UNIVERSITY OF LEEDS

Color Segmentation Features
and
Shared Parts for Deformable Part-based Models

Patrick Ott, Mark Everingham

Computer Vision Group, School of Computing

Object Detection?

UNIVERSITY OF LEEDS

- Goal: detect all instances of an object class (person) in an image

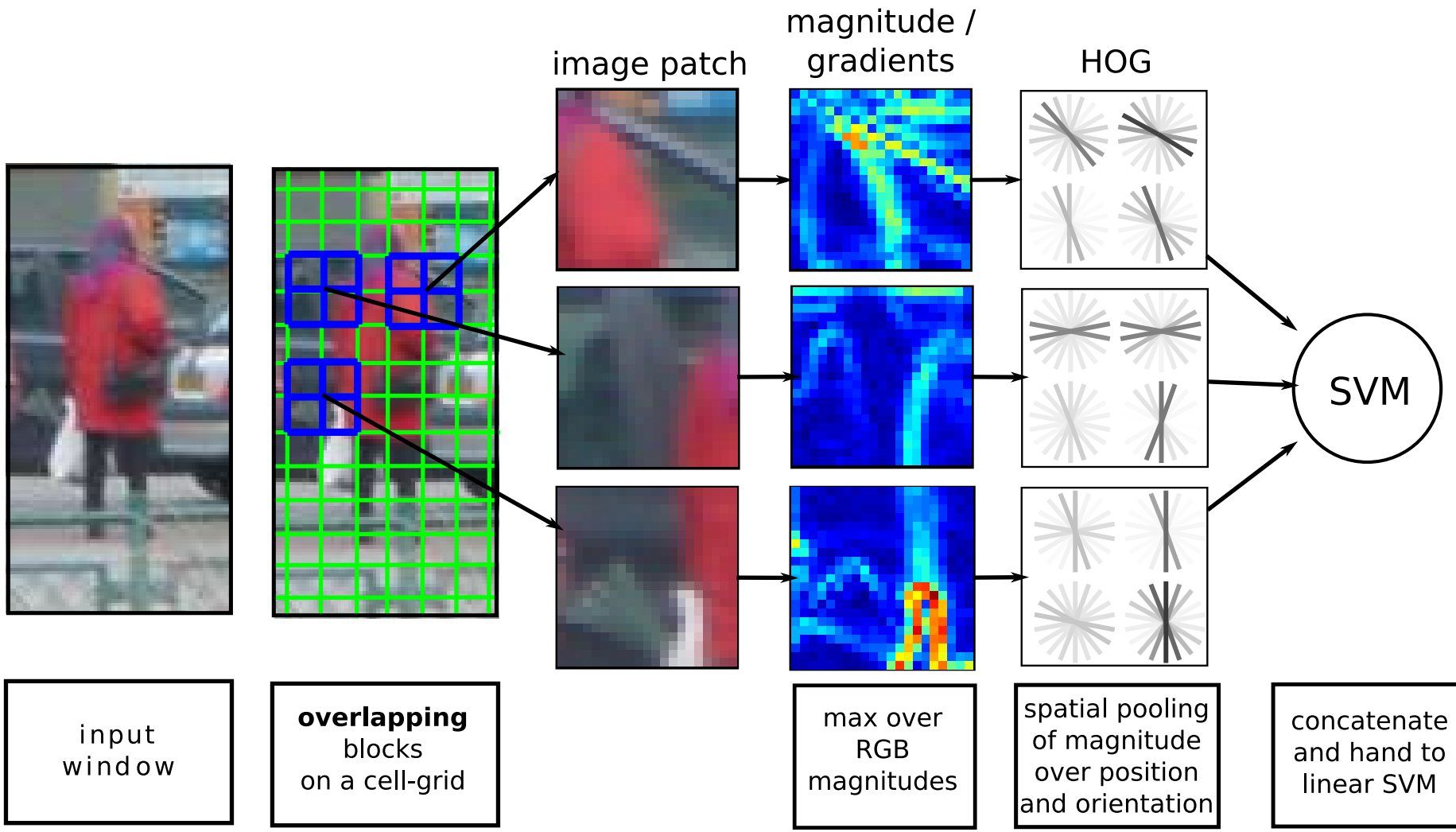
Overview

UNIVERSITY OF LEEDS

- Color Segmentation Features
  - From HOG to CHOG
  - Experiments
- Part Sharing for Deformable Part-based Models
  - Formulation of Shared Parts in a DPM
  - Experiments
- Outlook on future research directions

HOG, Dalal & Triggs 2005

UNIVERSITY OF LEEDS



HOG, Dalal & Triggs 2005

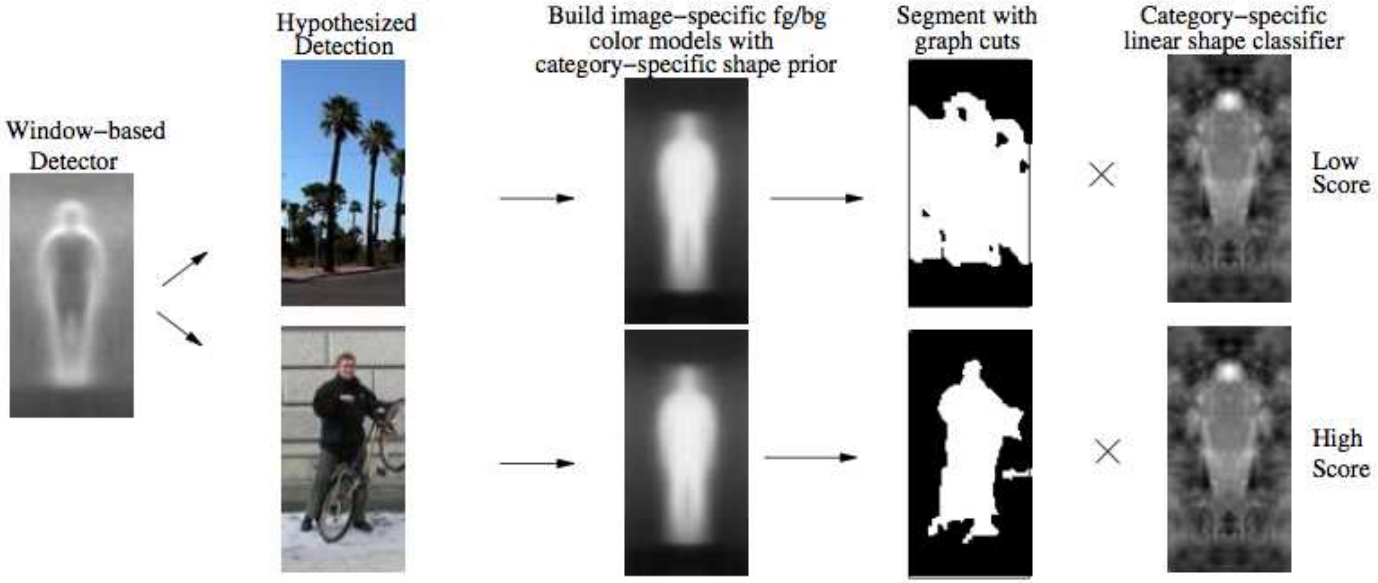
UNIVERSITY OF LEEDS

- Good
  - Local invariance due to spatial pooling
  - Invariance to brightness (gradients)
  - Invariance to contrast due to block normalization
- Not so good
  - Max-RGB gradients introduce background clutter
  - Object edges often not very strong
  - $\longrightarrow$ noisy HOG descriptors

Related Work

UNIVERSITY OF LEEDS

- Ramanan (CVPR 2007) used segmentation to verify detections



- Results indicate that segmentation can serve as a valuable cue to object detection
- Our goal: Incorporate segmentation into feature extraction stage

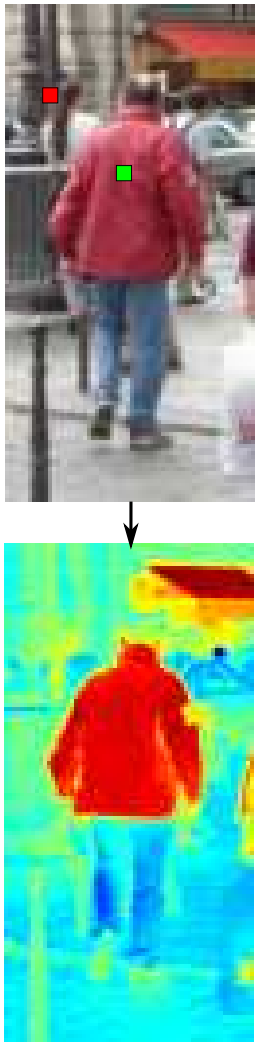
Obtaining soft segmentations

UNIVERSITY OF LEEDS

- Goal: Build a model for FG/BG, so we can ‘label’ pixels
- Local Gaussian assumption around $\mathbf{p}_{FG}$ (■) and $\mathbf{p}_{BG}$ (■) with means $\mathbf{m}_{FG}, \mathbf{m}_{BG}$
- Fast and simple projection:

