Combining Appearance and Structure from Motion Features for Road Scene Understanding

Paul Sturgess, Karteek Alahari, Ľubor Ladický, Phil Torr
Oxford Brookes University

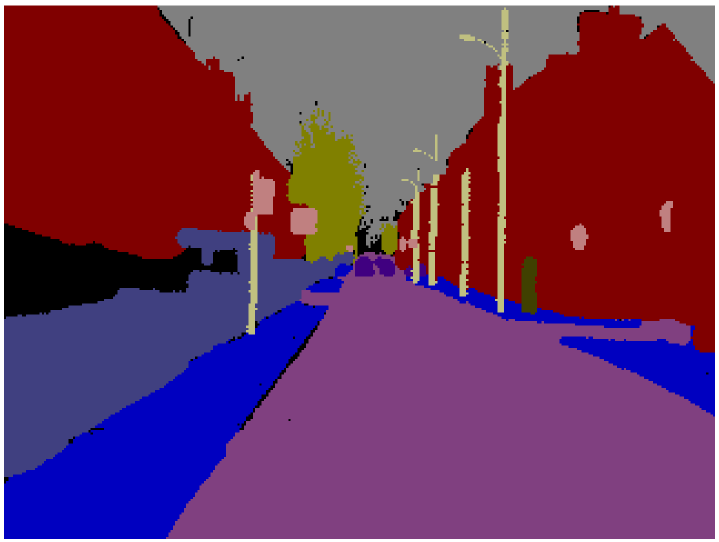
Oxford Brookes Vision Group

http://cms.brookes.ac.uk/research/visiongroup/

Goal: Classify ↔ Segment

Oxford Brookes Vision Group

- Abundance of street level imagery
- Classify every pixel in an image



Road	Building	Sky	Tree	Sidewalk	Car
Void	Column	Sign	Fence	Pedestrian	Cyclist

The Cambridge-driving Labeled Video Database

http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/

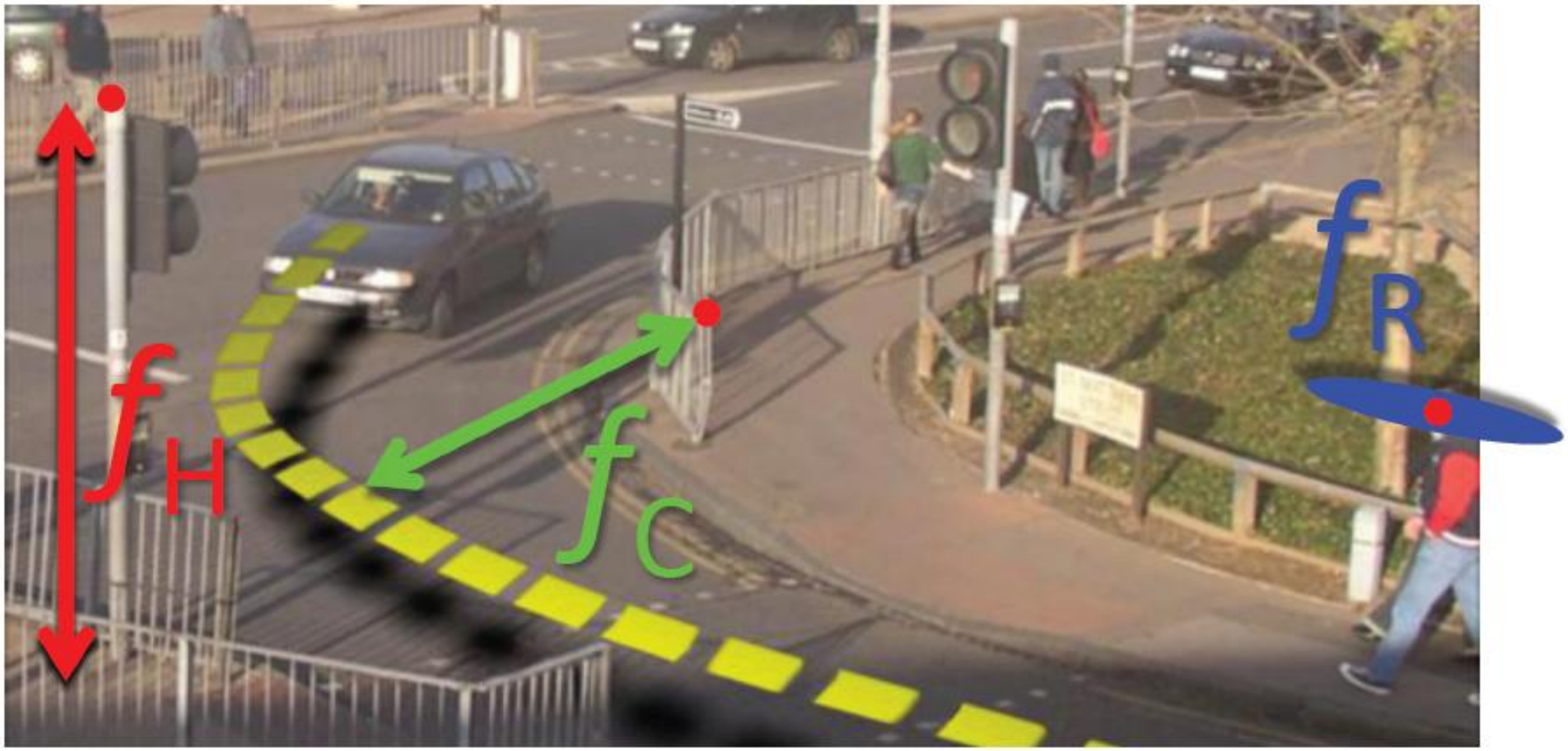
G. J. Brostow, J. Fauqueur, and R. Cipolla. Semantic object classes in video: A highdefinition ground truth database. Pattern Recognition Letters 2009.

Method

- A complementary set of features
  - Can describe a wide variety of object-classes
- Higher Order CRF
  - Produces high quality object-class boundaries
- Joint Boost for Unary Potentials
  - Single classifier for all features
- Evaluation
  - High quality annotated ground truth

Features

- Structure-from-motion
  - Moving Vs Static, 3D location cues, Texture



G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla.
Segmentation and recognition using structure from motion point clouds. ECCV 2008.

Features

Oxford Brookes Vision Group

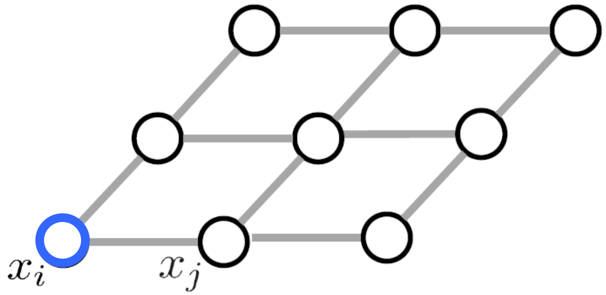
- HOG
- Colour
- Location
- Textons

Higher Order CRF

Oxford Brookes Vision Group

$$E(\mathbf{x}) = \underbrace{\sum_{i \in \mathcal{V}} \psi_i(x_i)}_{\text{Unary Potential}} + \sum_{(i,j) \in \mathcal{E}} \psi_{ij}(x_i, x_j) + \sum_{c \in \mathcal{S}} \psi_c(\mathbf{x}_c)$$

- Likelihood of a pixel taking a label
- Computed via a boosting approach



Boosting for Unary Potentials

Oxford Brookes Vision Group

- *TextonBoost*
  - Context exploited
  - Boosted combination of textons
  - Response defined by the pair

    [texton t, rectangular region r].

