Probabilistic Graphical Models

Michael Yang

September 12, 2017

Brief CV

- Since 2016, Assistant Professor, University of Twente
- 2008-2011, Ph.D, University of Bonn
- 2016-2020, Co-Chair ISPRS WG Dynamic Scene Analysis
- Main Research Areas: Photogrammetry, Computer Vision

Outline

- Introduction

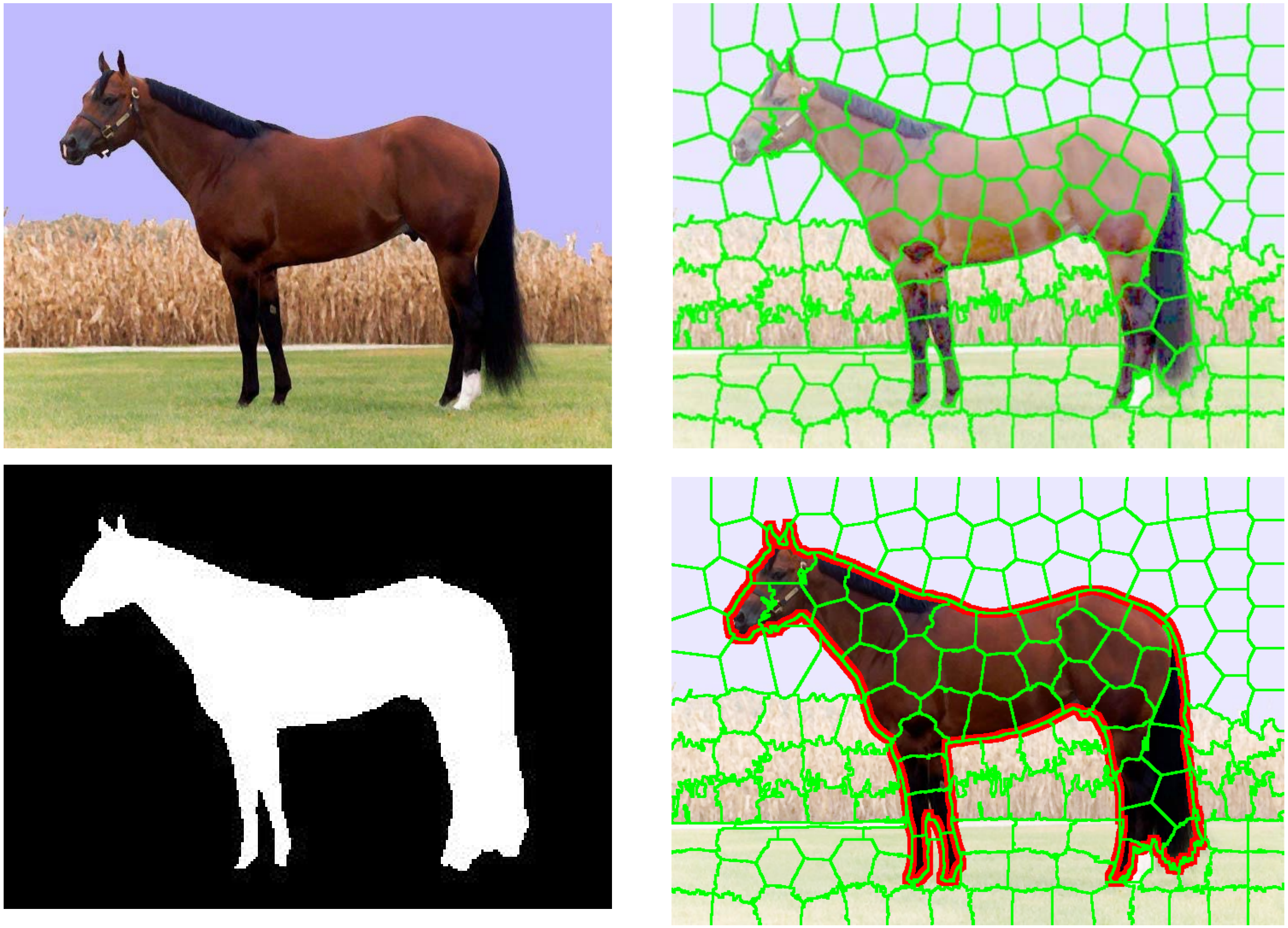
- Random Fields

  - Man-made Object Segmentation
  - Semantic Video Segmentation

Applications

- Medical diagnosis
- Social network models
- Speech recognition
- Robot localization
- Remote sensing
- Natural language processing
- Computer vision
  - Image segmentation
  - Tracking
  - Scene understanding
  - Image classification
  - 3D reconstruction
- ....

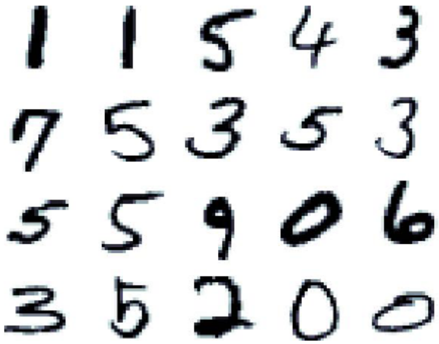
Segmentation



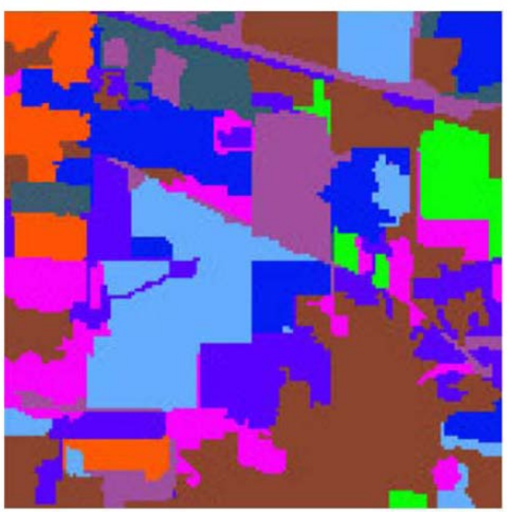
Yang, Rosenhahn, 2016

Applications

Classification



(MNIST benchmark data)



