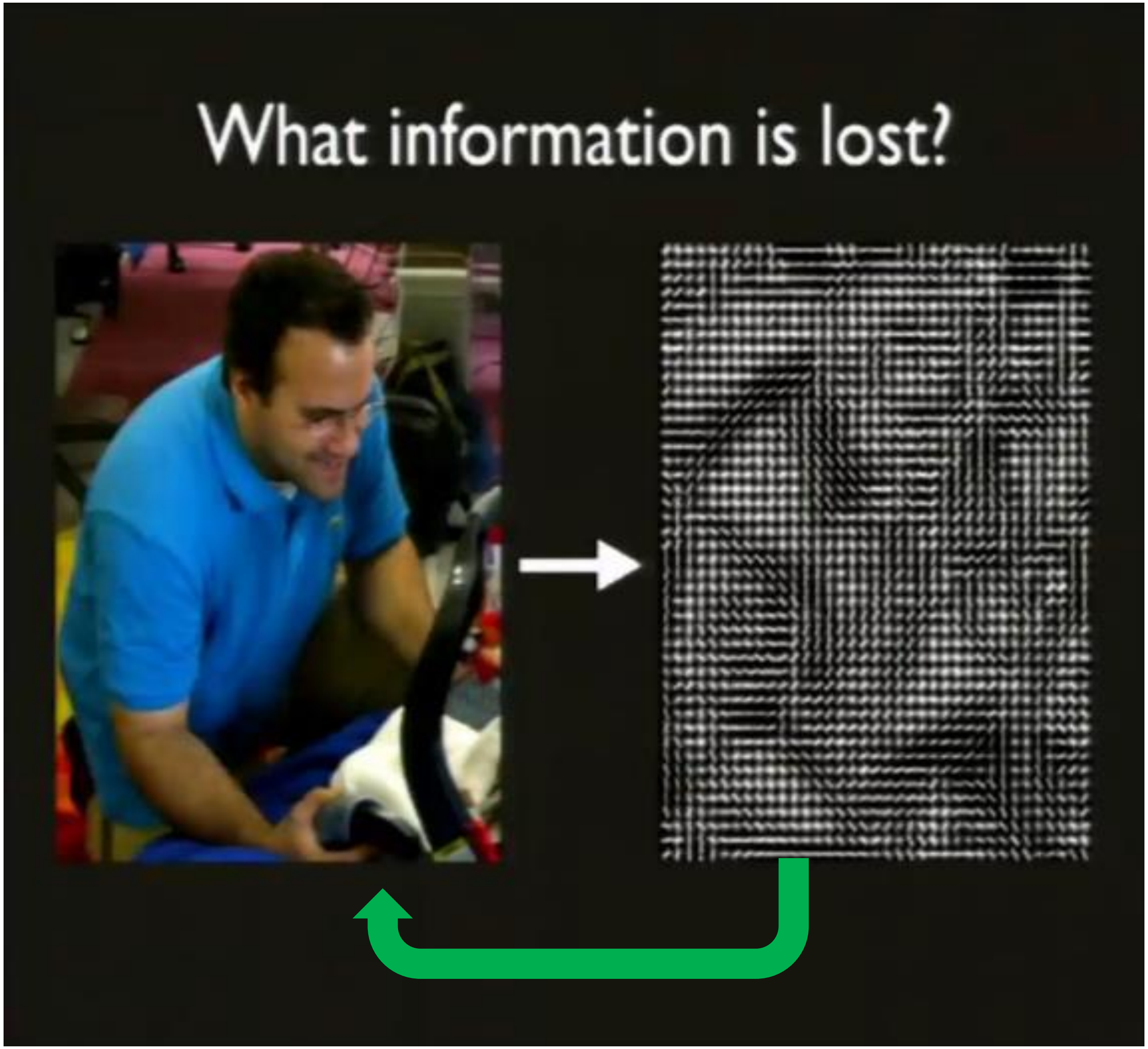
Image

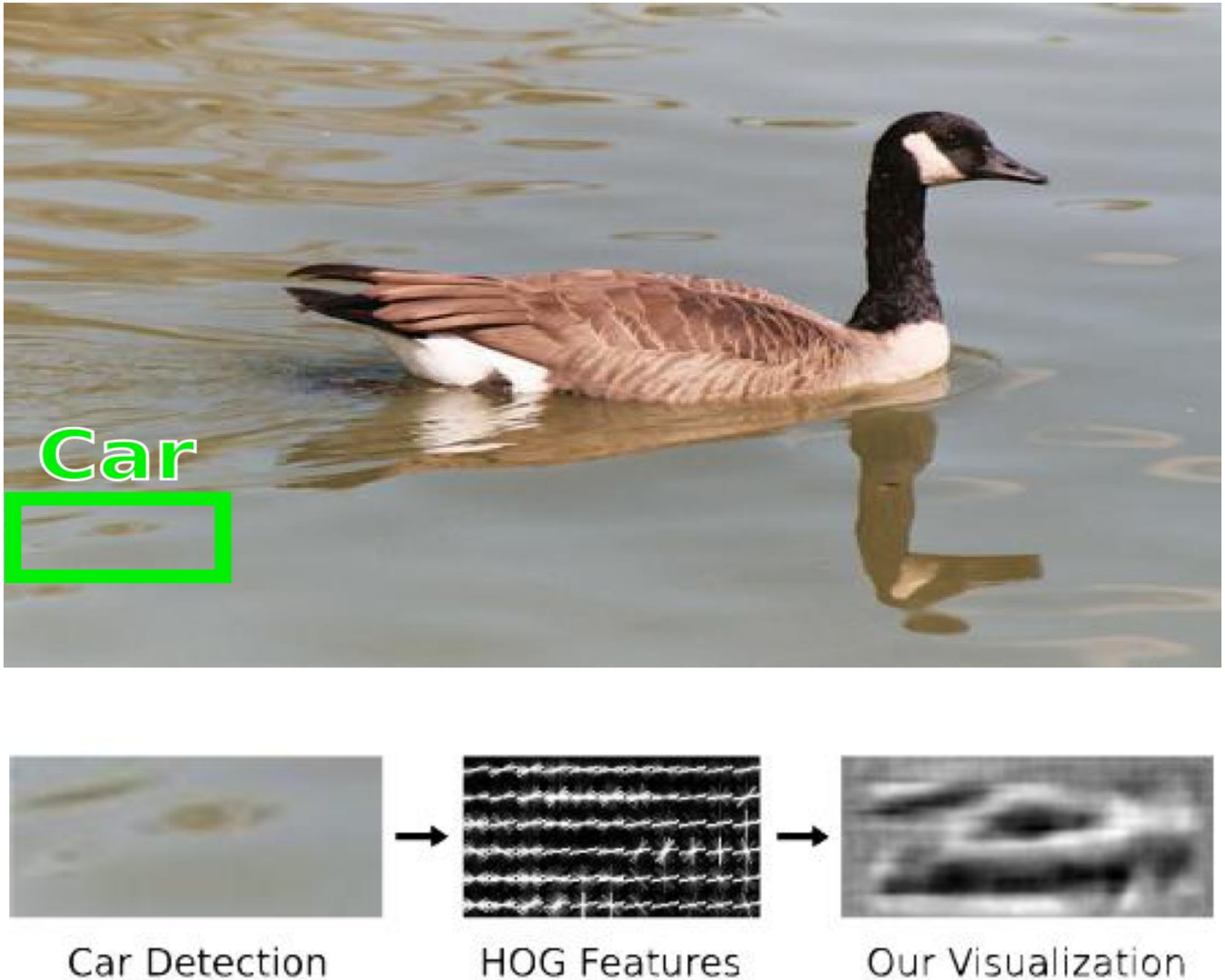
HOG



Nearest Neighbors

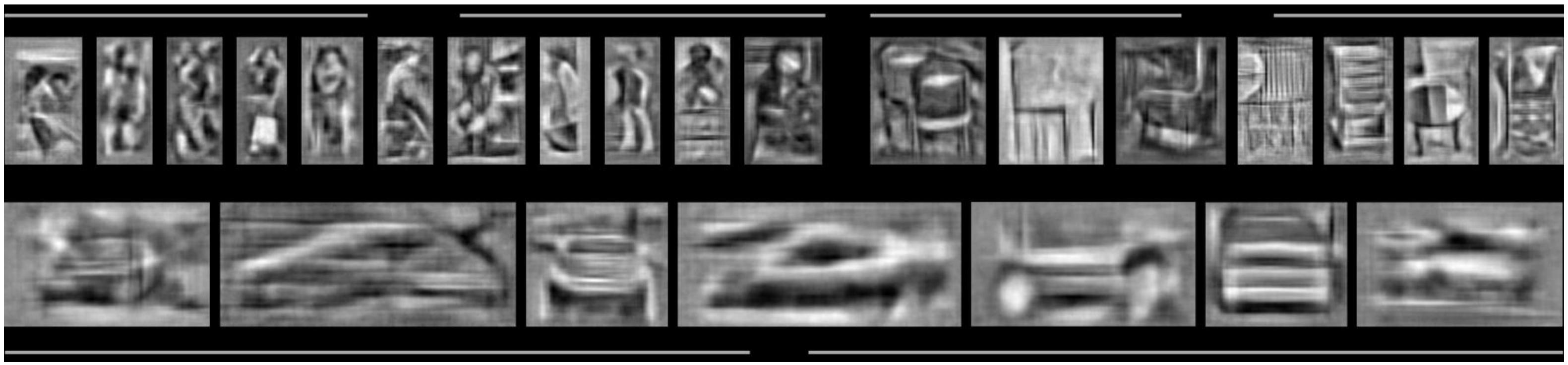