J. Shotton, J. M. Winn, C. Rother, and A. Criminisi.
TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. ECCV 2006.

Boosting for Unary Potentials

Oxford Brookes Vision Group

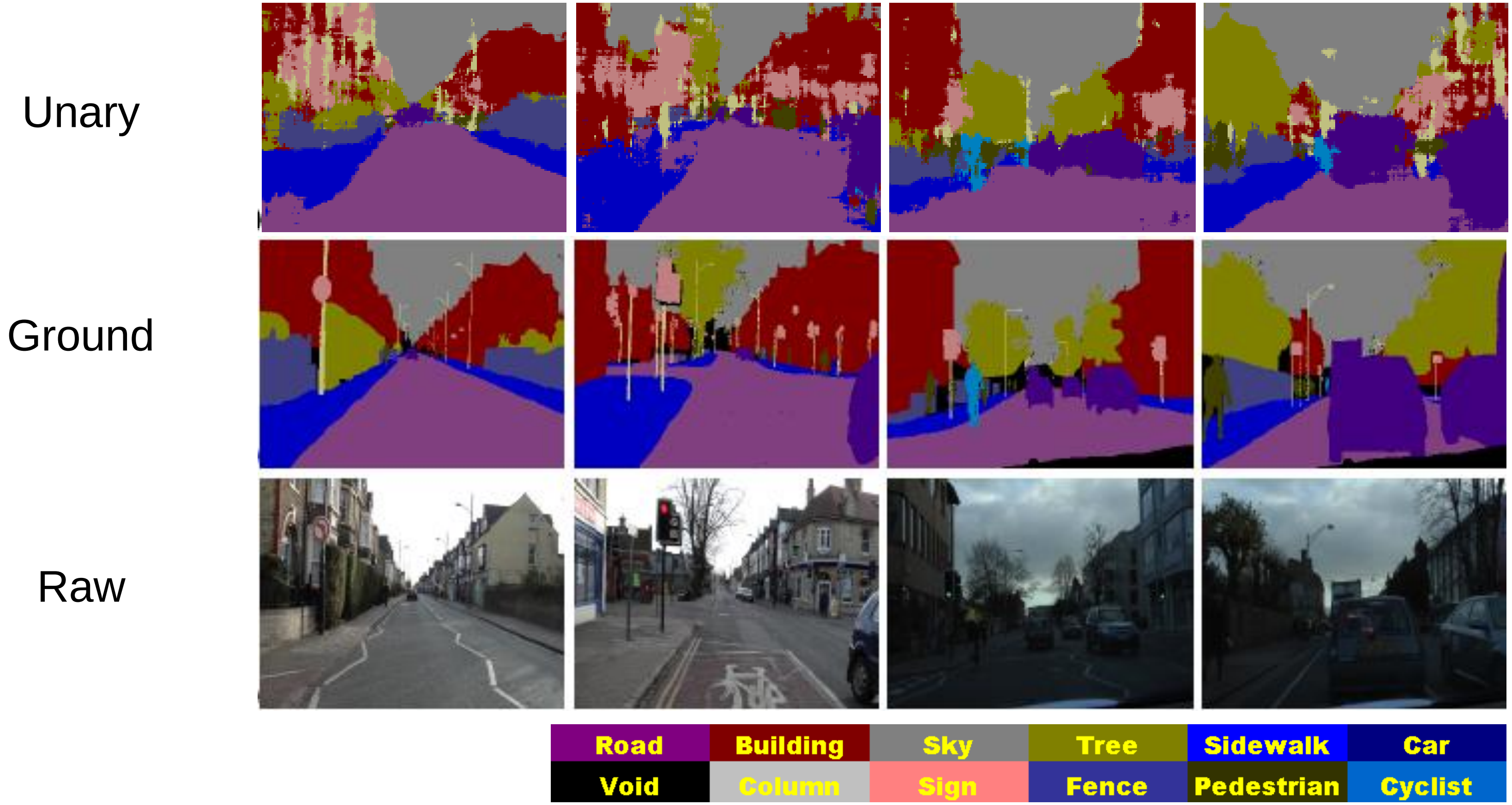
- Dense Boost
  - Response defined by the triplet

    [feature type f, feature cluster t, rectangular region r]

    f = {SfM, HOG, Colour, Location, Texton}

L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr.
Associative hierarchical crfs for object class image segmentation. ICCV 2009.

Unary Potential Result



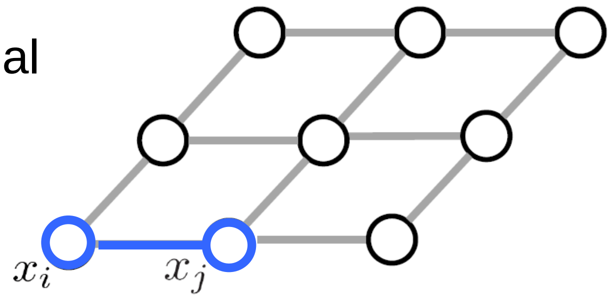
	Building	Tree	Sky	Car	Sign	Road	Pedestrian	Fence	Column	Sidewalk	Bicyclist	Average	Global
Brostow	46.2	61.9	89.7	68.6	42.9	89.5	53.6	46.6	0.7	60.5	22.5	53	69.1
Unary	**61.9**	**67.3**	**91.1**	**71.1**	58.5	**92.9**	49.5	37.6	**25.8**	**77.8**	**24.7**	**59.8**	**76.4**

Columns = Per-class recall, Average = Average recall, Global = Overall correctly labelled pixels

Higher Order CRF

Oxford Brookes Vision Group

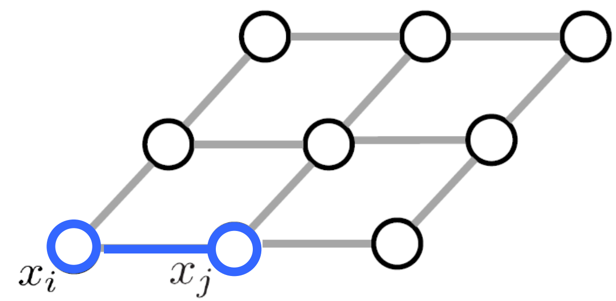
$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \underbrace{\sum_{(i,j) \in \mathcal{E}} \psi_{ij}(x_i, x_j)}_{\text{Pairwise Potential}} + \sum_{c \in \mathcal{S}} \psi_c(\mathbf{x}_c)$$



- Contrast sensitive Potts model
- Encourages label consistency in adjacent pixels

Higher Order CRF

$$E(\mathbf{x}) = \underbrace{\sum_{i\in\mathcal{V}} \psi_i(x_i) + \sum_{(i,j)\in\mathcal{E}} \psi_{ij}(x_i, x_j)}_{\textit{TextonBoost}} + \sum_{c\in\mathcal{S}} \psi_c(\mathbf{x}_c)$$



- Contrast sensitive Potts model
- Encourages label consistency in adjacent pixels