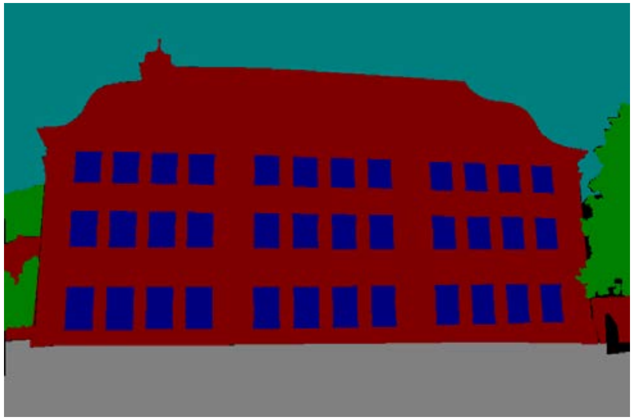
Zhong & Wang 2011

- Reading letters/numbers
- Land-cover classification in remote sensing

Interpretation

Chai et al., 2013

- Building and road extraction



Yang & Förstner, 2011

- Facade interpretation

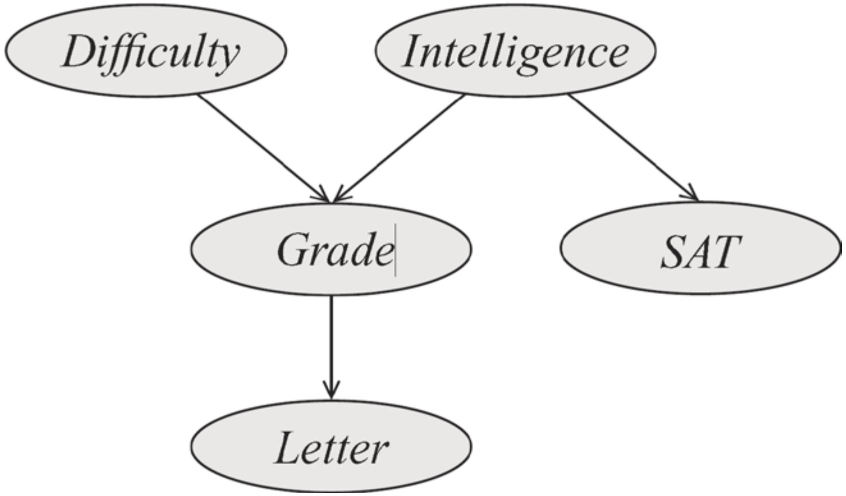
Probabilistic Graphical Models

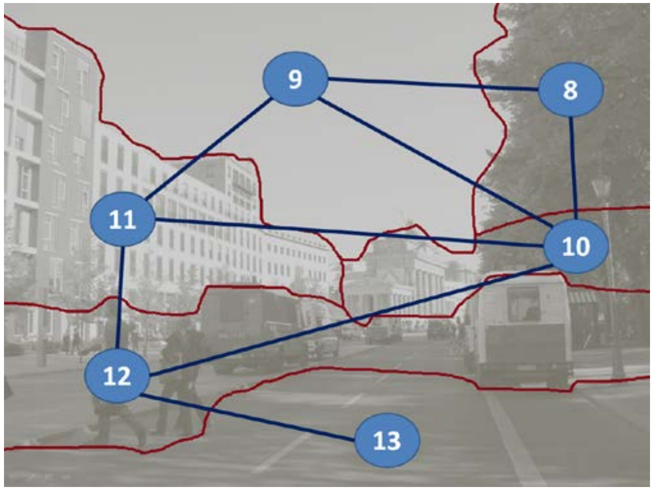
are a marriage between

probability theory & graph theory

Bayesian networks

Conditional/Markov random fields





- **Graph** $\mathcal{G}$

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$$

set of the nodes $\mathcal{V} = \{1, \cdots, i, \cdots, n\}$

set of the undirected edges

$$\mathcal{E} = \{\{i, j\} \mid i, j \in \mathcal{V}\}$$

set of the directed edges

$$\mathcal{A} = \{(i, j) \mid i, j \in \mathcal{V}\}$$

- **Graphical models**

A stochastical model represented by a graph $\mathcal{G}$

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$$

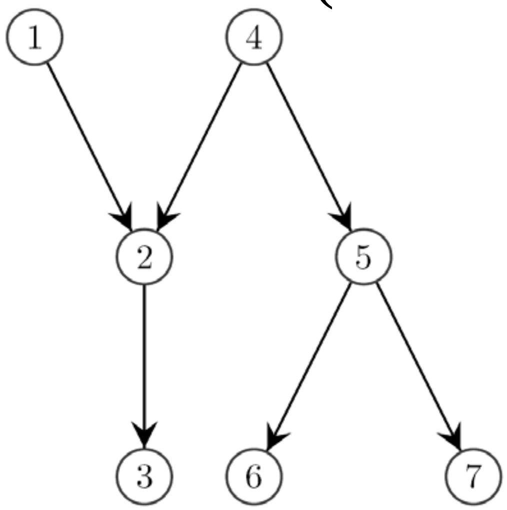
- Nodes $i \in \mathcal{V}$ represent random variables $\underline{\mathbf{x}}_i$
- Edges represent mutual relationships
  - ➢ Undirected edges $\{i, j\}$ model joint probabilities

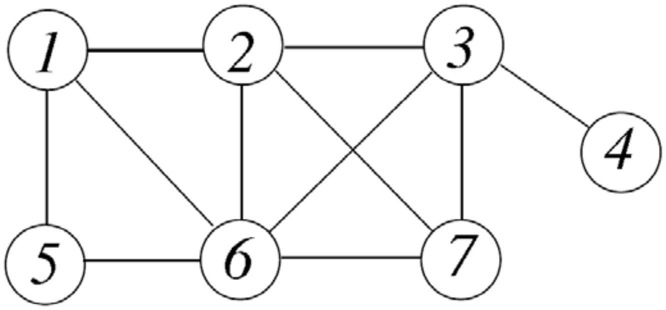
$$P(\mathbf{x}_i, \mathbf{x}_j)$$

  - ➢ Directed edges (i, j) model conditional dependencies

$$P(\mathbf{x}_j \mid \mathbf{x}_i)$$

- **Graphical models**
  - Visualization of dependencies
    - Conditional probabilities : directed edges (**Bayesian Networks**)
    - Joint probabilities: undirected edges (**Markov Random Field**)





Outline

- Introduction
- **Random Fields**
  - Man-made Object Segmentation
  - Semantic Video Segmentation

MRFs

- **Definition**

  Markov random field : graphical model over an undirected graph

  \+ positivity property + Markov property $\mathcal{H} = (\mathcal{V}, \mathcal{E})$

  $$P(\mathbf{x}) > 0$$

  ➢ Set of random variables linked to nodes

  $$\{\underline{x}_i, i \in \mathcal{V}\} \qquad \underline{\mathbf{x}} = [\underline{x}_i]$$

  ➢ Set of neighbored random variable

  $$\mathcal{N}(x_i) = \{x_j \mid j \in \mathcal{N}_i\}$$

  Markov property:

  $$P(x_i \mid \mathbf{x}_{\mathcal{V}-\{i\}}) = P(x_i \mid \mathbf{x}_{\mathcal{N}_i})$$