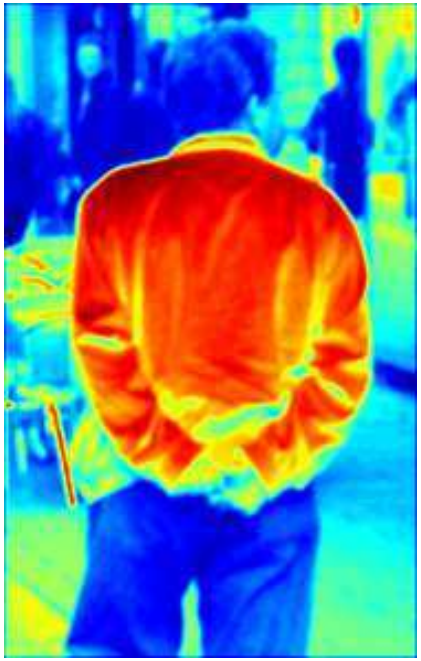
$$\hat{\mathbf{w}} \propto (\mathbf{m}_{FG} - \mathbf{m}_{BG})$$

  - ♦ Simplified form of Fisher discriminant



Motivation

UNIVERSITY OF LEEDS

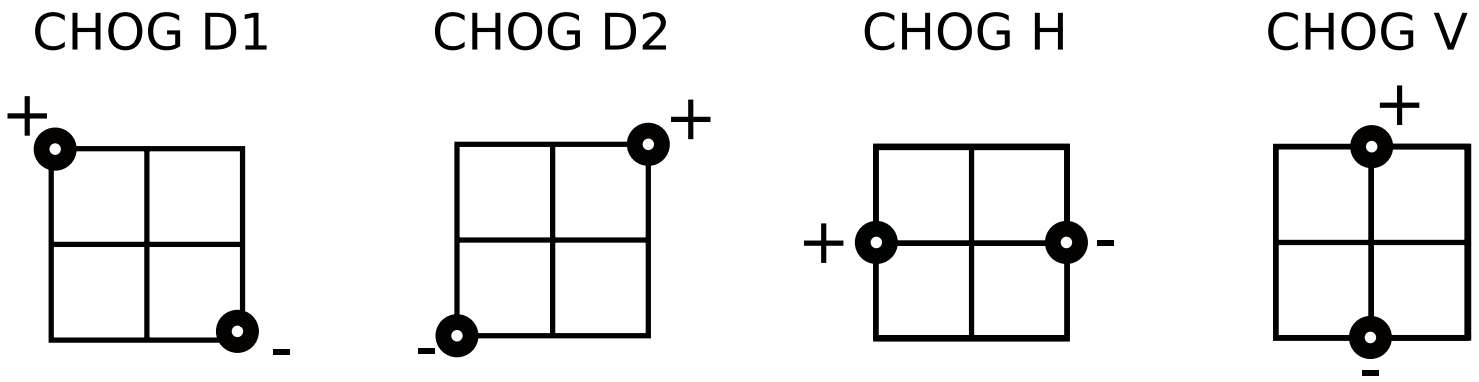


- Soft segmentation allows us to create an instance-specific cue about the object
- $\rightarrow$ ‘this jacket is white’ rather than ‘jackets are white’

Reference Points

UNIVERSITY OF LEEDS

- Our goal is to compute HOG features on the segmentation map S
- To compute $\hat{\mathbf{w}}$ locally we link the position of the reference points $\mathbf{p}_{FG}$ and $\mathbf{p}_{BG}$ to the position of the HOG block



Color Normalization

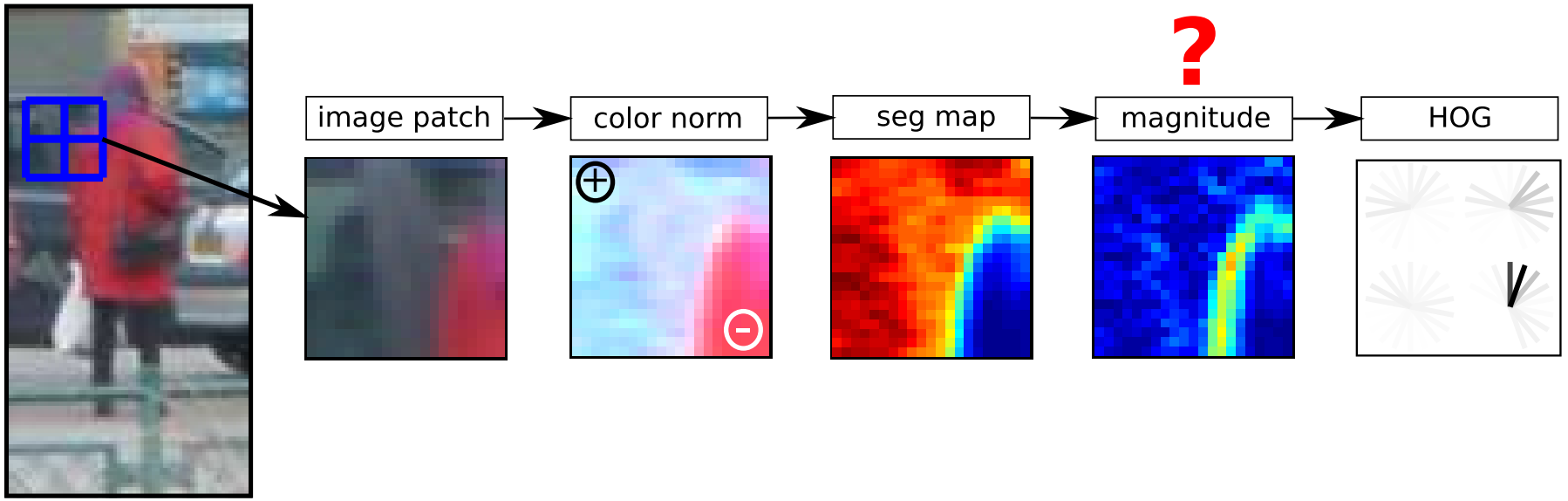
UNIVERSITY OF LEEDS

- Difference in means is dominated by intensity
  $\longrightarrow$ intensity invariant color normalization