What information does HOG have?
Image
HOG
Nearest Neighbors

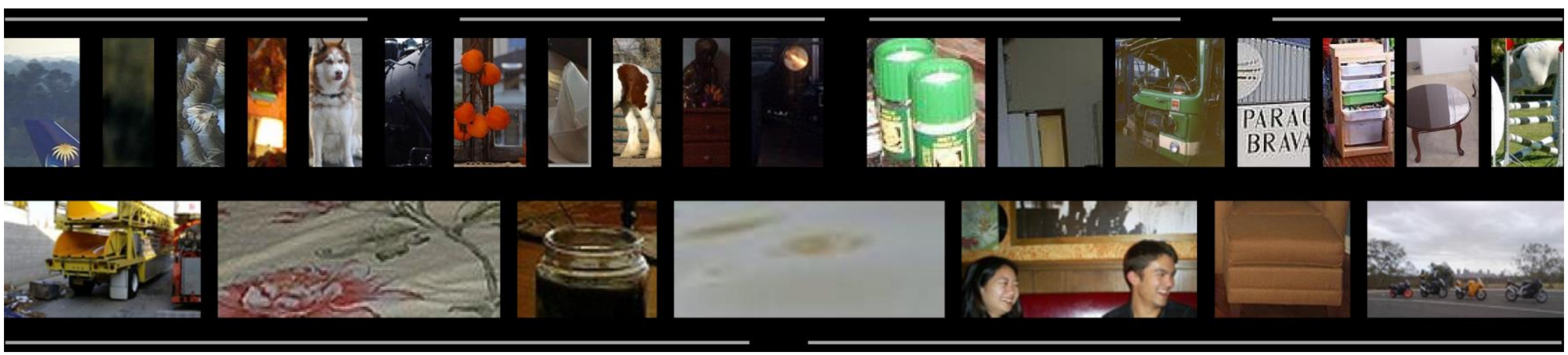

What information is lost?


Car
Car Detection
HOG Features
Our Visualization

Person, chair, and car
Can you guess which are **false** alarms?

High scoring detections from the deformable parts model (DPM) for person, chair, and car.