MRFs

- **Pairwise MRFs**

  popular

$$P(\mathbf{x}) = \frac{1}{Z}\exp(-E(\mathbf{x}))$$

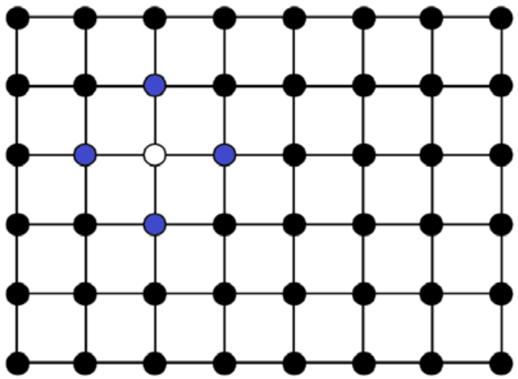
with energy function

$$E = \sum_{i\in\mathcal{V}} \underbrace{E_1(x_i)}_{Unary} + \alpha \sum_{\{i,j\}\in\mathcal{N}} \underbrace{E_2(x_i, x_j)}_{Pairwise}$$

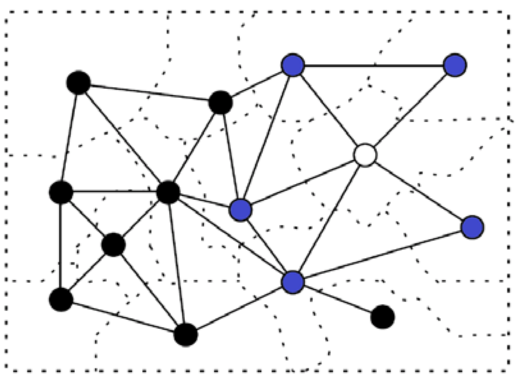
MRFs

- **Structure of MRFs**

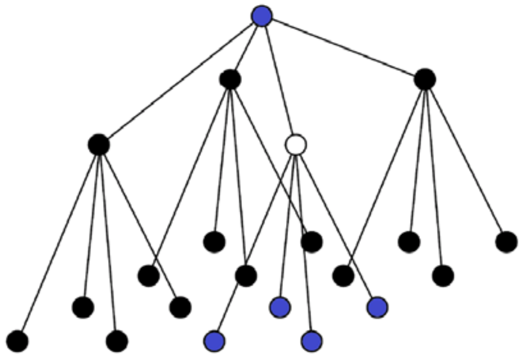
  Typical graph structures



rectangular grid



irregular graph

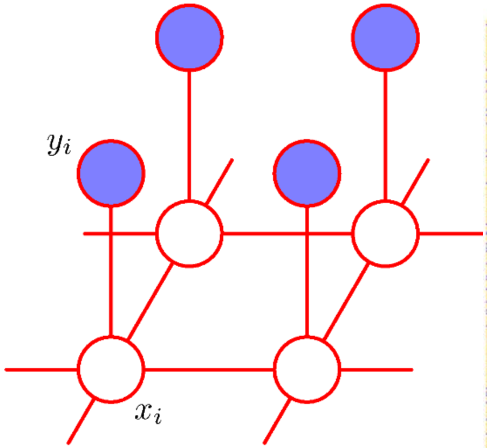


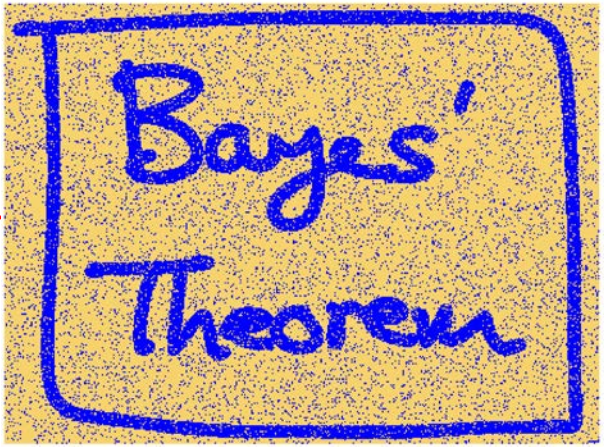
pyramid structure

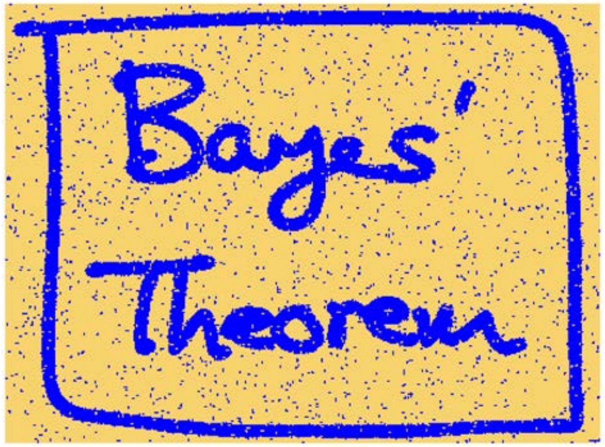
Figure courtesy of P. Perez

MRFs

- **Image Denoising using Pairwise MRFs**







[From Bishop PRML] noisy image result

CRFs

- **Definition: conditioanl random fields**

A CRF is an MRF globally conditioned on observed data

CRFs

- **Definition: conditioanl random fields**

A CRF is an MRF globally conditioned on observed data

MRF

Joint distribution

$$P(\boldsymbol{x}, \boldsymbol{d}) = \frac{1}{Z} \exp\left(-E(\boldsymbol{x})\right) = \frac{1}{Z} \exp\left(-\sum_{c \in C} \phi_c(\boldsymbol{x}_c)\right)$$

CRF

Conditional distribution

$$P(\boldsymbol{x} \mid \boldsymbol{d}) = \frac{1}{Z} \exp\left(-E(\boldsymbol{x} \mid \boldsymbol{d})\right) = \frac{1}{Z} \exp\left(-\sum_{c} \phi_c(\boldsymbol{x}_c \mid \boldsymbol{d})\right)$$