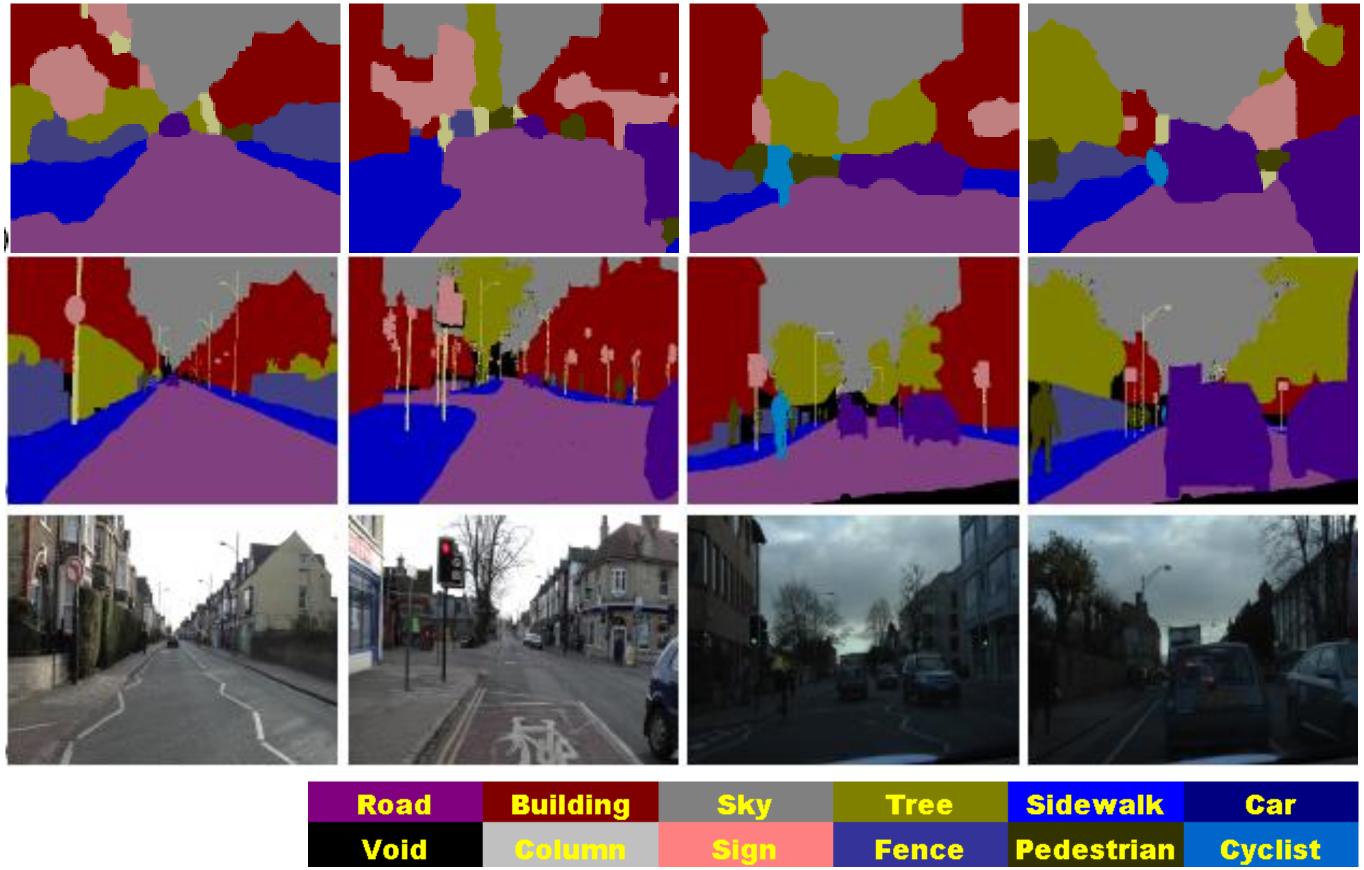
Pairwise Potential Result

Oxford Brookes Vision Group

Unary + Pairwise

Ground

Raw



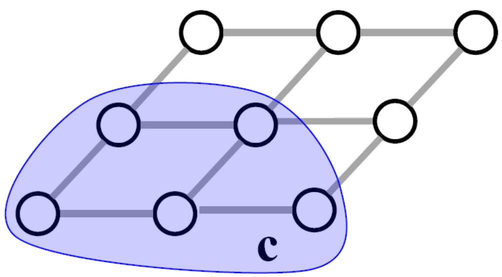
	Building	Tree	Sky	Car	Sign	Road	Pedestrian	Fence	Column	Sidewalk	Bicyclist	Average	Global
Brostow	46.2	61.9	89.7	68.6	42.9	89.5	**53.6**	**46.6**	0.7	60.5	22.5	53	69.1
Unary	61.9	67.3	91.1	71.1	**58.5**	92.9	49.5	37.6	**25.8**	77.8	**24.7**	59.8	76.4
+Pairwise	**70.7**	**70.8**	**94.7**	**74.4**	55.9	**94.1**	45.7	37.2	13	**79.3**	23.1	**59.9**	**79.8**

Columns = Per-class recall, Average = Average recall, Global = Overall correctly labelled pixels

Higher Order CRF

Oxford Brookes Vision Group

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \psi_{ij}(x_i, x_j) + \underbrace{\sum_{c \in \mathcal{S}} \psi_c(\mathbf{x}_c)}_{\text{Higher Order Potential}}$$

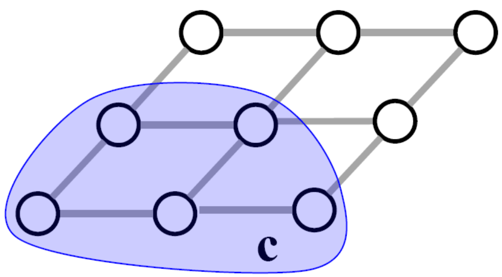


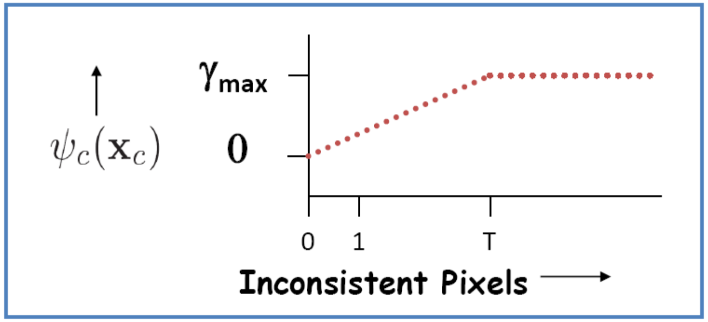
- Potential takes the form of a robust P^N model
- Encourages label consistency within a super-pixel
- Super-pixels computed using meanshift

Pushmeet Kohli, Lubor Ladicky, Philip H.S. Torr.
Robust Higher Order Potentials for Enforcing Label Consistency. IJCV 2009.

Robust P^N model

$$\psi_c(\mathbf{x}_c) = \begin{cases} \underbrace{\overbrace{N_i(\mathbf{x}_c)}^{\text{Number of inconsistent pixels}}\frac{1}{Q}\gamma_{\max}} & \text{if } N_i(\mathbf{x}_c) \leq \overbrace{Q}^{\text{Slope}} \\ \underbrace{\gamma_{\max}}_{\text{label inconsistency cost}} & \text{otherwise,} \end{cases}$$





Ensures cost of breaking a good segment is higher than that of a bad segment

Robust P^N code: *http://sots.brookes.ac.uk/lubor/*

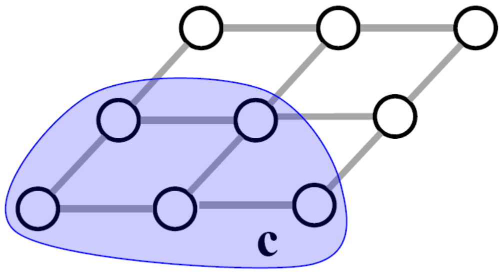
Slide adapted from P. Kohli

Segment Quality

Oxford Brookes Vision Group

- Label inconsistency cost depends on segment quality

$$\gamma_{\max} = |c|^{\theta_\alpha}\left(\theta_p^h + \theta_v^h \underbrace{G(c)}_{\text{variance of intensities}}\right)$$