$$\mathbf{x}' = \frac{\mathbf{x}}{\max(x_r, x_g, x_b) + \epsilon}$$

Gradient Computation

UNIVERSITY OF LEEDS

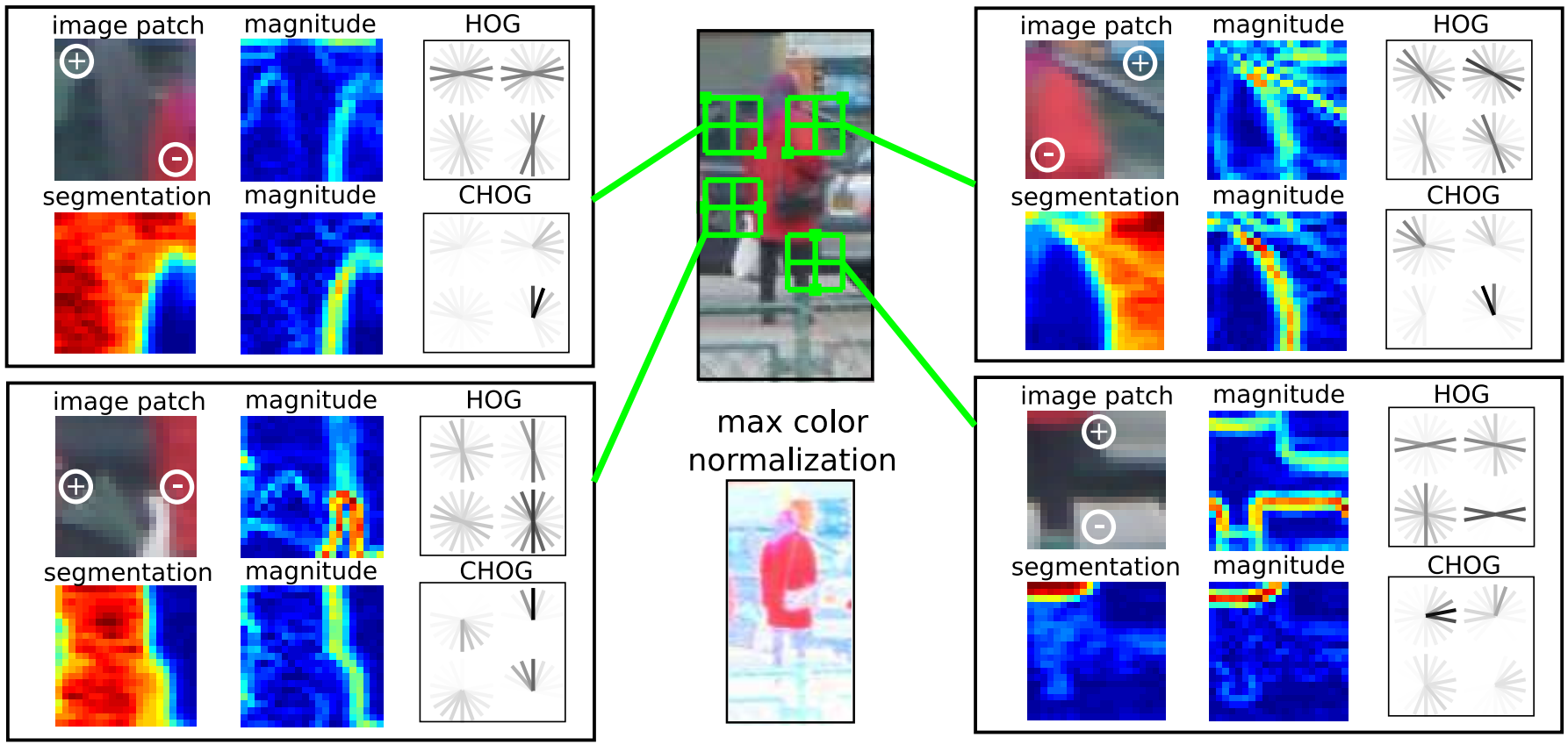


- Gradient of S as weighted sum of original image gradients

$$\frac{\partial}{\partial x}S = \hat{\mathbf{w}}^T \left\langle \frac{\partial}{\partial x}I_R, \frac{\partial}{\partial x}I_G, \frac{\partial}{\partial x}I_B \right\rangle^T$$

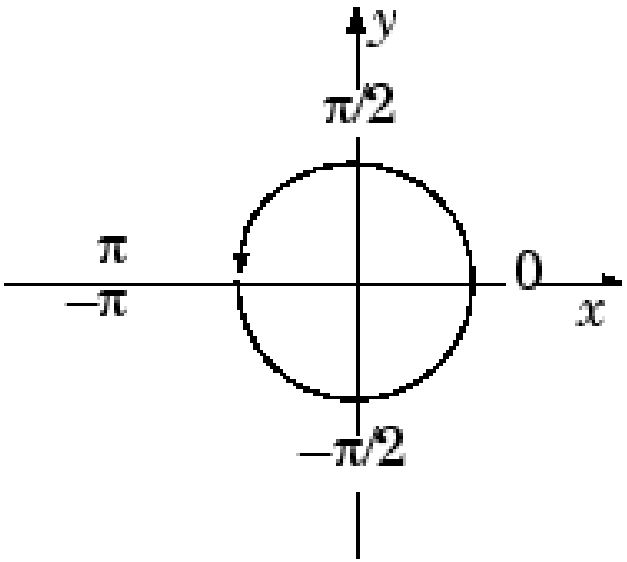
Discussion: CHOG

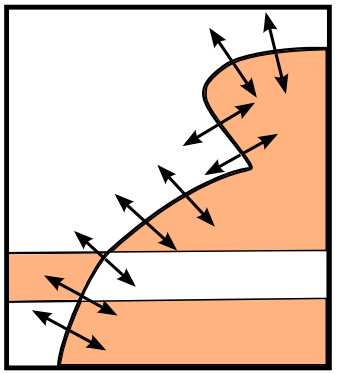
UNIVERSITY OF LEEDS

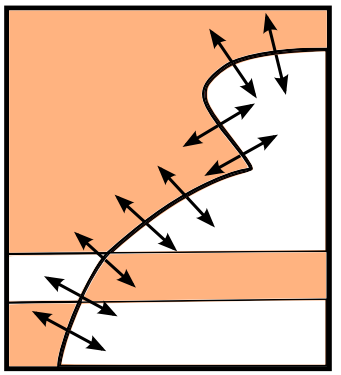


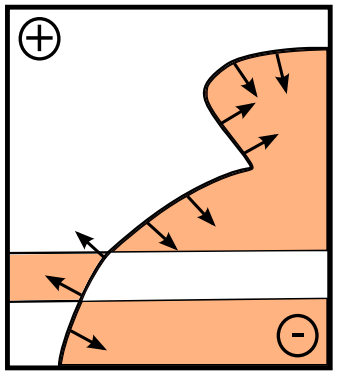
Discussion: Gradient Sign

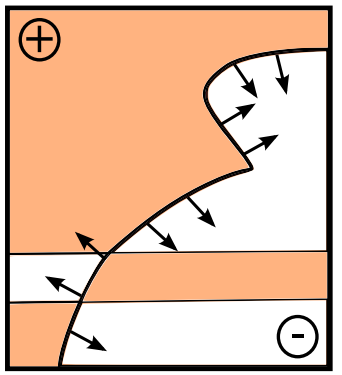
UNIVERSITY OF LEEDS











(a) HOG unsigned (b) CHOG (signed)

Classifier

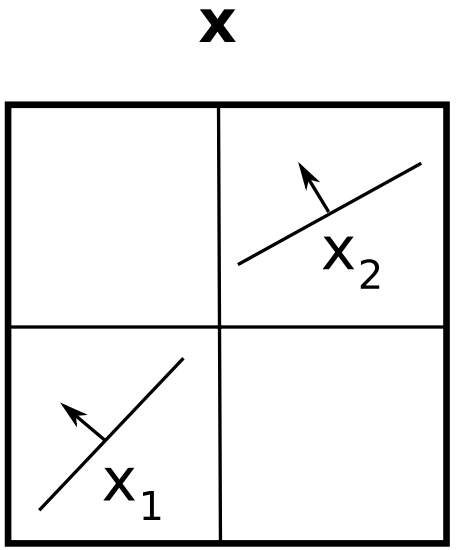
UNIVERSITY OF LEEDS

- Linear SVM & Kernel SVM with quadratic kernel

$$K(\mathbf{x}, \mathbf{y}) = \left(\mathbf{x}^T\mathbf{y} + 1\right)^2 = \phi(\mathbf{x}) \cdot \phi(\mathbf{y})$$

- The kernel is capable of modeling dependencies between any two features in the feature vector, i.e.

$$\phi(\mathbf{x}) = \left\langle \mathbf{x}_1^2, \mathbf{x}_2^2, \mathbf{x}_1\mathbf{x}_2, \ldots \right.$$
$$\left. \sqrt{2}\mathbf{x}_1, \sqrt{2}\mathbf{x}_2, 1 \right\rangle$$