Outline

- Introduction

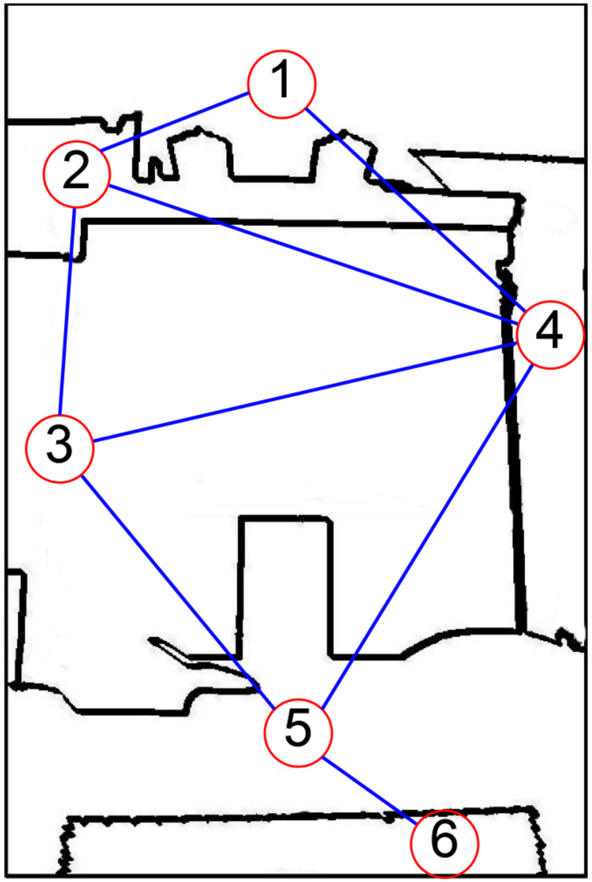
- **Random Fields**

  - **Man-made Object Segmentation**
  - Semantic Video Segmentation

CRFs

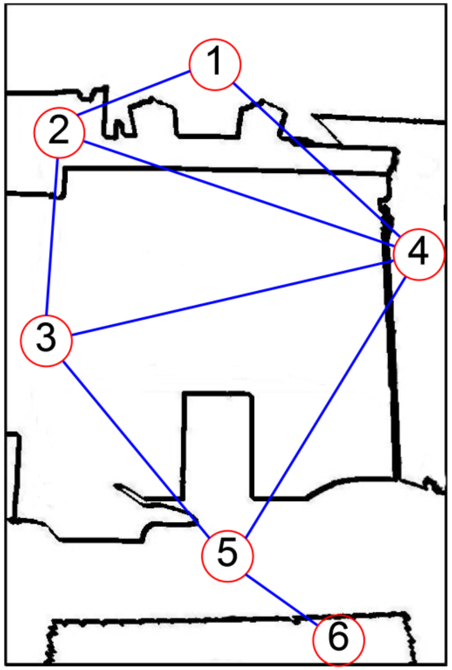
Yang & Förstner, 2011

Building facade image



Region adjacency graph

CRFs



CRF has a Gibbs distribution

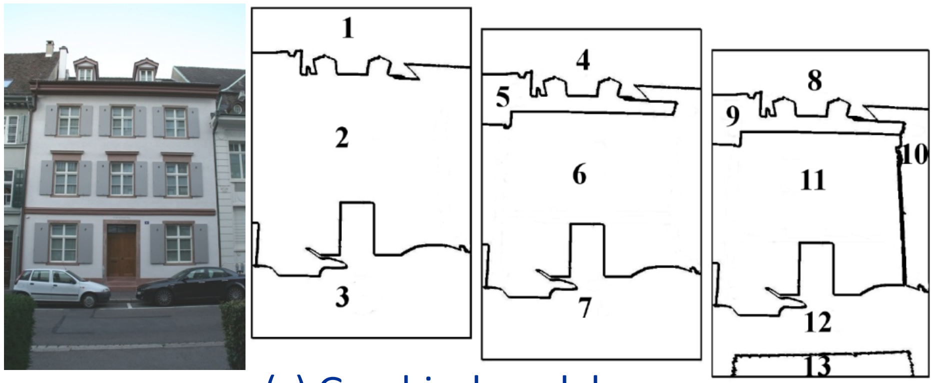
$$P(\boldsymbol{x} \mid \boldsymbol{d}) = \frac{1}{Z} \exp\left(-E(\boldsymbol{x} \mid \boldsymbol{d})\right)$$

Gibbs energy function (*all dependent on data*)

$$E = \sum_{i \in \mathcal{V}} \underbrace{E_1(x_i)}_{Unary} + \alpha \sum_{\{i,j\} \in \mathcal{N}} \underbrace{E_2(x_i, x_j)}_{Pairwise}$$