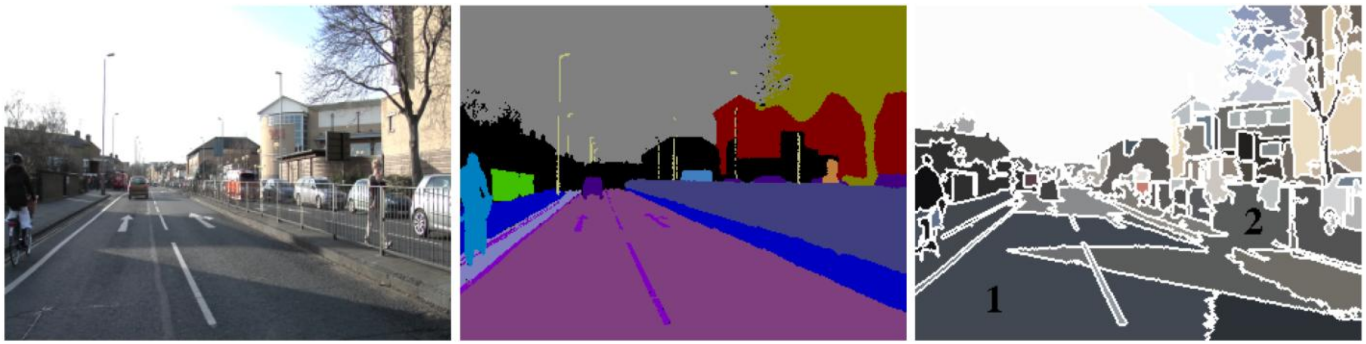


- Low variance indicates good quality
- High variance indicates poor quality

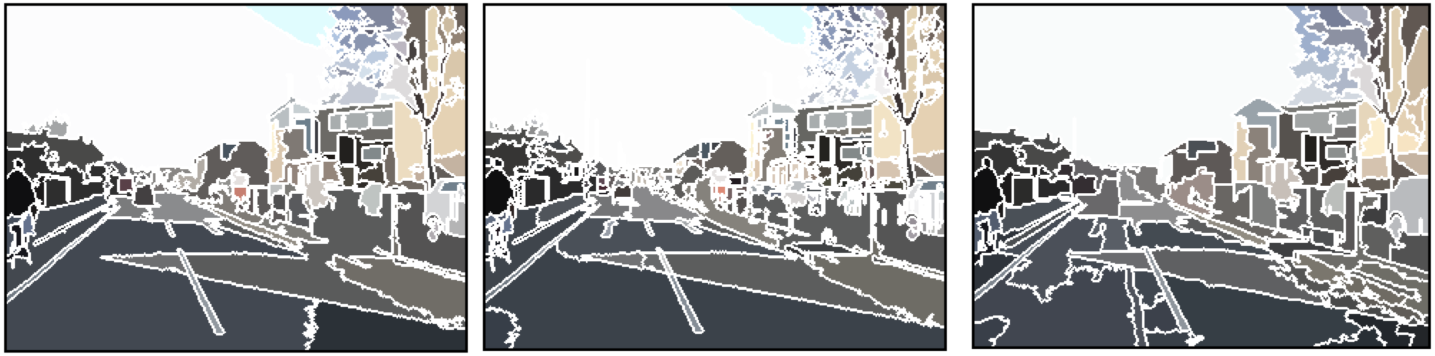
Multiple Segmentations

Oxford Brookes Vision Group

- Single Segmentation?



- Combine multiple segmentations



HO Potential Result

Unary
+Pairwise
+Higher Order

Ground

Raw

	Building	Tree	Sky	Car	Sign	Road	Pedestrian	Fence	Column	Sidewalk	Bicyclist	Average	Global
Brostow	46.2	61.9	89.7	68.6	42.9	89.5	**53.6**	**46.6**	0.7	60.5	22.5	53	69.1
Unary	61.9	67.3	91.1	71.1	**58.5**	92.9	49.5	37.6	**25.8**	77.8	24.7	59.8	76.4
+Pairwise	70.7	70.8	94.7	**74.4**	55.9	94.1	45.7	37.2	13	**79.3**	23.1	**59.9**	79.8
+HO	**84.5**	**72.6**	**97.5**	72.7	34.1	**95.3**	34.2	45.7	8.1	77.6	**28.5**	59.2	**83.8**

Columns = Per-class recall, Average = Average recall, Global = Overall correctly labelled pixels