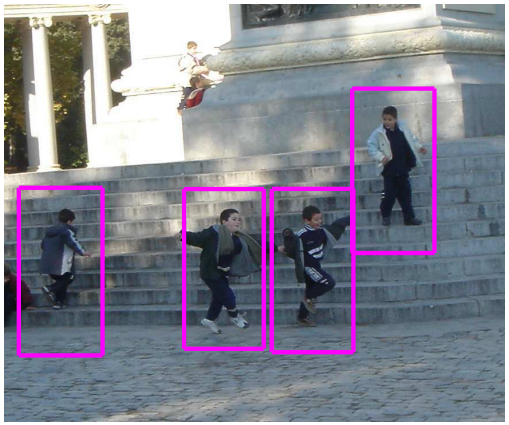
Experiments

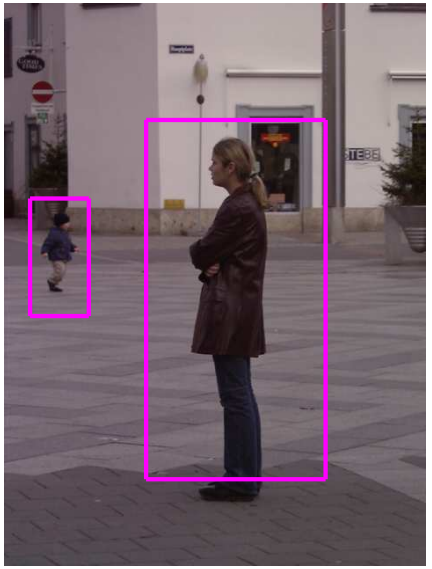
UNIVERSITY OF LEEDS

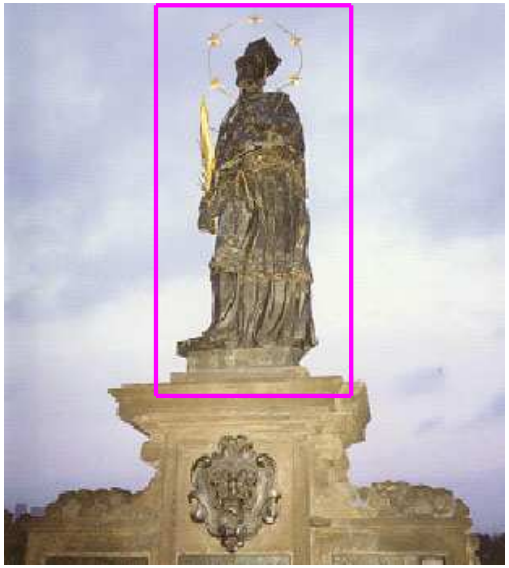
- INRIAPerson dataset $\rightarrow$ followed Dalal & Triggs evaluation protocol
  - Detection-Error-Tradeoff (DET) curves: FPPW vs miss-rate
- PASCAL VOC 2006 dataset $\rightarrow$ followed evaluation protocol
  - Precision-Recall curves

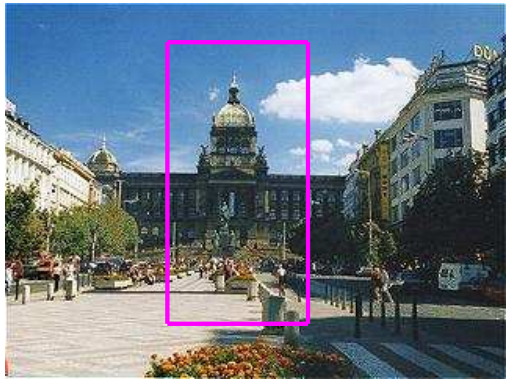
Detection Examples

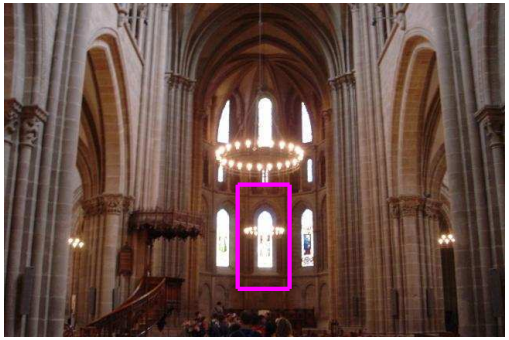
UNIVERSITY OF LEEDS





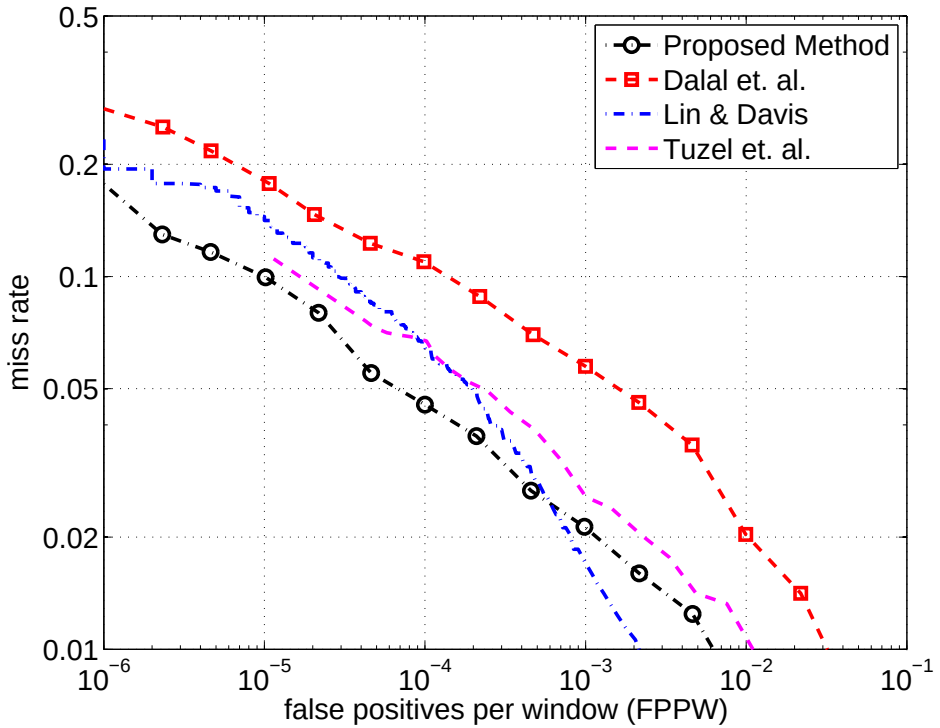






Quantitative Results

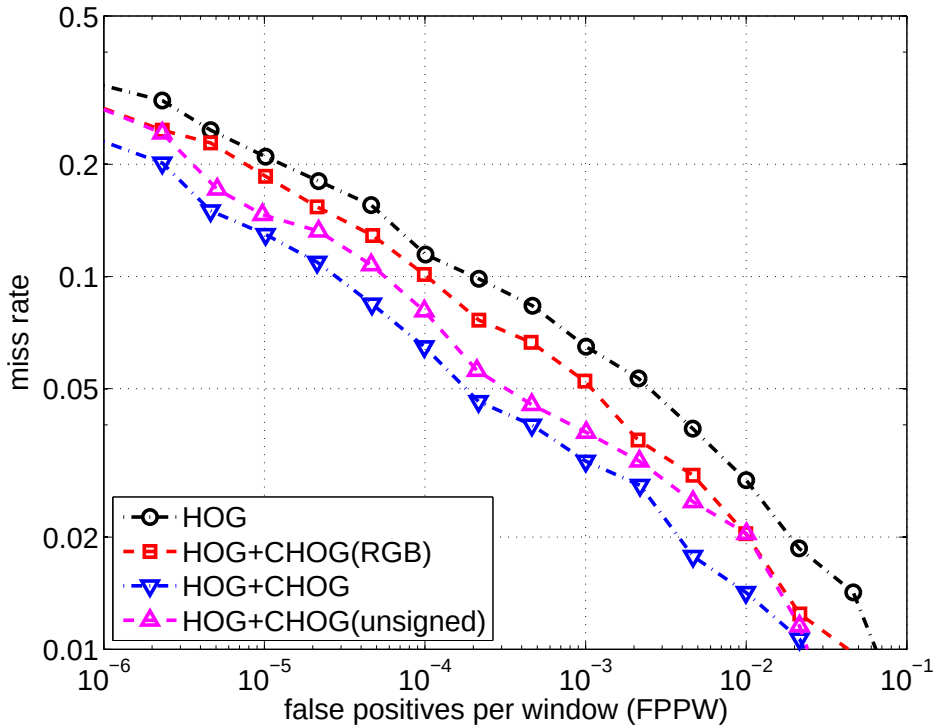
UNIVERSITY OF LEEDS



- For FPPW rate $\leqslant 10^{-4}$ we outperform all other methods
- $\sim 60\%$ improvement compared to Dalal&Triggs at 10^{-4} FPPW.