<u>Yang</u> & Förstner, 2011

(a) Test image (b) Multi-scale segmentation



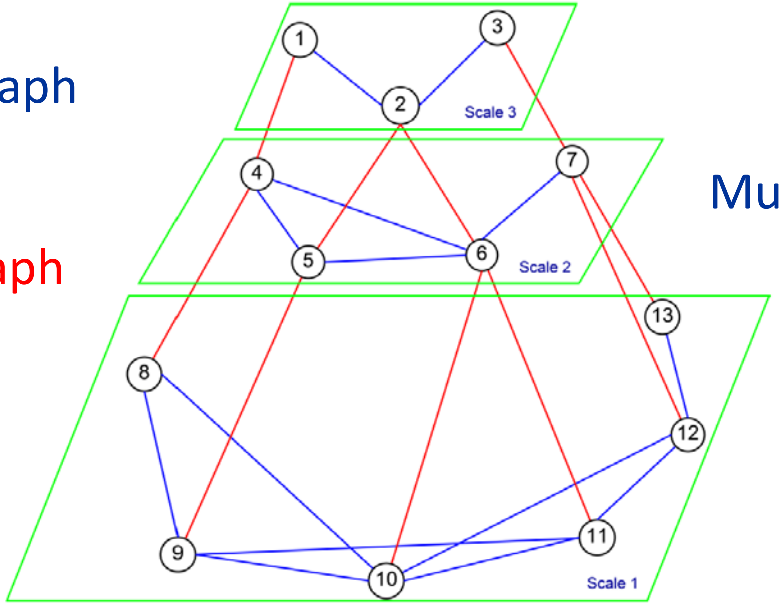
(c) Graphical model

Region adjacency graph

Blue edges

Region hierarchy graph

Red edges



Multi-layer CRF

Hierarchical CRFs

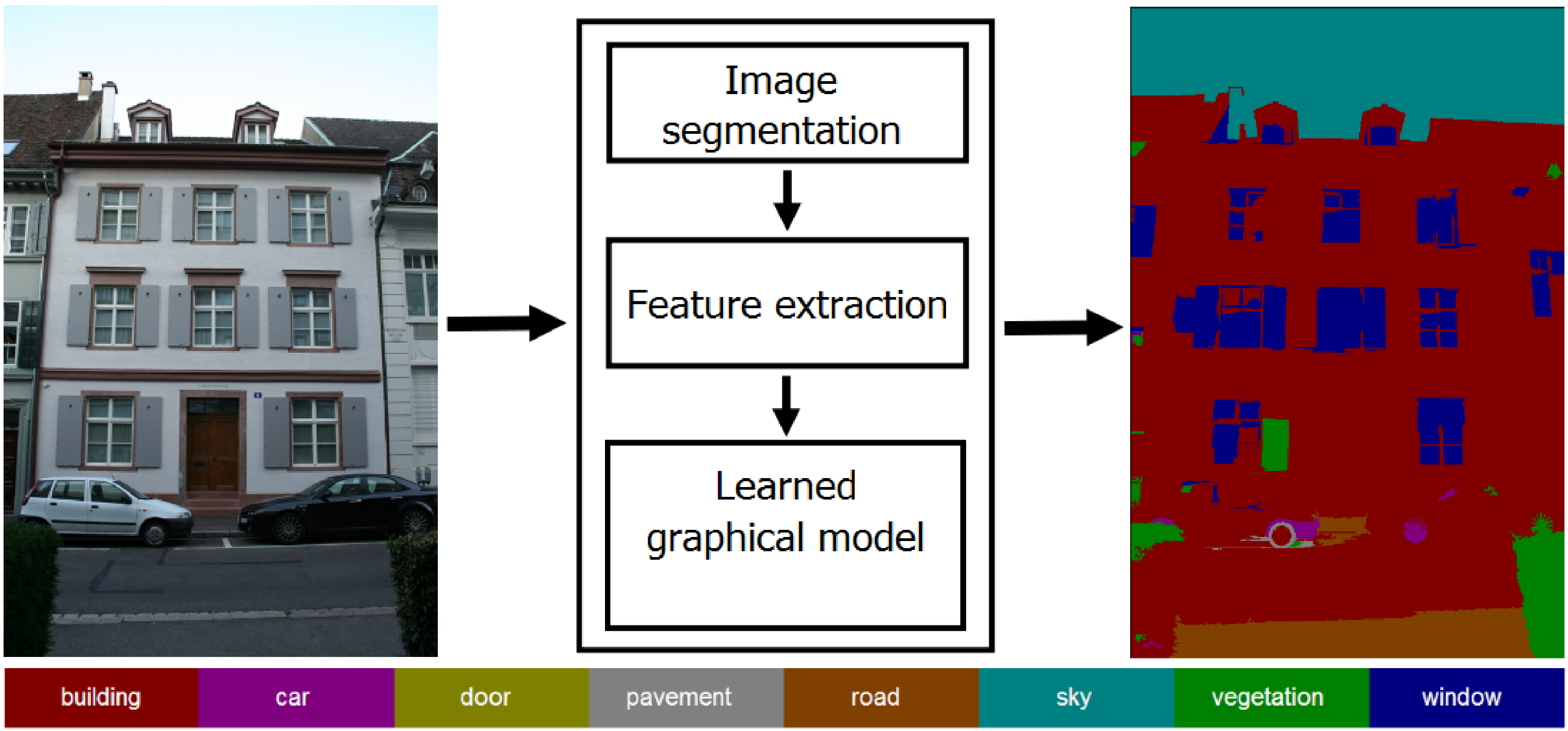
Energy function

$$E = \sum_{i \in \mathcal{V}} \underbrace{E_1(x_i)}_{Unary} + \alpha \sum_{\{i,j\} \in \mathcal{N}} \underbrace{E_2(x_i, x_j)}_{Pairwise} + \beta \sum_{\{i,k\} \in \mathcal{H}} \underbrace{E_3(x_i, x_k)}_{Hierarchical}$$

- Unary potential: classifier output (RF)
- Pairwise potential: (Data-dependent) Potts
- Hierarchical potential: (Data-dependent) Potts