ALL ARE FALSE ALARMS: Consequently, even with a better learning algorithm or more data, these false alarms will likely persist. In other words, **the features are to blame.**

Inverting HOG descriptor

Let $x \in \mathbb{R}^D$ be an image
$y = \phi(x)$ be the corresponding HOG feature descriptor.

$$\phi^{-1}(y) = \ ?$$

HOG is highly sensitive to noise and the equation has frequent local minima ☹

The paper presents 4 algorithms to invert HOG

- 3 base lines
- Paired dictionary learning

Paired dictionary learning

first K eigenvectors of $\Sigma_{XX} \in \mathbb{R}^{D\times D}$

Let $x \in \mathbb{R}^D$ be an image and $y \in \mathbb{R}^d$ be its HOG descriptor.

Suppose we write x and y in terms of bases $U \in \mathbb{R}^{D\times K}$ and $V \in \mathbb{R}^{d\times K}$ respectively, but with shared coefficients $\alpha \in \mathbb{R}^K$:

$$x = U\alpha \quad \text{and} \quad y = V\alpha \tag{5}$$

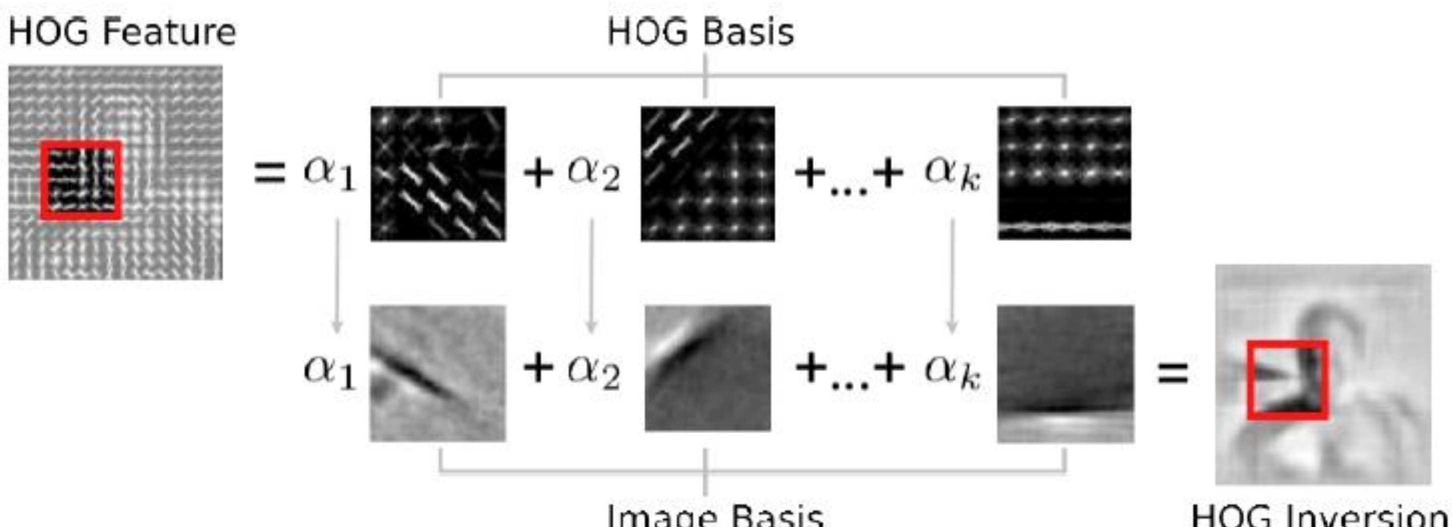
Paired dictionaries require finding appropriate bases U and V such that above equation holds. We solve a paired dictionary learning problem, inspired by recent super resolution ***sparse*** *coding work.*

$$\begin{aligned} &\underset{U,V,\alpha}{\operatorname{argmin}} \sum_{i=1}^{N} \left(||x_i - U\alpha_i||_2^2 + ||\phi(x_i) - V\alpha_i||_2^2\right) \\ &\text{s.t.} \quad ||\alpha_i||_1 \leq \lambda \ \forall i, \ ||U||_2^2 \leq \gamma_1, \ ||V||_2^2 \leq \gamma_2 \end{aligned} \tag{7}$$