Descriptor Scheme

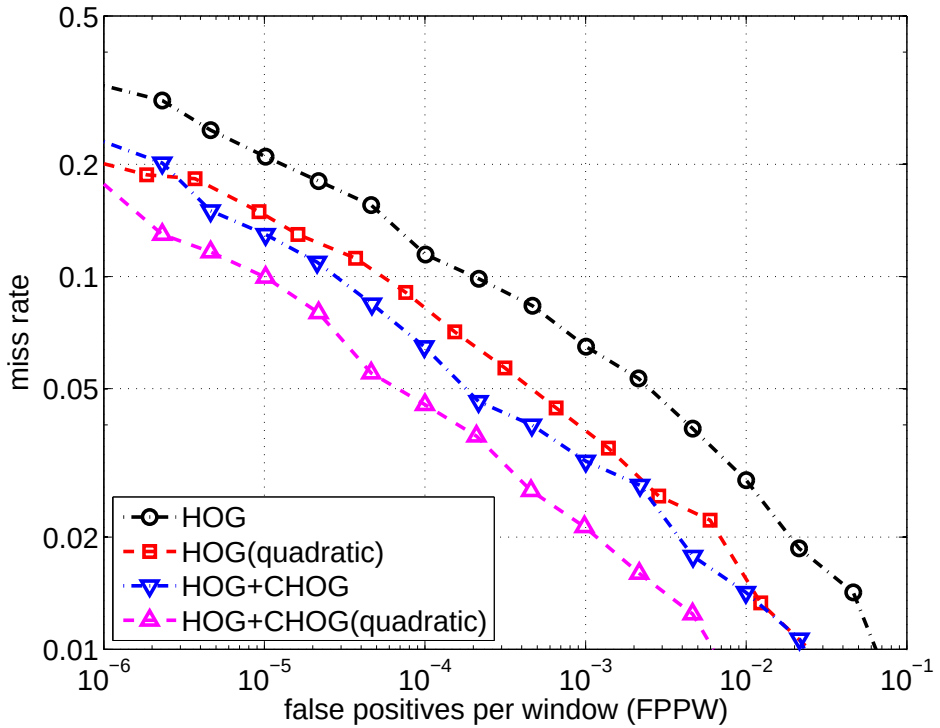
UNIVERSITY OF LEEDS



- 43% relative improvement at 10^{-4} FPPW when using HOG+CHOG
- Signed gradients for CHOG account for 30% of this improvement

Choice of Kernel

UNIVERSITY OF LEEDS

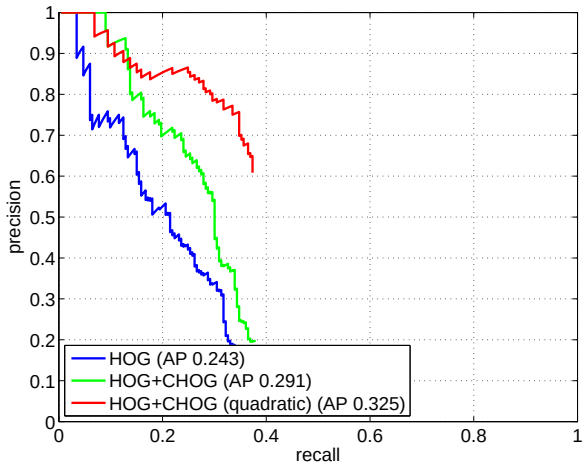


- 31% relative improvement for both, HOG and HOG+CHOG, with quadratic kernel

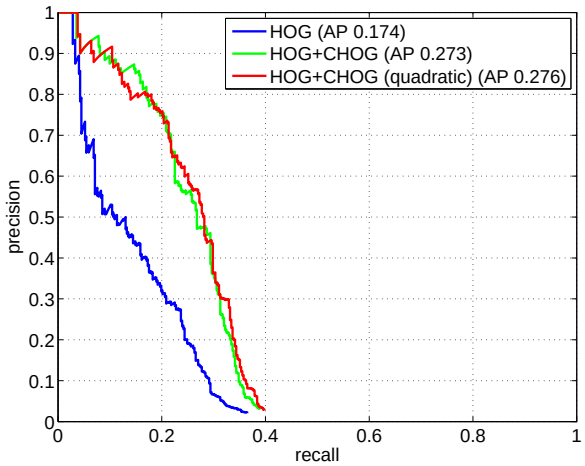
Experiments

UNIVERSITY OF LEEDS

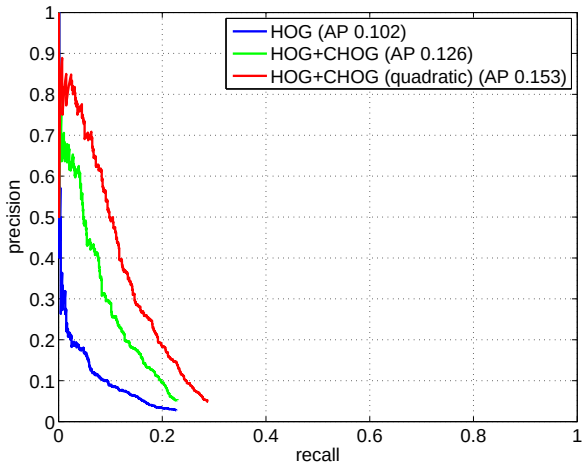
- PASCAL VOC 2006



(a) Bus



(b) Sheep



(c) Person

- Improvements in Average Precision up to 50% on various classes.

Summary of CHOG features

UNIVERSITY OF LEEDS

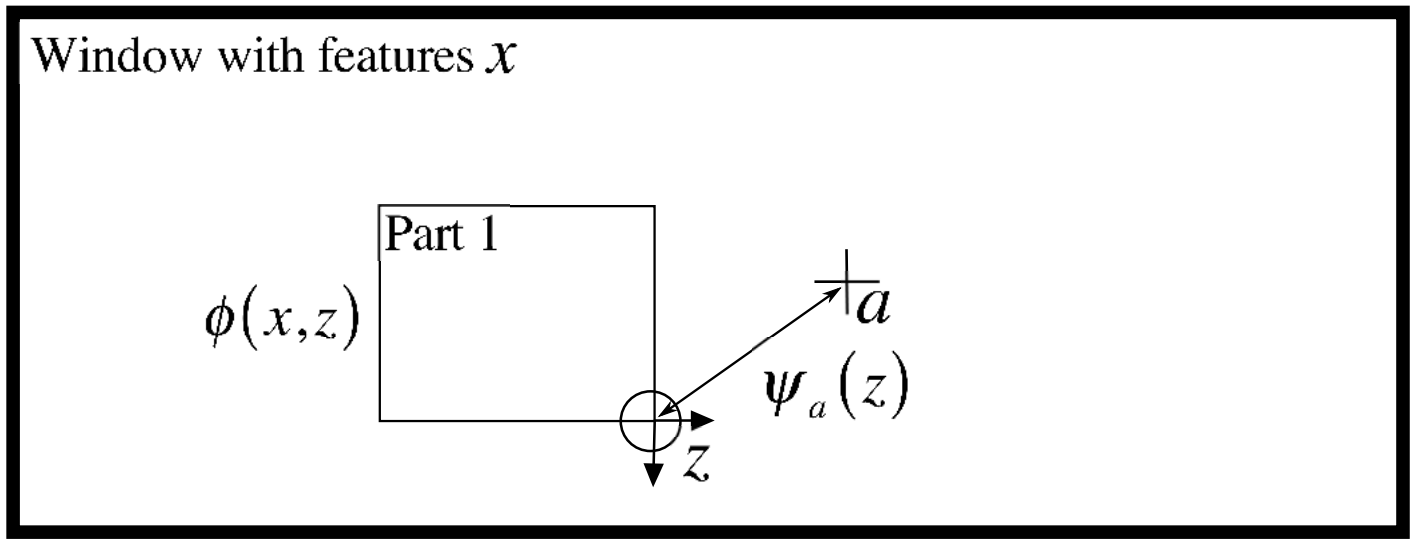
- Incorporated segmentation cues directly in the feature extraction process
- CHOG is an extension of HOG to segmentation and color information
- $\rightarrow$ captures stronger and coherent edges
- Substantial performance improvement over HOG for pedestrian and object detection
- State-of-the-art results on INRIAPerson

Deformable Part-based Models

UNIVERSITY OF LEEDS

- Standard SVM formulation might not offer the flexibility required to solve the general object detection task (PASCAL VOC)
- DPMs (Felzenszwalb et. al., PAMI 2010) offer a higher degree of *learnt* invariance by
  - partitioning the object into a set of local and agile parts, *allowing the detector to adapt to the underlying image structure*
  - using a mixture model to *capture large differences in appearance of an object class*