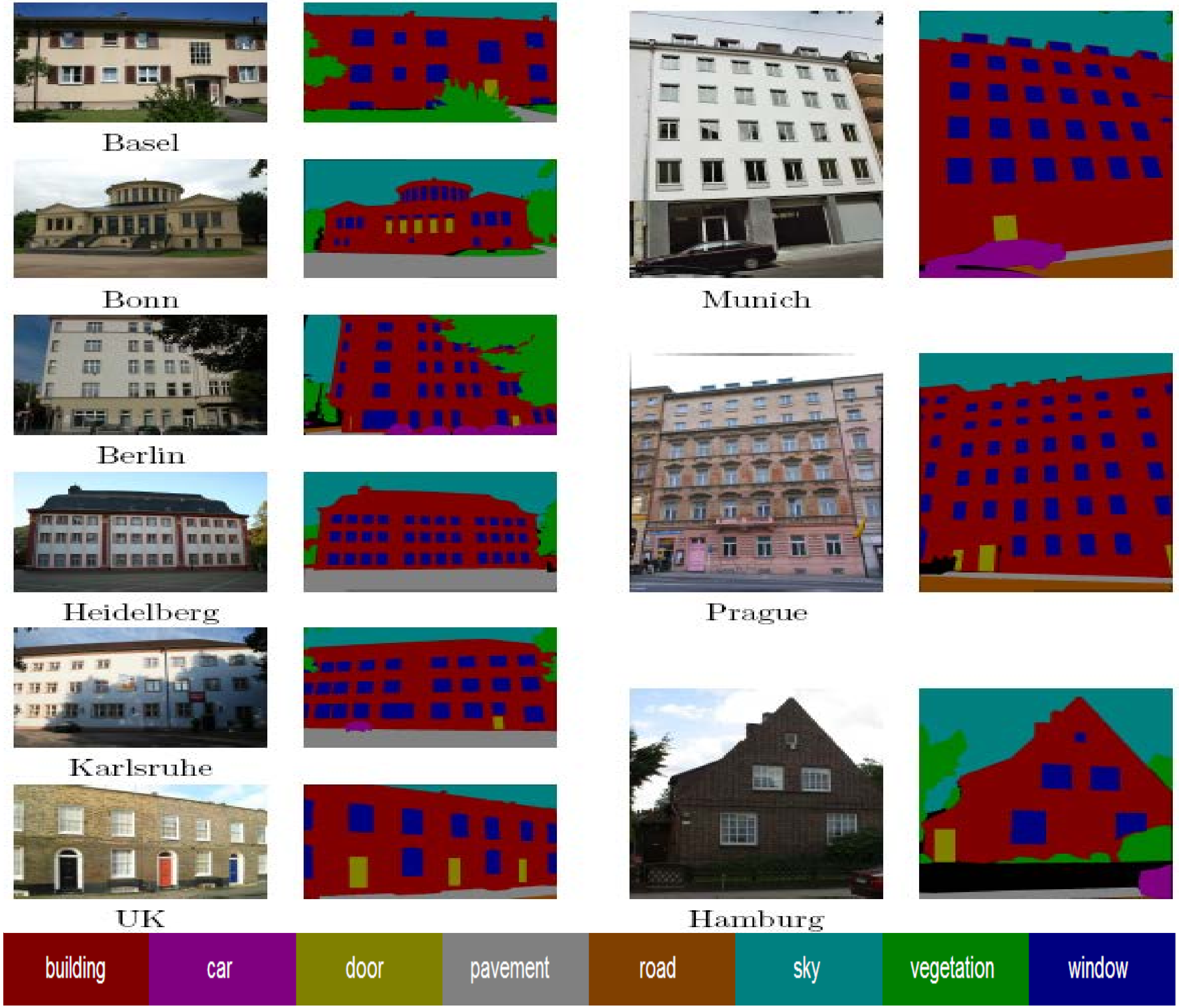
Scene Interpretation

Framework



Workflow for image interpretation of man-made scenes

ETRIMS Database

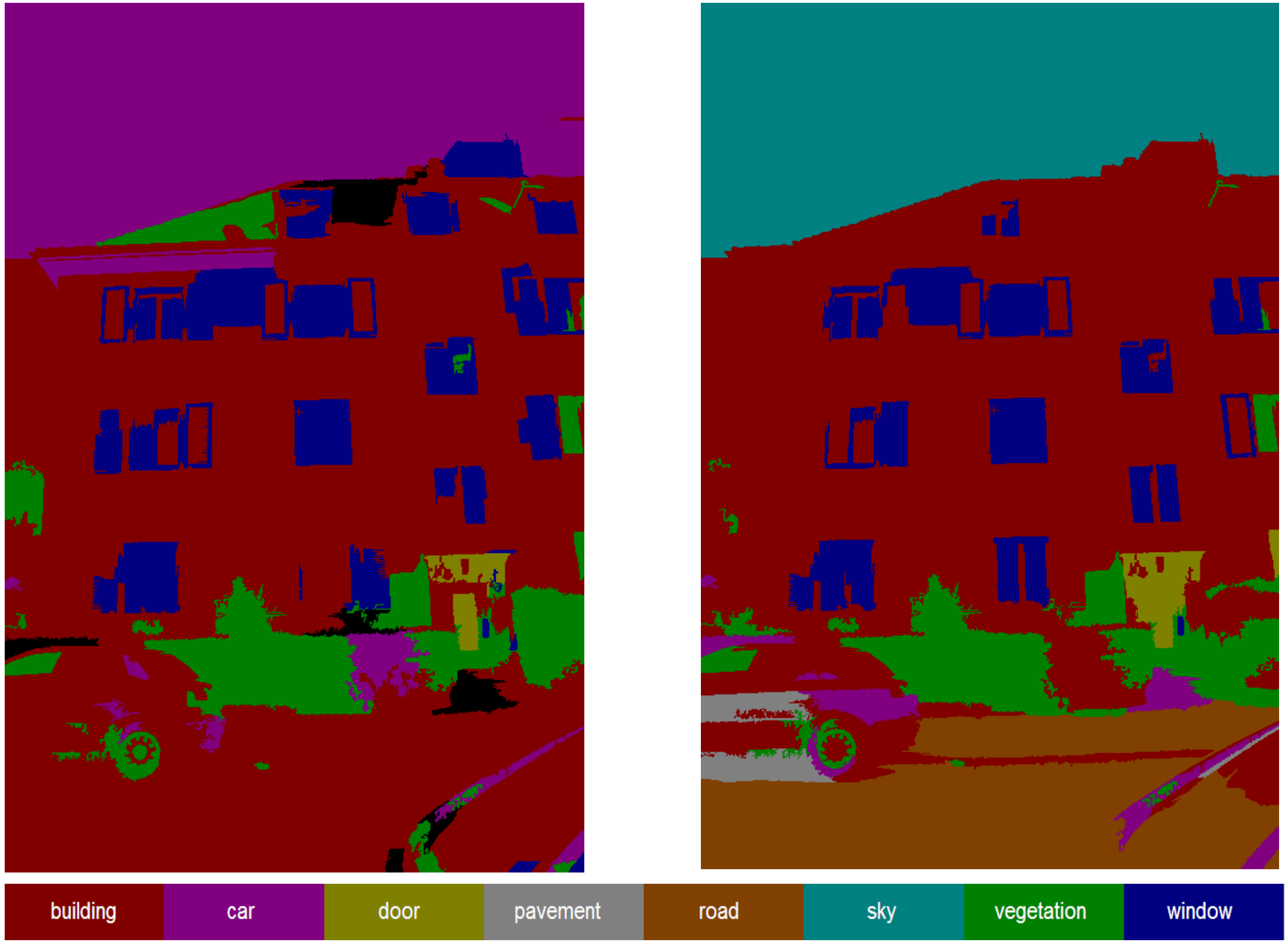


Example Image

One example image

Ground truth labeling

Classification Results



Region classifier (RDF)   Pairwise CRF

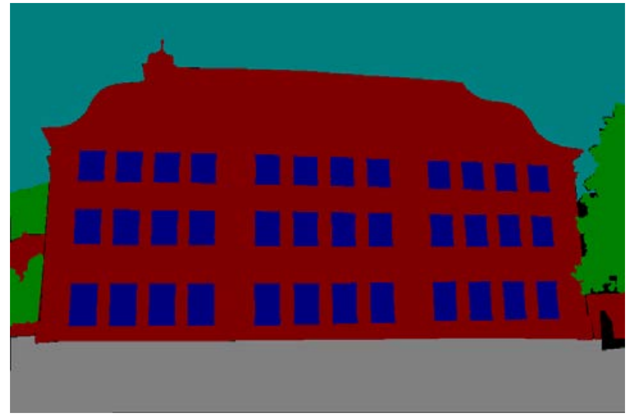
HCRF Results

Image RDF CRF HCRF GT

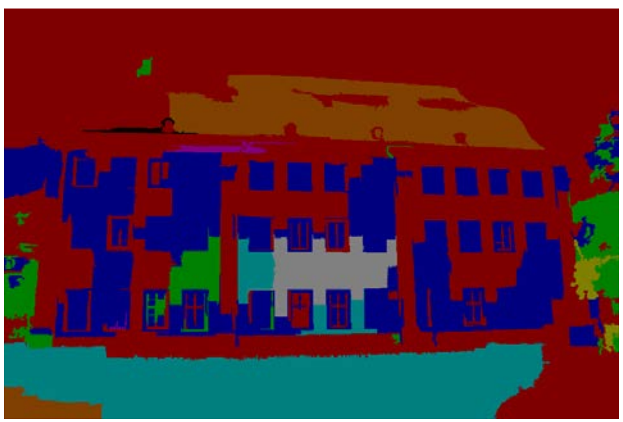
HCRF Results

Image

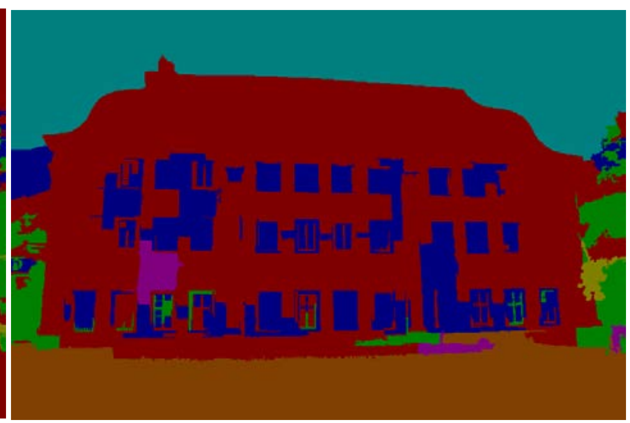
GT



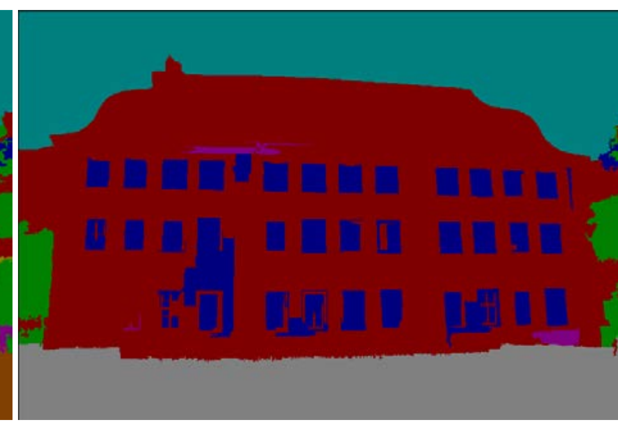
RDF



CRF



HCRF