Brostow et al ECCV 08

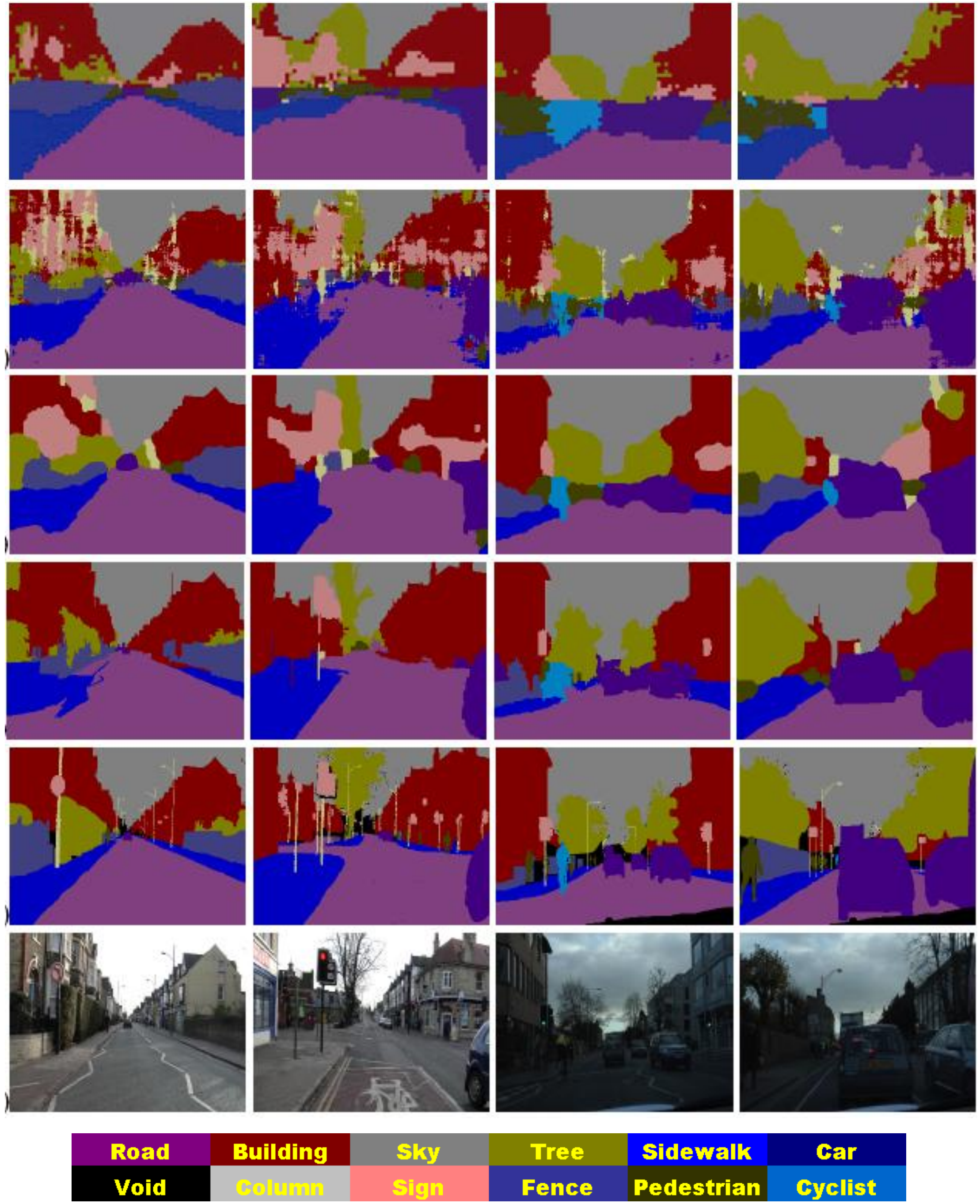
Unary

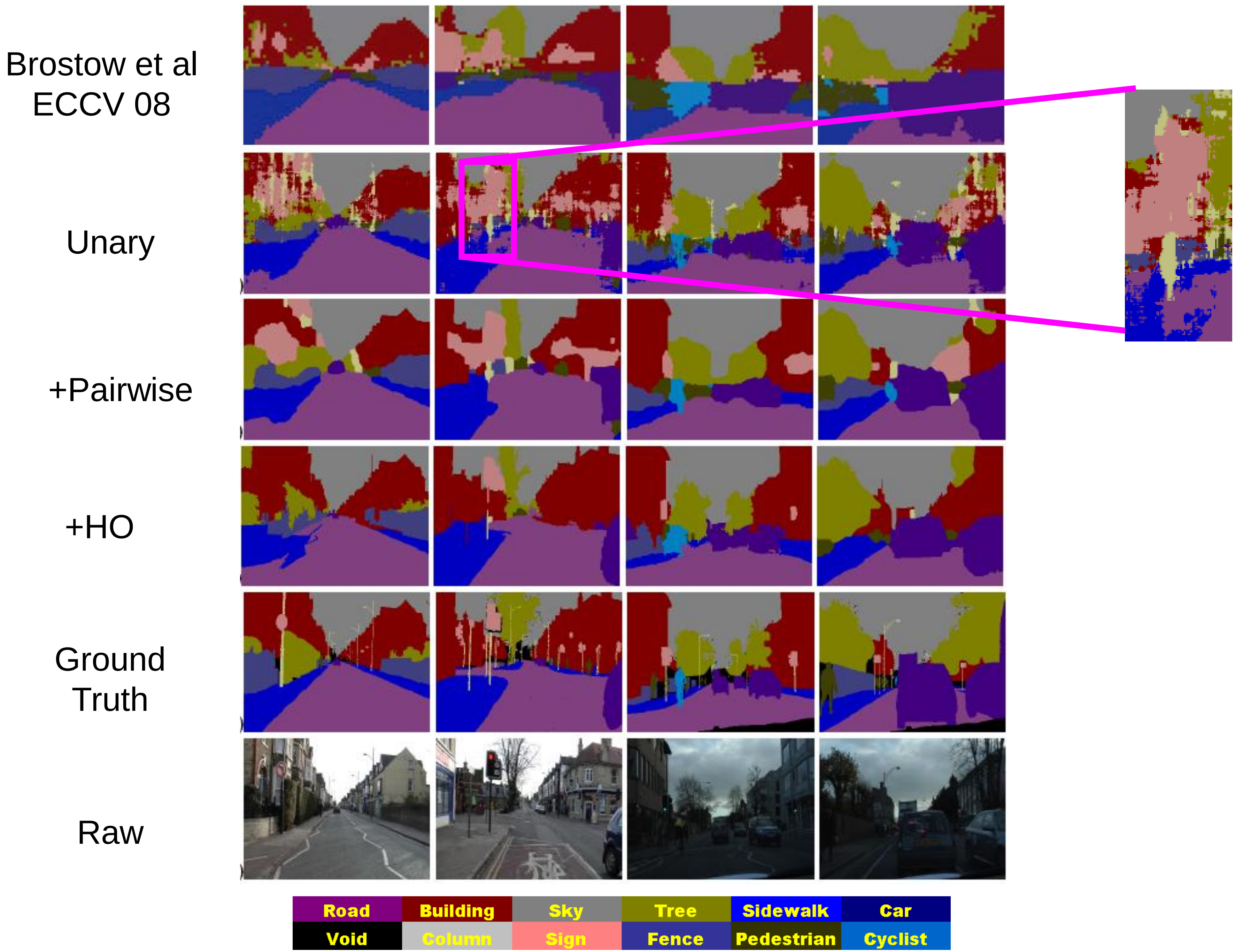
+Pairwise

+HO

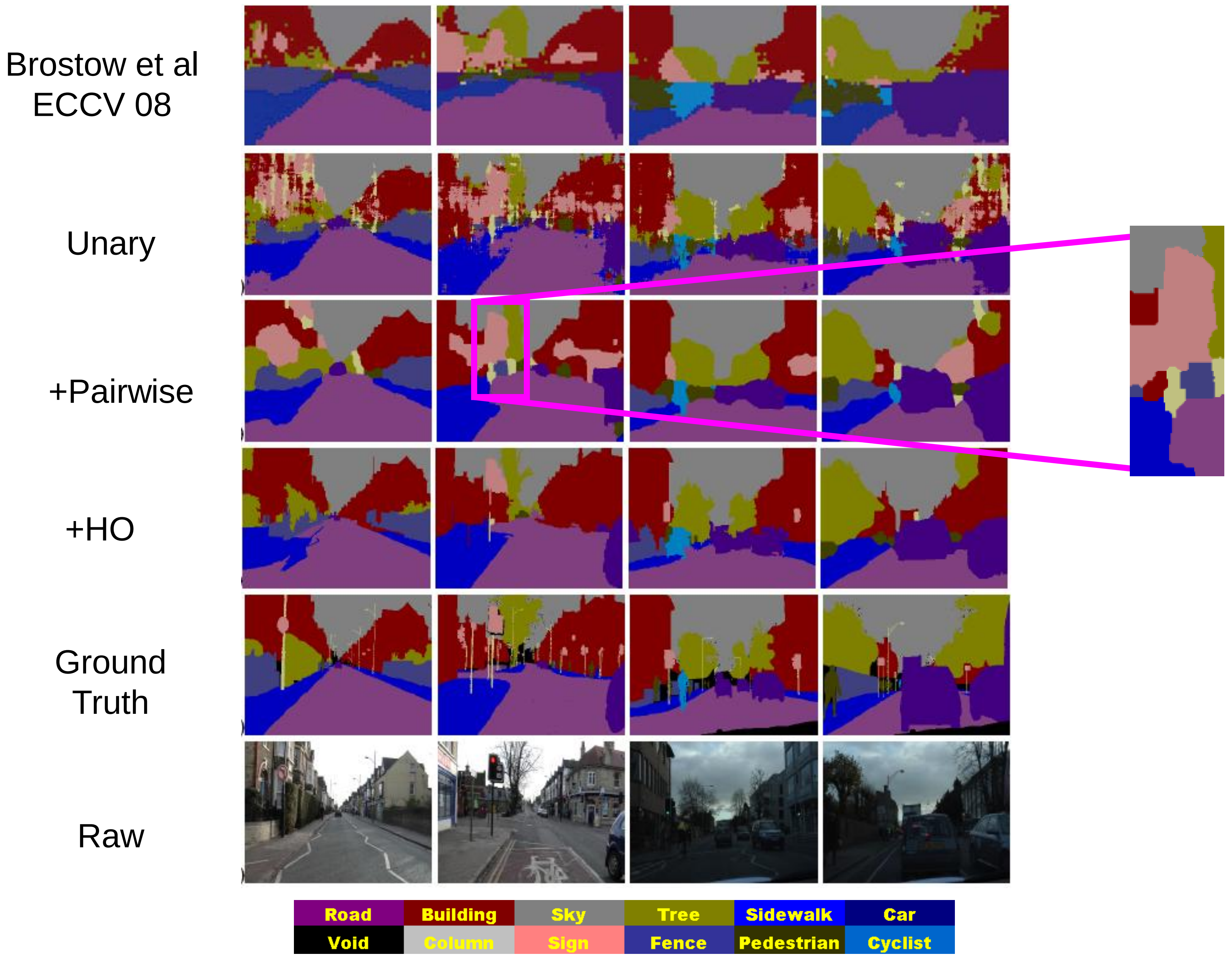
Ground Truth

Raw

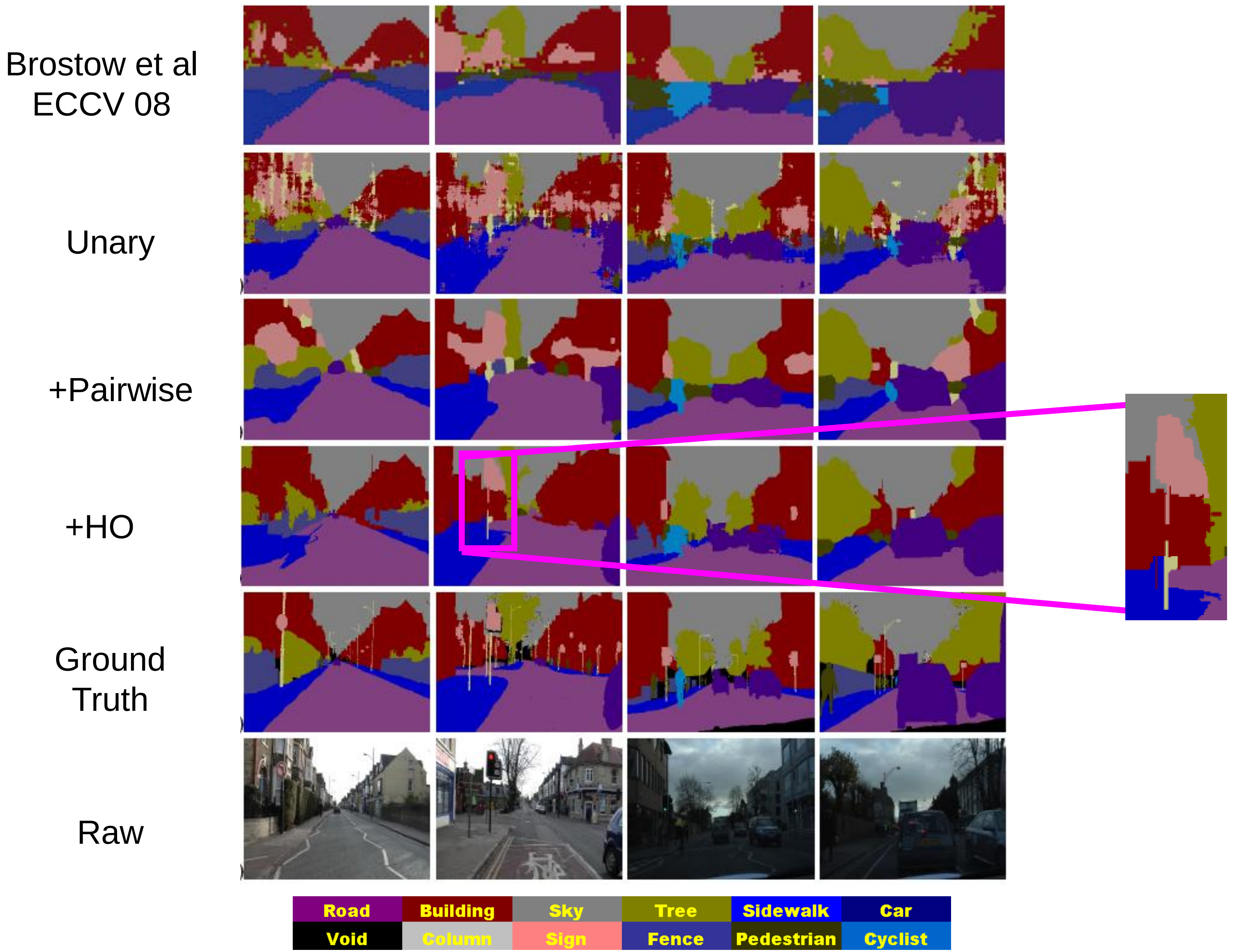




Brostow et al ECCV 08
Unary
+Pairwise
+HO
Ground Truth
Raw
Road
Building
Sky
Tree
Sidewalk
Car
Void
Column
Sign
Fence
Pedestrian