Parts



- $\phi\left(\boldsymbol{x}, \boldsymbol{z}\right)$ extracts features from window given part position $z = \langle z_x, z_y \rangle$
- $\psi_{\boldsymbol{a}}\left(\boldsymbol{z}\right)$ is the displacement of a part from its anchor $\boldsymbol{a} = \langle a_x, a_y \rangle$

$$\psi_{\boldsymbol{a}}\left(\boldsymbol{z}\right) = \left[\left(a_x - z_x\right)^2, \left(a_x - z_x\right), \left(a_y - z_y\right)^2, \left(a_y - z_y\right)\right]$$

Modular Definition

UNIVERSITY OF LEEDS

- Part Response via inference of part position $\boldsymbol{z}$ on a window with features $\boldsymbol{x}$:

$$\vartheta\left(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{v}, \boldsymbol{a}\right) = \max_{\boldsymbol{z} \in \mathcal{Q}} \left\{ \boldsymbol{w}^{\mathsf{T}} \phi\left(\boldsymbol{x}, \boldsymbol{z}\right) - \boldsymbol{v}^{\mathsf{T}} \psi_{\boldsymbol{a}}\left(\boldsymbol{z}\right) \right\}$$

- Part appearance $\boldsymbol{w}^{\mathsf{T}} \phi\left(\boldsymbol{x}, \boldsymbol{z}\right)$ with part filter $\boldsymbol{w}$ and features $\phi\left(\boldsymbol{x}, \boldsymbol{z}\right)$ given part position $\boldsymbol{z}$
- Spatial configuration $\boldsymbol{v}^{\mathsf{T}} \psi_{\boldsymbol{a}}\left(\boldsymbol{z}\right)$ with spatial priors $\boldsymbol{v}$

Modular Definition II

UNIVERSITY OF LEEDS

- Response to the lth mixture component:

$$g^l(\boldsymbol{x}) = b_l + \sum_{i=1}^{p} \vartheta\left(\boldsymbol{x}, \boldsymbol{w}^{l,i}, \boldsymbol{v}^{l,i}, \boldsymbol{a}^{l,i}\right)$$

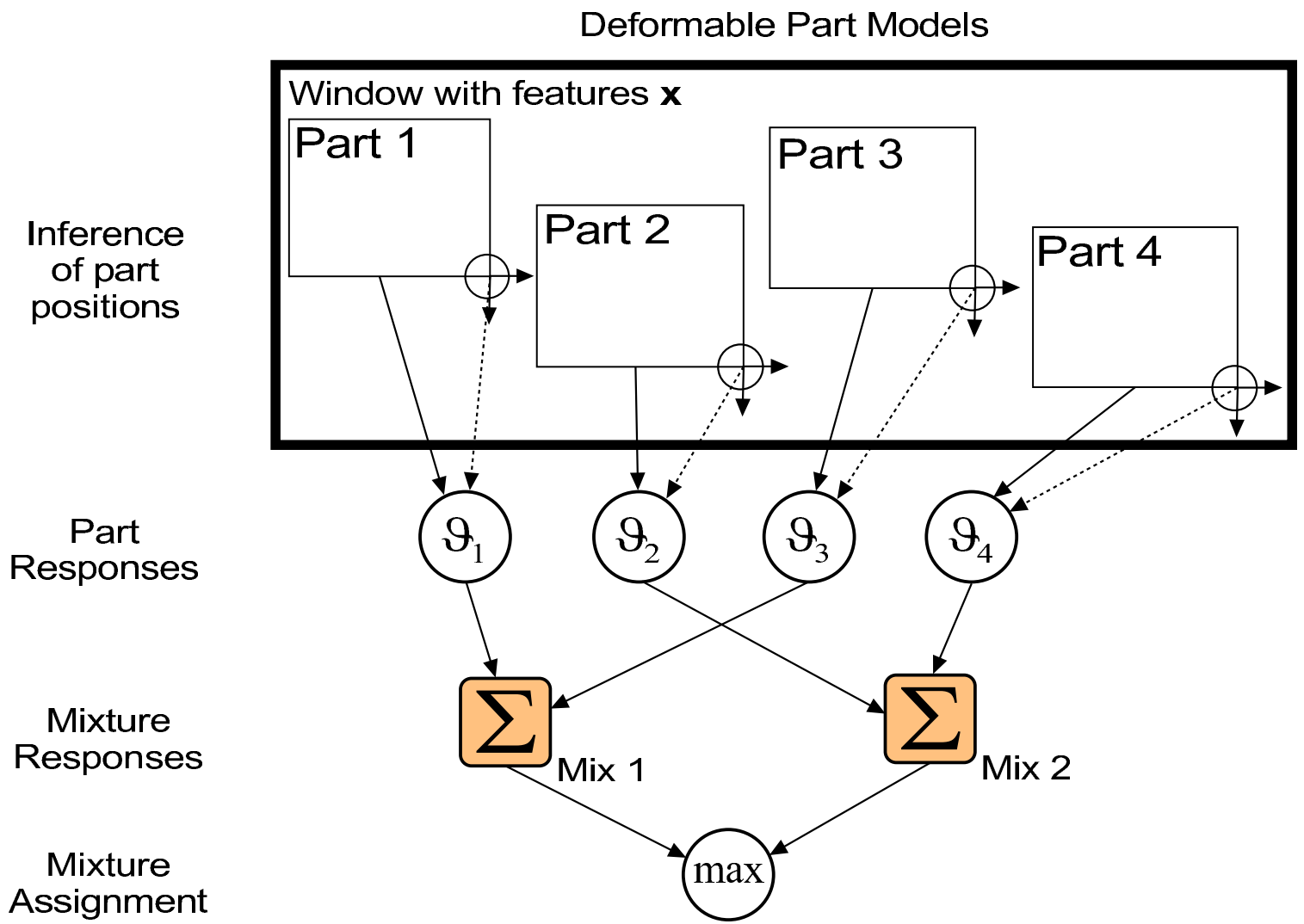
- Detector response by obtaining max over d experts

$$h(\boldsymbol{x}) = \max\left\{g^1(\boldsymbol{x}), g^2(\boldsymbol{x}), ..., g^d(\boldsymbol{x})\right\}$$

- Positive training examples for each expert are (usually) mutually exclusive

Overview DPMs

UNIVERSITY OF LEEDS



Problems

UNIVERSITY OF LEEDS

- Increasing the number of mixtures does not necessarily lead to better performing detectors
  - Parts are bijectively linked to mixture components $\rightarrow$ *increasing mixtures results in linear growth of number of parameters*
  - Positive training examples are assigned to a single mixture component $\rightarrow$ *linear decrease in available training data per mixture*
- $\Rightarrow$ Bad generalization, high computational requirements

Solutions

UNIVERSITY OF LEEDS

- Sharing part responses across multiple mixture components has various positive effects
  - Parts can be reused $\rightarrow$ number of parameters is reduced
  - Training examples are shared across all relevant parameters $\rightarrow$ paucity of available training data will have less negative impact
  - Computational expense is reduced $\rightarrow$ potential scaling to large numbers of mixtures/classes.