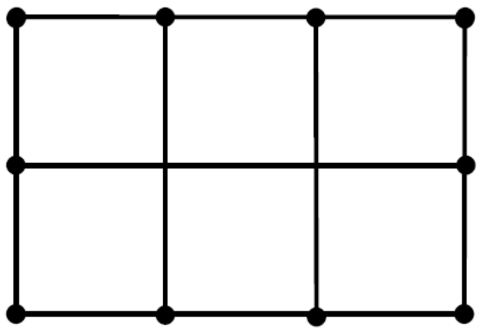
building	car	door	pavement	road	sky	vegetation	window

HCRF Results

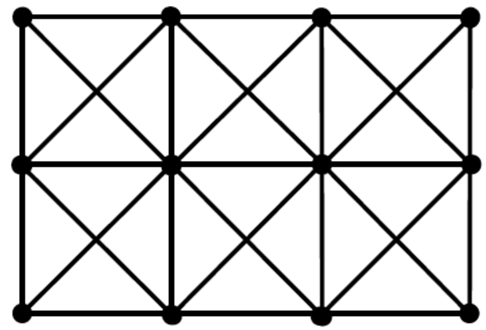
Pixelwise accuracy comparison

C \ S	watershed	mean shift
RDF	55.4%	58.8%
CRF	61.8%	65.8%
HCRF	65.3%	69.0%

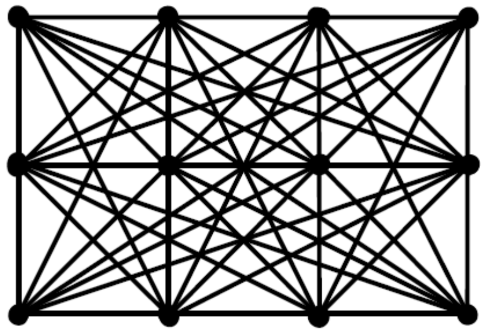
Fully Connected CRF



4-connected CRF



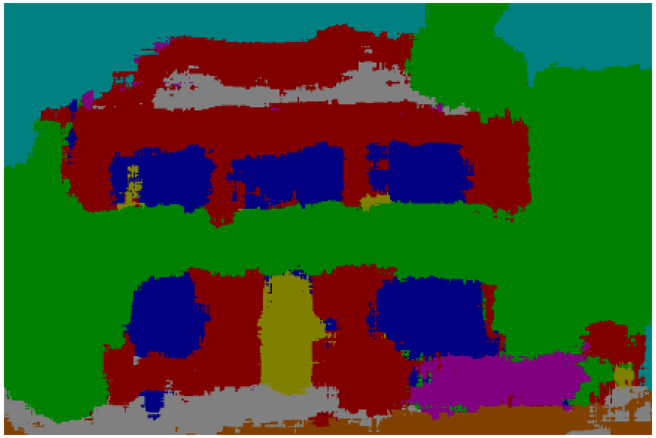
8-connected CRF



Fully-connected CRF

Fully Connected CRF

Image



Unary

Final

Li, Yang, 2016

Fully Connected CRF

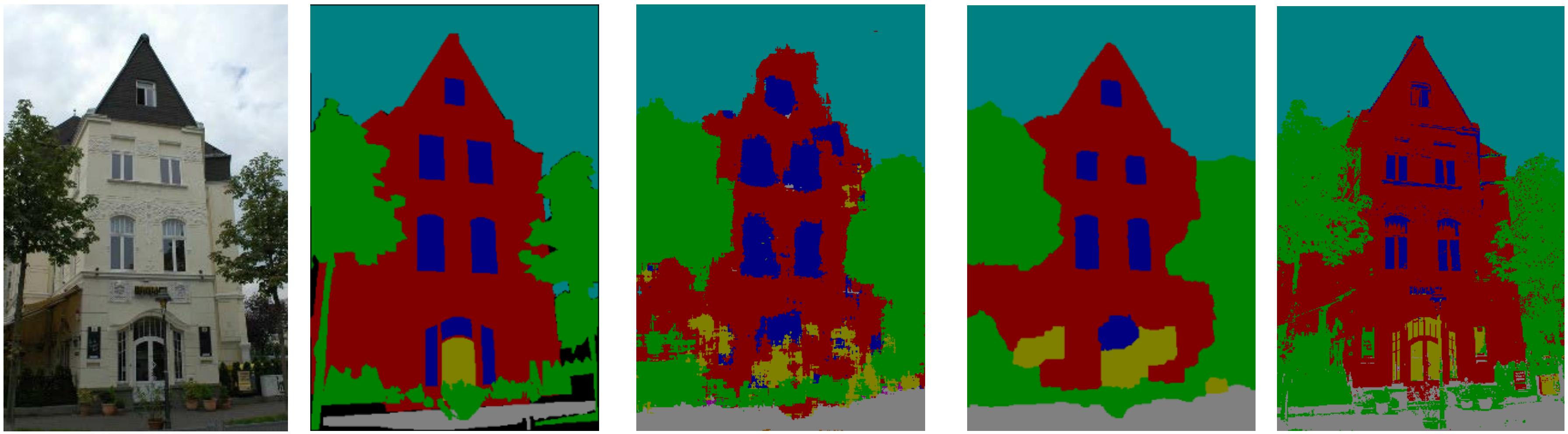
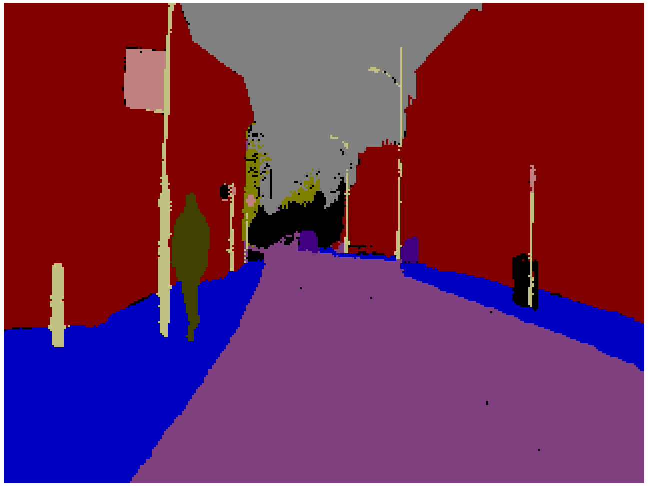