Cyclist



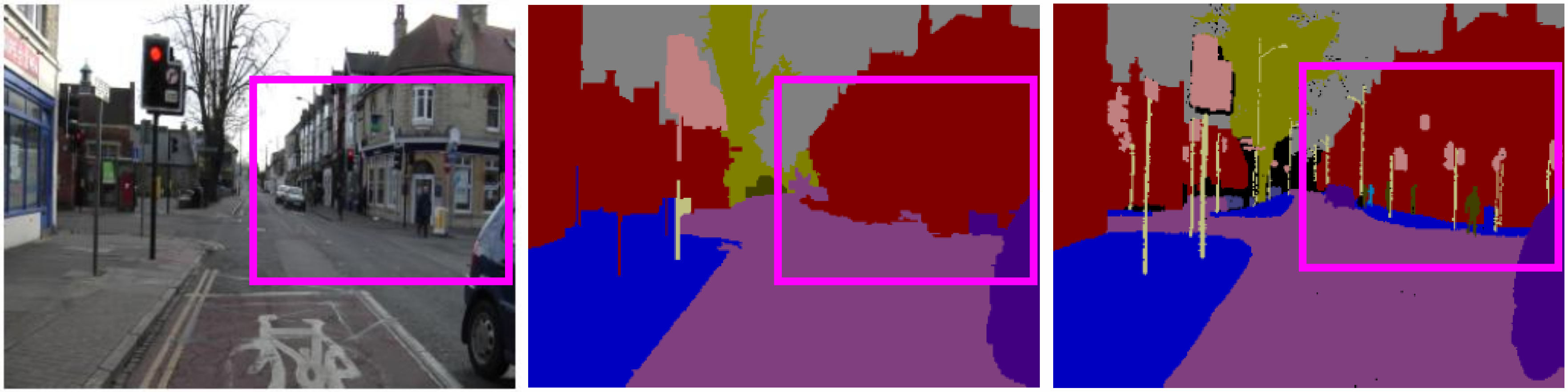
Brostow et al
ECCV 08
Unary
+Pairwise
+HO
Ground
Truth
Raw
Road
Building
Sky
Tree
Sidewalk
Car
Void
Column
Sign
Fence
Pedestrian
Cyclist



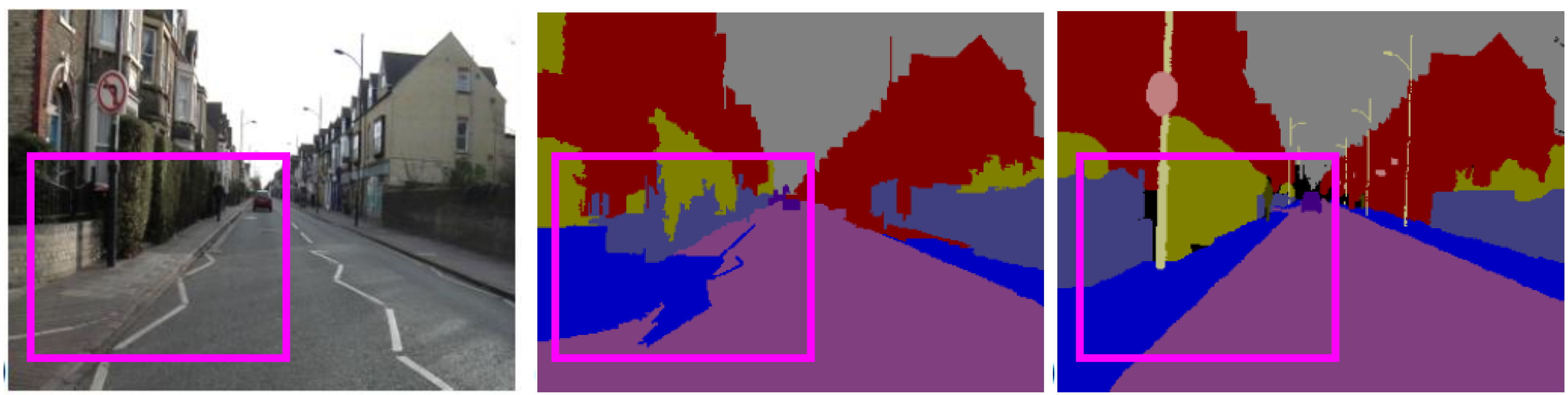
Brostow et al ECCV 08
Unary
+Pairwise
+HO
Ground Truth
Raw
Road
Building
Sky
Tree
Sidewalk
Car
Void
Column
Sign
Fence
Pedestrian
Cyclist