Part Sharing

UNIVERSITY OF LEEDS

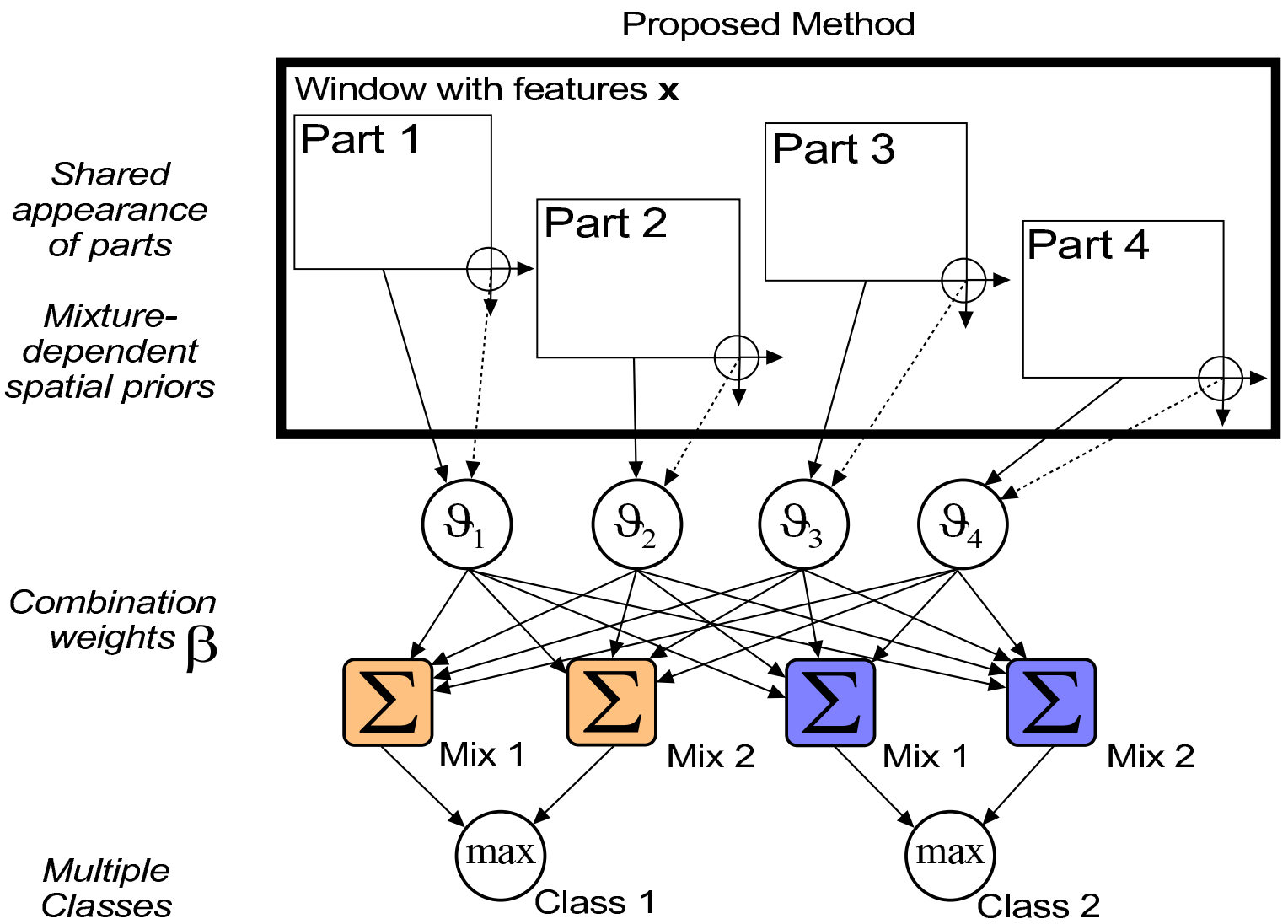
- Response to the lth mixture component:

$$g^l(\boldsymbol{x}) = b_l + \sum_{i=1}^{p} \beta_i^l \vartheta\left(\boldsymbol{x}, \boldsymbol{w}^i, \boldsymbol{v}^{l,i}, \boldsymbol{a}^{l,i}\right)$$

- Weights $\boldsymbol{\beta}$ give importance to individual part responses per mixture
- Note: Shared part appearance but independent spatial configuration!

DPMs with Part Sharing

UNIVERSITY OF LEEDS



Learning

UNIVERSITY OF LEEDS

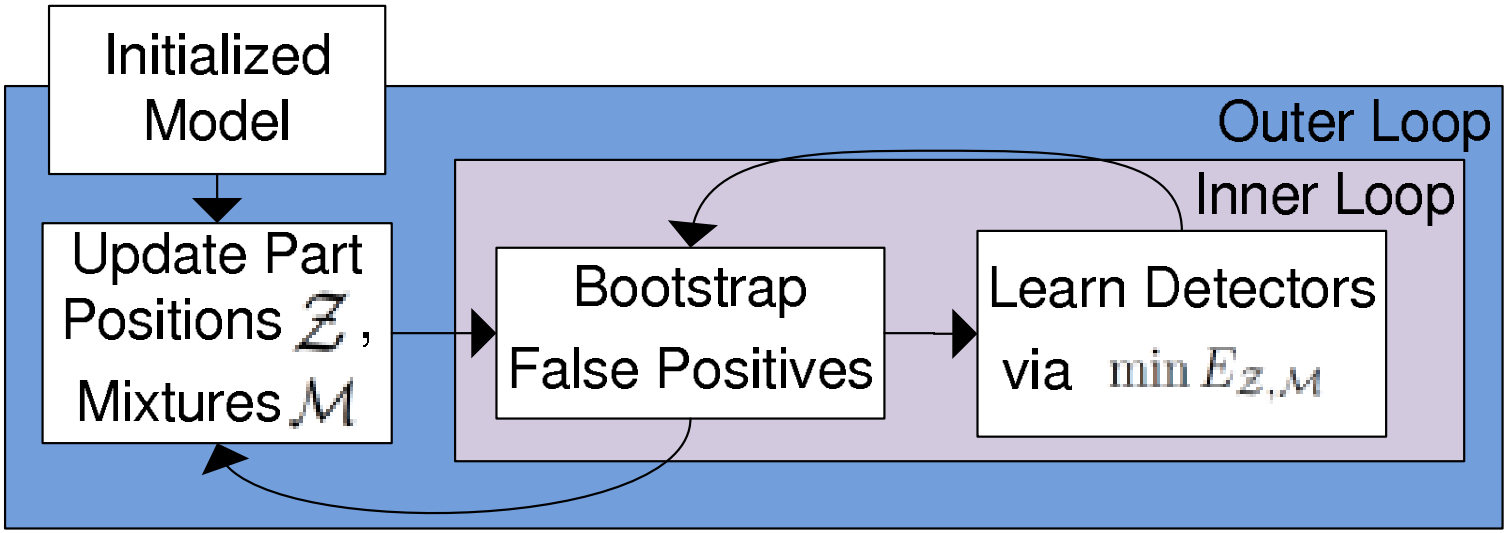
- Given a set of part filters $\mathcal{W}$, spatial priors $\mathcal{V}$ and combination weights $\boldsymbol{\beta}$
- Learning via Energy Minimization:

$$E\left(\mathcal{W},\mathcal{V},\boldsymbol{\beta}\right)=\lambda R\left(\mathcal{W},\boldsymbol{\beta}\right)+\sum_{k=1}^{n}L\left(y^{k},h\left(\boldsymbol{x}^{k}\right)\right)$$

- L is the hinge loss, R is a ℓ_2 regularization term on the part responses

Optimization

UNIVERSITY OF LEEDS

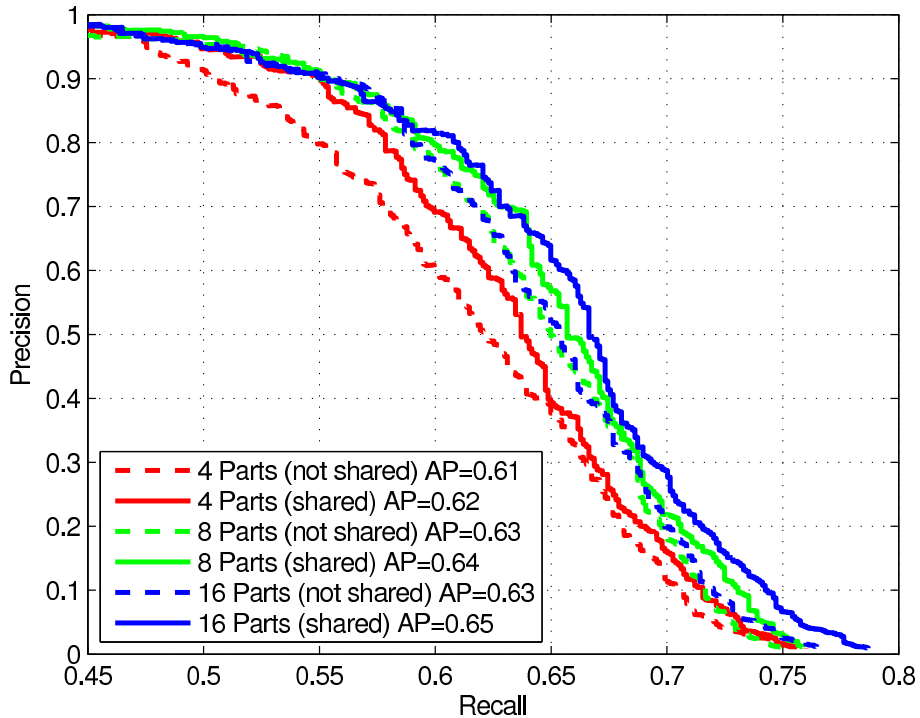


- In the Inner Loop, repeat:
  - Alternate until convergence:
    - Optimize wrt part filters $\mathcal{W}$, spatial priors $\mathcal{V}$
    - Optimize wrt combination weights β
  - Bootstrap False Positives