Paired dictionary learning

The key observation is that inversion can be obtained by first projecting the HOG features y onto the HOG basis V, then projecting α into the natural image basis U:

$$\phi_D^{-1}(y) = U\alpha^* \\ \text{where} \quad \alpha^* = \underset{\alpha \in \mathbb{R}^K}{\operatorname{argmin}} ||V\alpha - y||_2^2 \quad \text{s.t.} \quad ||\alpha||_1 \leq \lambda \tag{6}$$



Paired dictionary learning

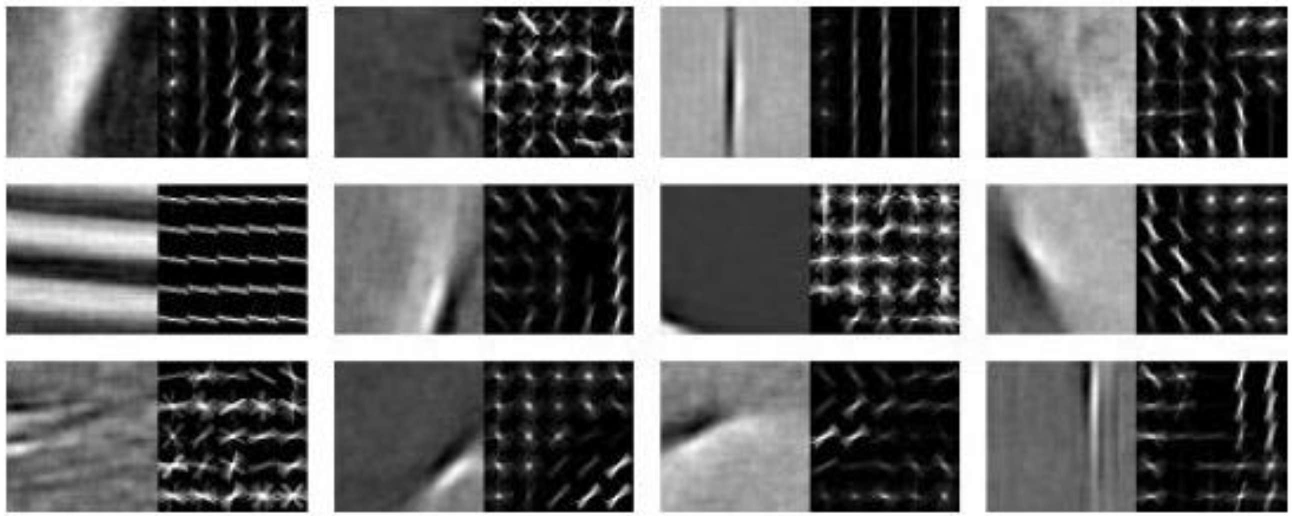
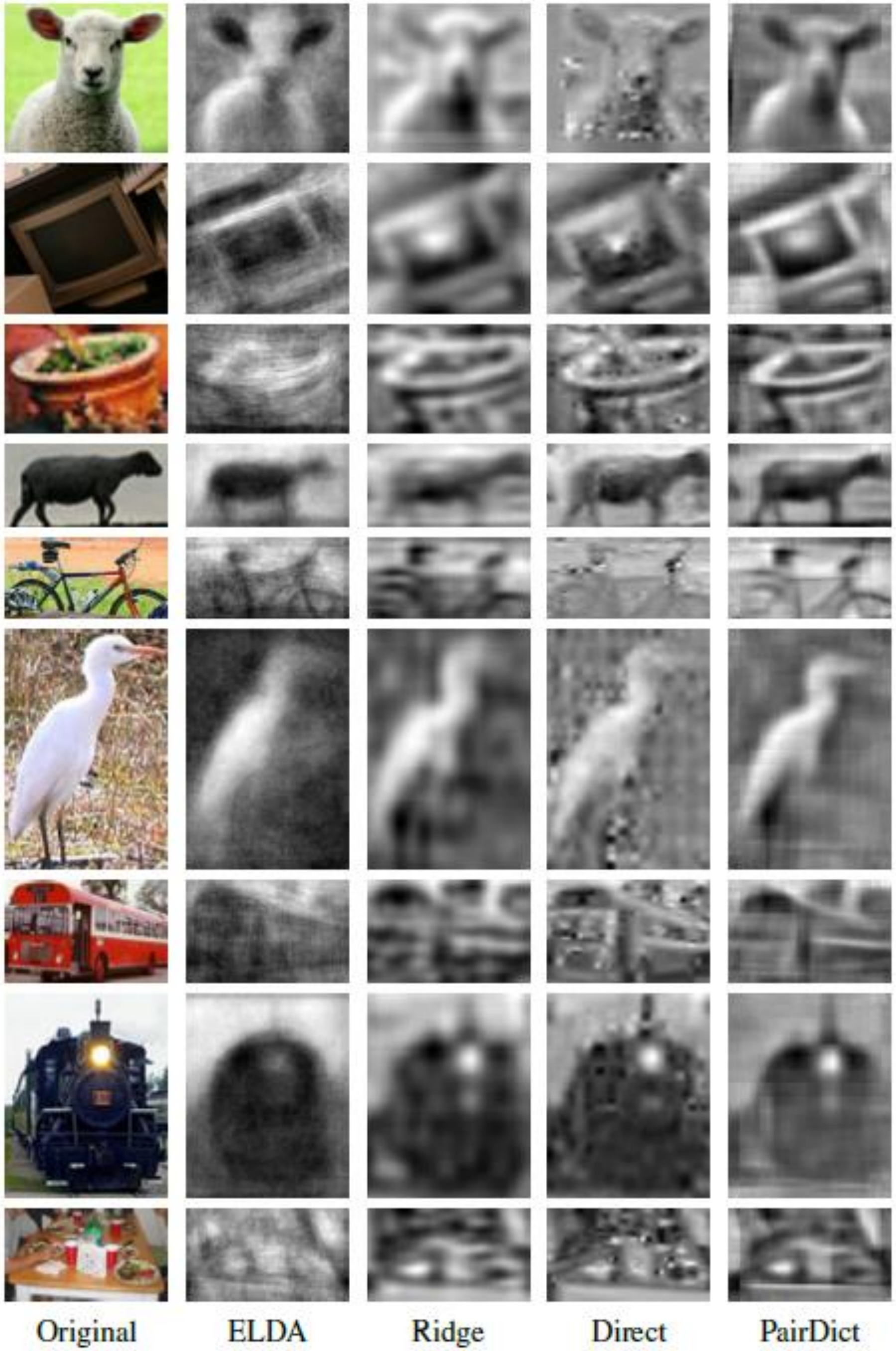