HO Problems

Oxford Brookes
Vision Group



Raw Higher Order Ground



Evaluation Summary

Oxford Brookes Vision Group

	Building	Tree	Sky	Car	Sign-Symbol	Road	Pedestrian	Fence	Column-Pole	Sidewalk	Bicyclist	**Average**	**Global**
Mot. [8]	43.9	46.2	79.5	44.6	19.5	82.5	24.4	**58.8**	0.1	61.8	18.0	43.6	61.8
App. [8]	38.7	60.7	90.1	71.1	51.4	88.6	**54.6**	40.1	1.1	55.5	23.6	52.3	66.5
Combined [8]	46.2	61.9	89.7	68.6	42.9	89.5	53.6	46.6	0.7	60.5	22.5	53.0	69.1
ψ_i	61.9	67.3	91.1	71.1	**58.5**	92.9	49.5	37.6	**25.8**	77.8	24.7	59.8	76.4
$\psi_i + \psi_{ij}$	70.7	70.8	94.7	**74.4**	55.9	94.1	45.7	37.2	13.0	**79.3**	23.1	**59.9**	79.8
$\psi_i + \psi_{ij} + \psi_c$	**84.5**	**72.6**	**97.5**	72.7	34.1	**95.3**	34.2	45.7	8.1	77.6	**28.5**	59.2	**83.8**

Columns = Per-class recall, Average = Average recall, Global = Overall correctly labelled pixels

- Improvement in 9 out of 11 classes
- Pairwise terms improve most classes
- Higher order terms further improve most classes

Evaluation Summary

Oxford Brookes Vision Group

	Building	Tree	Sky	Car	Sign-Symbol	Road	Pedestrian	Fence	Column-Pole	Sidewalk	Bicyclist	**Average**	**Global**
Mot. [8]	43.9	46.2	79.5	44.6	19.5	82.5	24.4	**58.8**	0.1	61.8	18.0	43.6	61.8
App. [8]	38.7	60.7	90.1	71.1	51.4	88.6	**54.6**	40.1	1.1	55.5	23.6	52.3	66.5
Combined [8]	46.2	61.9	89.7	68.6	42.9	89.5	53.6	46.6	0.7	60.5	22.5	53.0	69.1
ψ_i	61.9	67.3	91.1	71.1	**58.5**	92.9	49.5	37.6	**25.8**	77.8	24.7	59.8	76.4
$\psi_i + \psi_{ij}$	70.7	70.8	94.7	**74.4**	55.9	94.1	45.7	37.2	13.0	**79.3**	23.1	**59.9**	79.8
$\psi_i + \psi_{ij} + \psi_c$	**84.5**	**72.6**	**97.5**	72.7	34.1	**95.3**	34.2	45.7	8.1	77.6	**28.5**	59.2	**83.8**

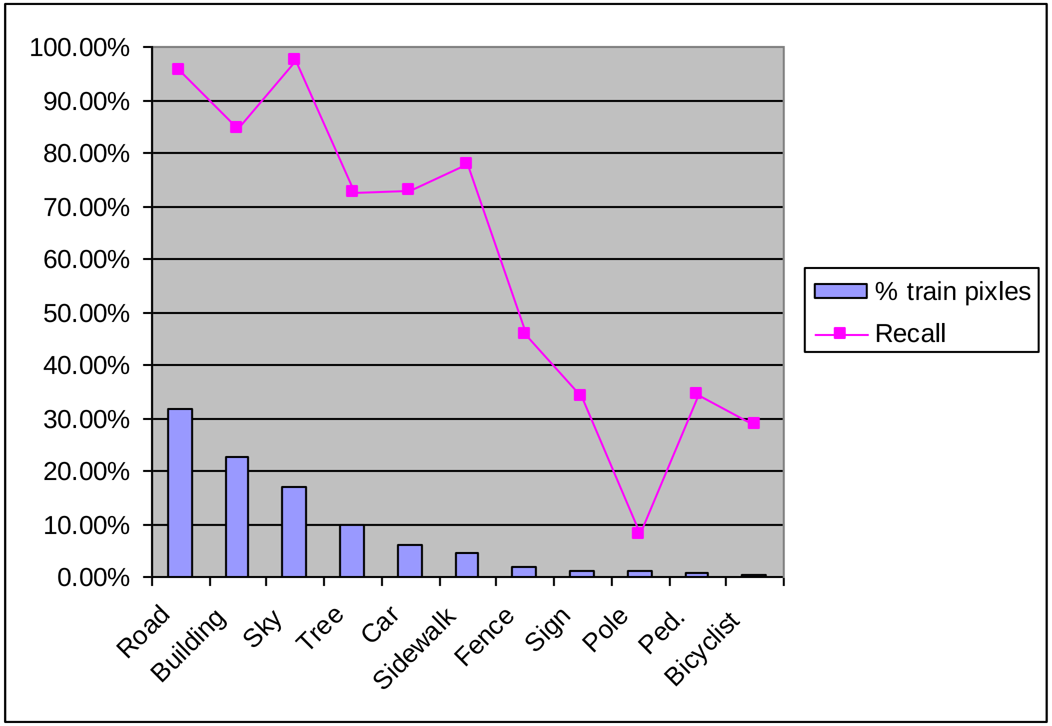
Columns = Per-class recall, Average = Average recall, Global = Overall correctly labelled pixels

- Improvement in 9 out of 11 classes
- Pairwise terms improve most classes
- Higher order terms further improve most classes
- Brostow et al ECCV08 better for 2 classes

Discussion: data

Oxford Brookes Vision Group

Recall Vs percent of class pixels in training data



- Column/pole=2,536,704 << building =57,583,181
- Poorer on all classes bellow 2% training pixels

Discussion: HO Problems

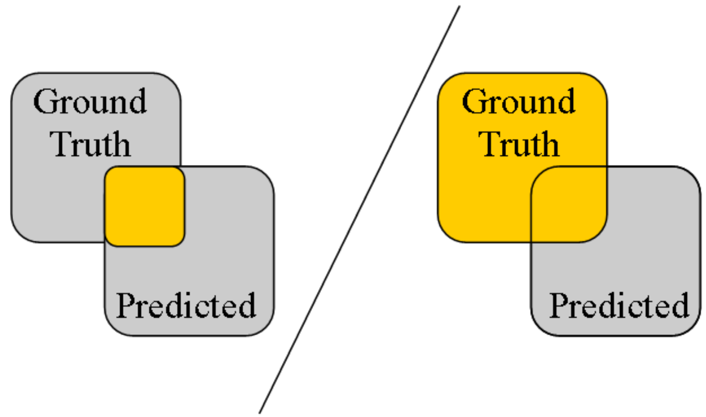
	Building	Tree	Sky	Car	Sign-Symbol	Road	Pedestrian	Fence	Column-Pole	Sidewalk	Bicyclist	**Average**	**Global**
ψ_i	61.9	67.3	91.1	71.1	**58.5**	92.9	49.5	37.6	**25.8**	77.8	24.7	59.8	76.4
$\psi_i + \psi_{ij}$	70.7	70.8	94.7	**74.4**	55.9	94.1	45.7	37.2	13.0	**79.3**	23.1	**59.9**	79.8
$\psi_i + \psi_{ij} + \psi_c$	**84.5**	**72.6**	**97.5**	72.7	34.1	**95.3**	34.2	45.7	8.1	77.6	**28.5**	59.2	**83.8**

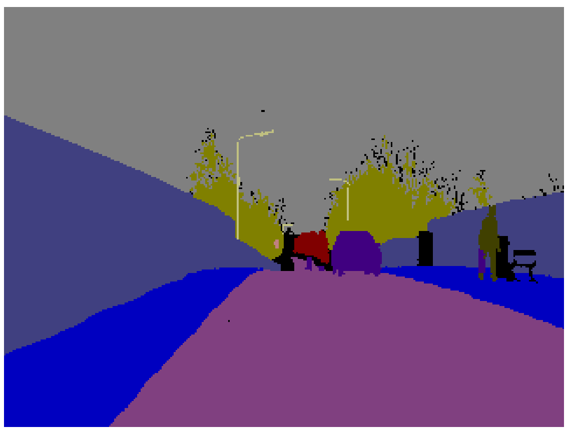
Columns = Per-class recall, Average = Average recall, Global = Overall correctly labelled pixels