Quantitative results

UNIVERSITY OF LEEDS

- VOC 2006 car, 2 mixtures, parts vary

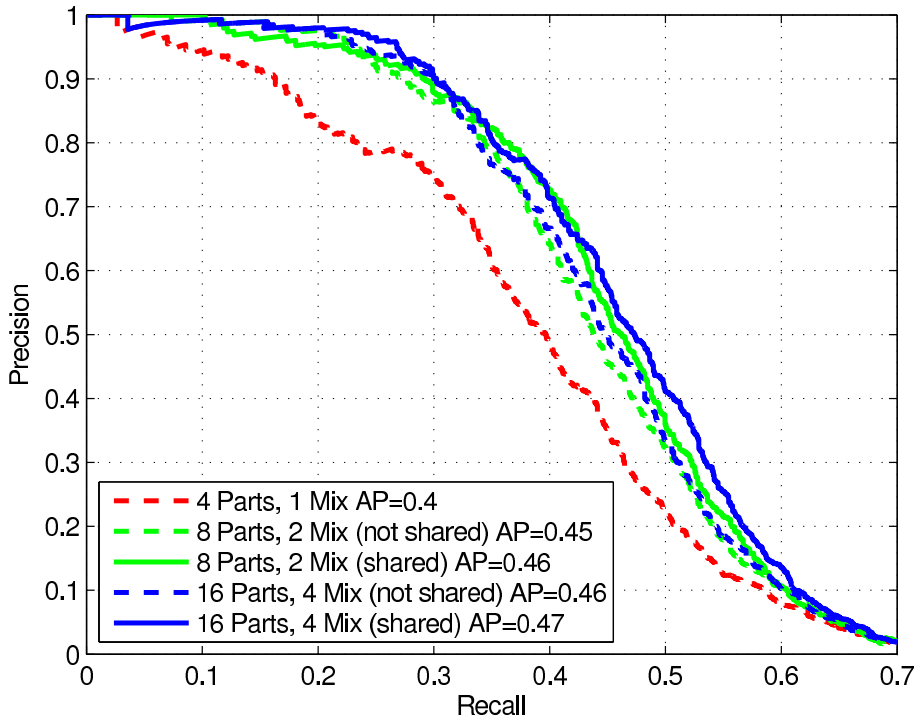


- Sharing parts is always beneficial
- 8 shared parts outperform 16 non-shared parts $\longrightarrow$ doing more with less

Quantitative results

UNIVERSITY OF LEEDS

- VOC 2007 car, mixtures vary, parts vary

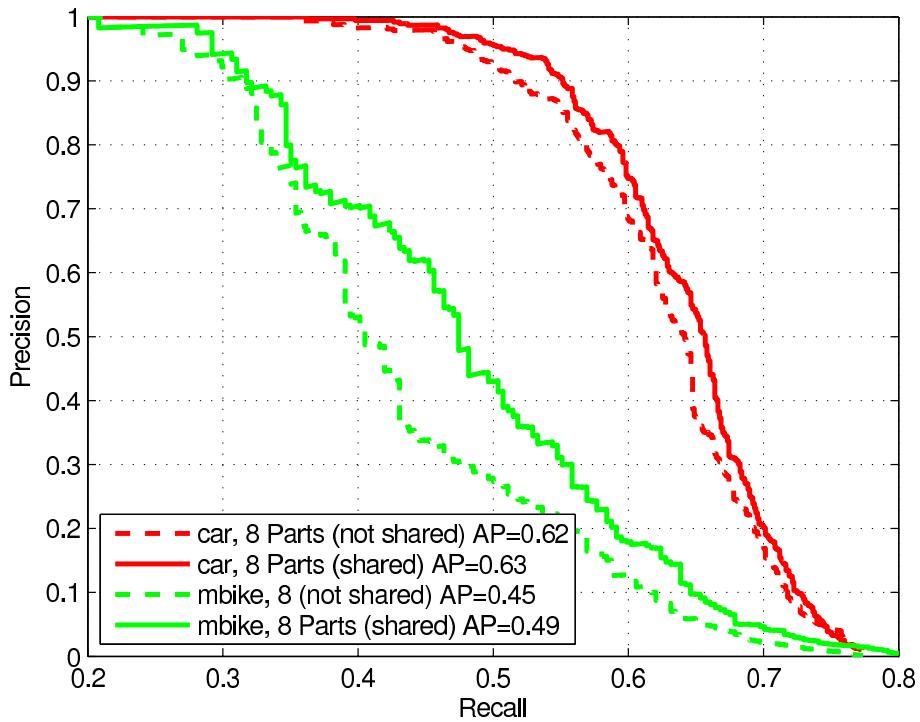


- Sharing parts is always beneficial
- Increasing number of mixtures gives small improvement in AP

Quantitative results

UNIVERSITY OF LEEDS

- VOC 2006 car+motorbike, 2 mixtures each



- Sharing parts is beneficial for both classes

Summary of DPMs with Part Sharing

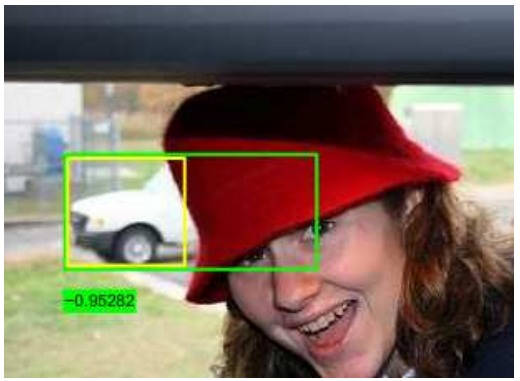
UNIVERSITY OF LEEDS

- Part sharing allows for a more efficient usage of parameters and training data
- We can do “more with less”
- Extension of the DPM learning scheme to multi-class learning
- A step towards a holistic object detection framework

Outlook on future research

UNIVERSITY OF LEEDS

- A lack of sufficient techniques to model occlusion is apparent
- Ideally occlusion modeling will provide
  - Higher detection accuracy
  - Superior scene interpretation (visibility)



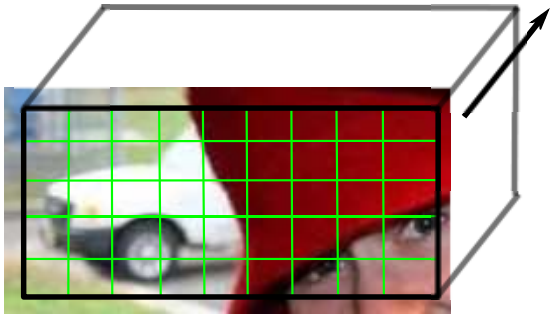
Binary Flags

UNIVERSITY OF LEEDS

- Discard parts $\rightarrow$ sliding window + SVM
- Binary flag per feature-block indicates whether the image region is occluded or not

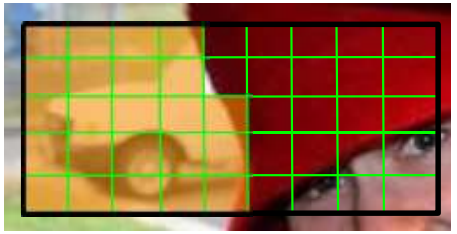
$$f\left(\boldsymbol{x}, \boldsymbol{o}\right) = \sum_{i=1}^{q} \left[\!\left[o_i = 0 \right]\!\right] \boldsymbol{w}^i \cdot \boldsymbol{x}^i + \left[\!\left[o_i \neq 0 \right]\!\right] u_i$$

$\boldsymbol{x}$



n features
per block
(HOG, CHOG, LBP, ...)

$\boldsymbol{o}$



binary occlusion
features
per block

Binary Flags

UNIVERSITY OF LEEDS

- How to infer occlusion $\boldsymbol{o}$?
- Good start:

$$g(\boldsymbol{x}) = \max_{\boldsymbol{o}} f(\boldsymbol{x}, \boldsymbol{o})$$

- Updates on $\boldsymbol{o}$ can be performed in the outer loop of the DPM learning scheme
- Next step: Move inference of $\boldsymbol{o}$ into the non-maximum-suppression scheme to infer a scene interpretation

Thank you

UNIVERSITY OF LEEDS

Any questions?