Figure 7: Some pairs of dictionaries for U and V. The left of every pair is the gray scale dictionary element and the right is the positive components elements in the HOG dictionary. Notice the correlation between dictionaries.


Original
ELDA
Ridge
Direct
PairDict

What object detectors see?

(a) Human Vision

(b) HOG Vision

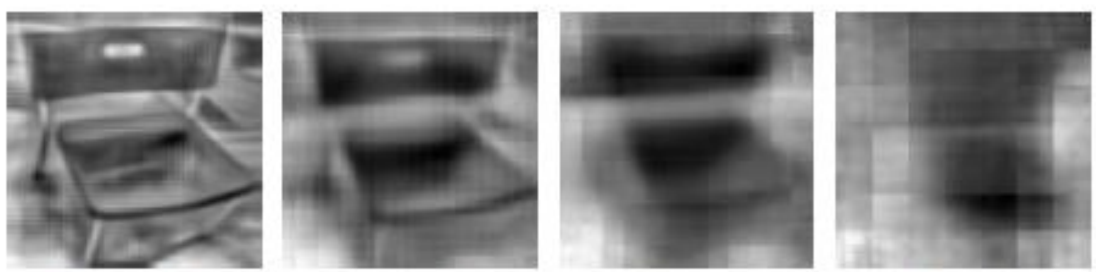
Estimating color images

Limitations

- Not optimal

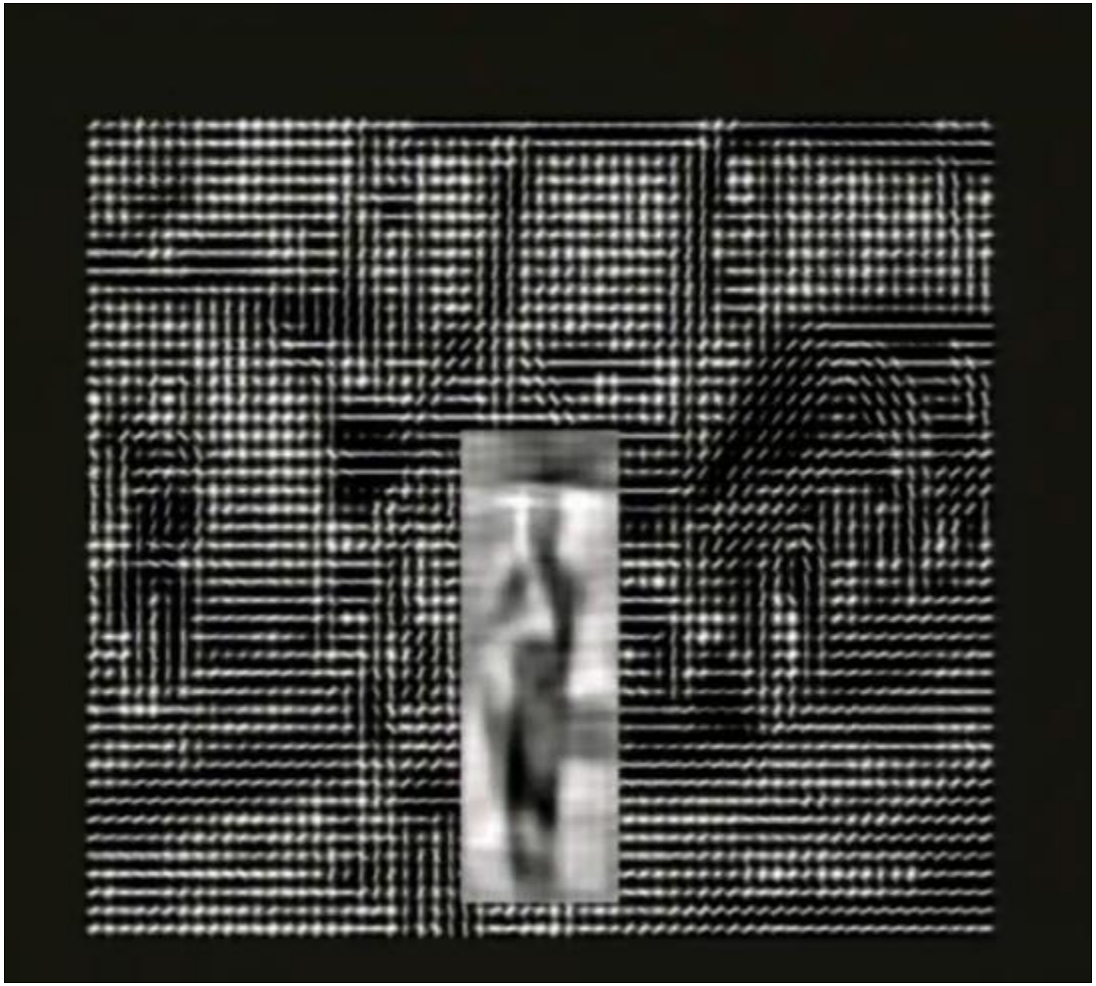
PairDict (seconds) | Greedy (days) | Original