- Decrease doesn't match with qualitative results

Discussion : Error

Recall =




Ground Truth

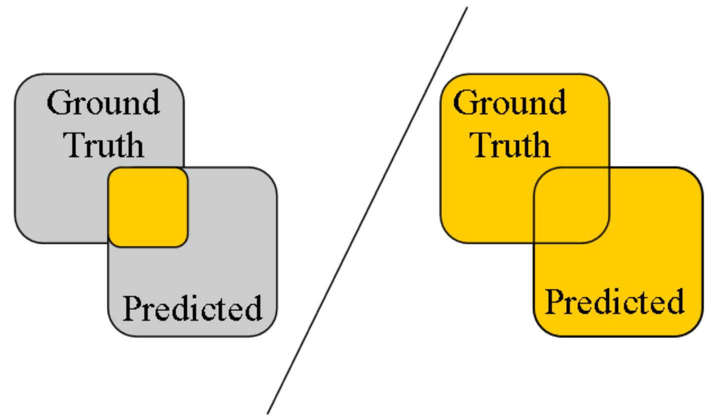
Labelling

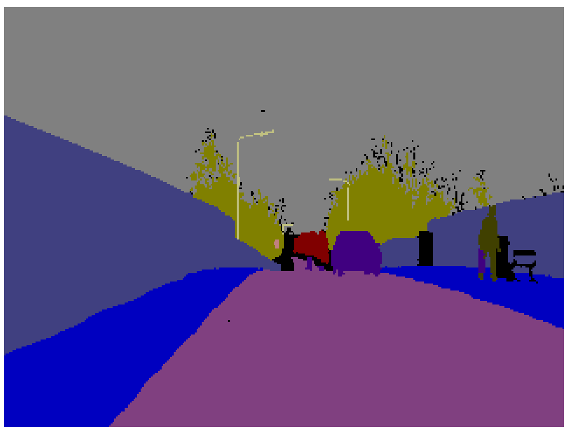
= 100% for column/pole

- Favours over estimates

Discussion : Error

Intersection/union =





Ground Truth

Labelling

= Almost 0% for column/pole

- Allows for an independent per-class error measurement
- Penalises both over- and under-estimates

Slide adapted from

Discussion : Error

Oxford Brookes Vision Group

- Intersection/union table

	Building	Tree	Sky	Car	Sign-Symbol	Road	Pedestrian	Fence	Column-Pole	Sidewalk	Bicyclist	**Average**
ψ_i	55.3	54.3	84.8	51.8	11.9	85.5	15.6	27.4	**7.5**	60.0	15.7	42.71
$\psi_i + \psi_{ij}$	63.6	58.0	87.8	55.9	13.6	86.4	16.9	27.6	6.1	61.9	18.1	45.07
$\psi_i + \psi_{ij} + \psi_c$	**71.6**	**60.4**	**89.5**	**58.3**	**19.4**	**86.6**	**26.1**	**35.0**	7.2	**63.8**	**22.6**	**49.15**

- Higher Order terms improve performance in all classes

Conclusion

Oxford Brookes Vision Group

- Strong unary potential from boosting
- HO terms yield more precise boundaries
- Improvement in 9 out of 11 classes
- Intersection/union error more informative
- Directions
  - Balance training data
  - Potentials for thin structures
  - Use Associative hierarchical CRFs

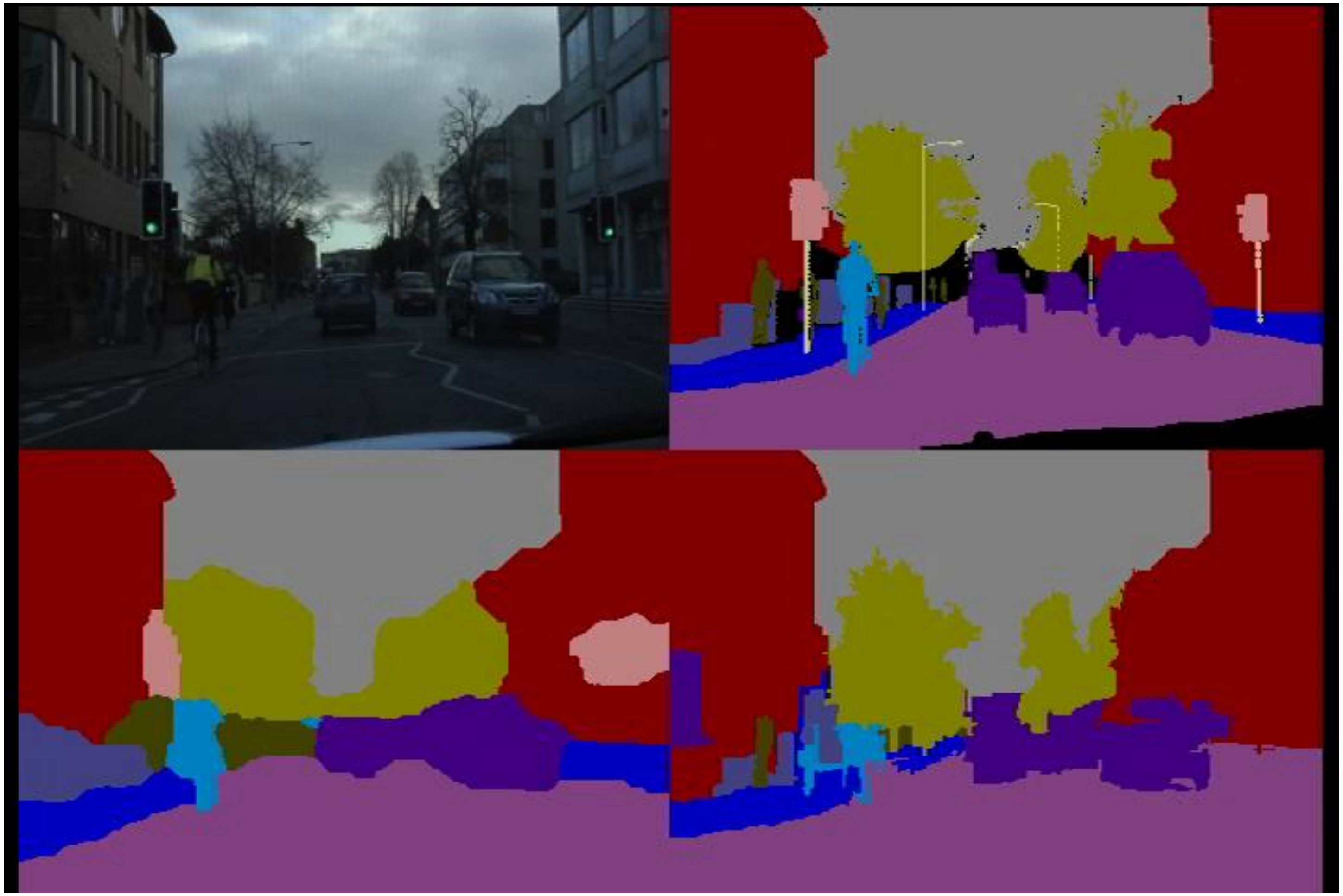
L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr.
Associative hierarchical crfs for object class image segmentation. ICCV 2009.

Questions

Oxford Brookes Vision Group

Raw Image

Ground Truth



Unary + Pairwise

Unary + Pairwise + Higher Order

Road	Building	Sky	Tree	Sidewalk	Car
Void	Column	Sign	Fence	Pedestrian	Cyclist