- Template size dependency

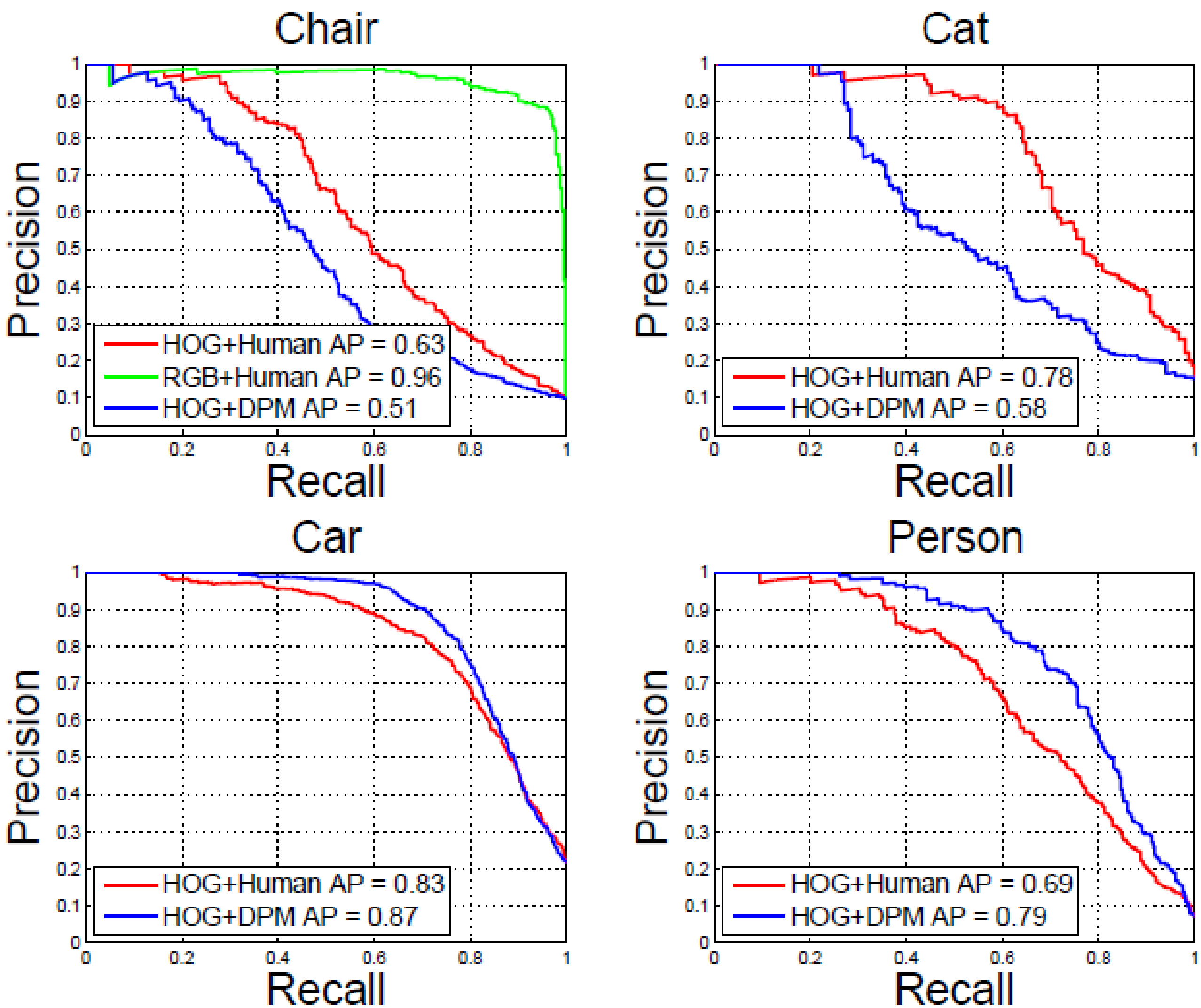


40×40 | 20×20 | 10×10 | 5×5

HOG+Human vs HOG+DPM



HOG+Human vs HOG+DPM



Message to Go

- tool for visualizing object detection features
- choice of feature matters
- DPM is close to the performance limit of HOG

THE END

Thank you for your attention