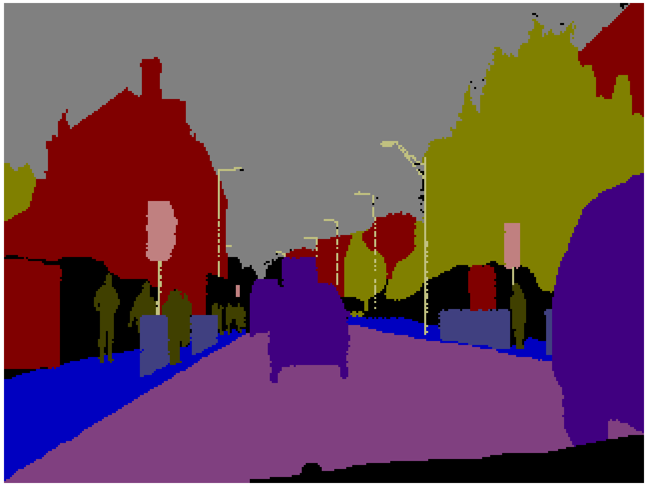
Image GT Texonboost CRF FC-CRF

Outline

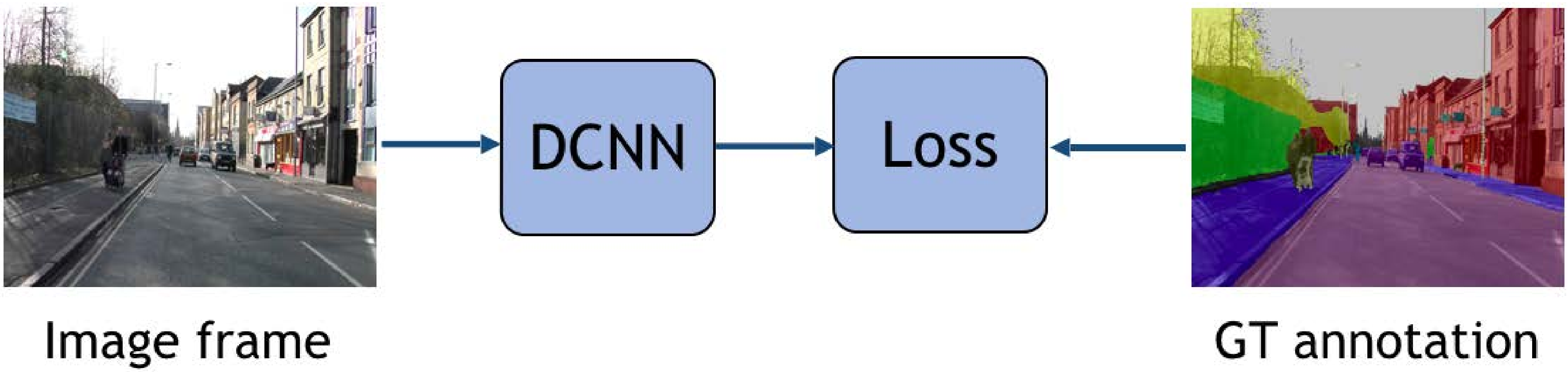
- Introduction
- Random Fields
  - Man-made Object Segmentation
  - **Semantic Video Segmentation**

Semantic Video Segmentation





Deep Learning for Semantic Video Segmentation

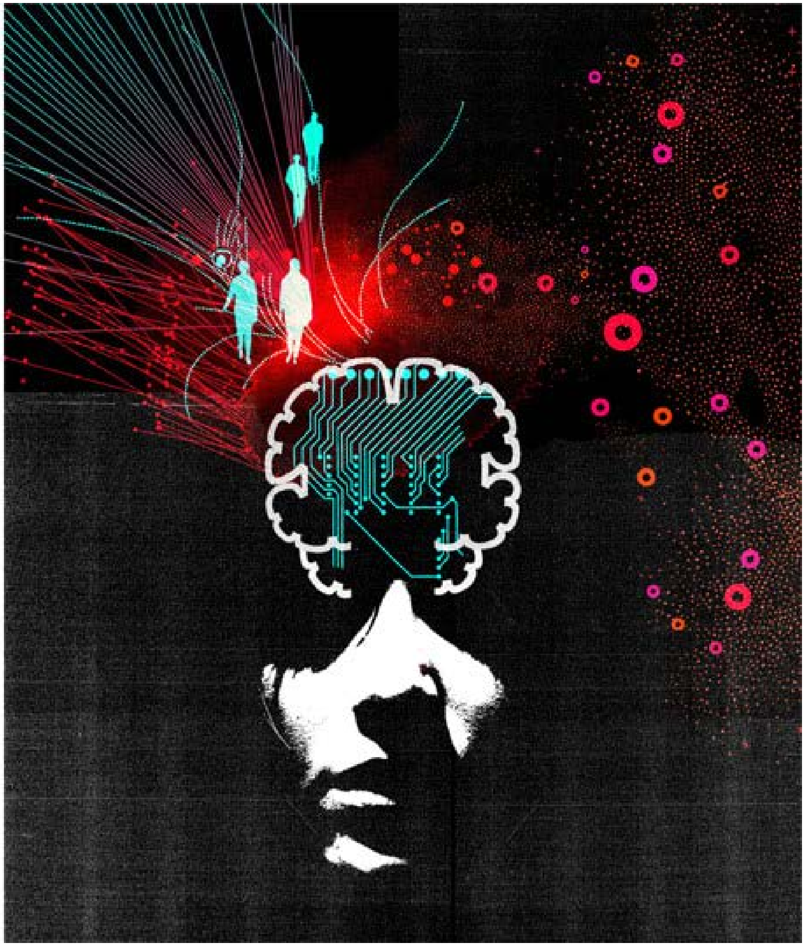


Badrinarayanan, Handa, Cipolla, arXiv 2015
SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling

MIT Technology Review

10 BREAKTHROUGH TECHNOLOGIES 2013

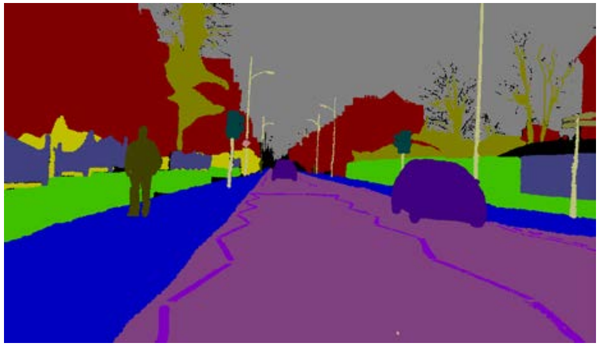
Deep Learning



Semantic Video Segmentation

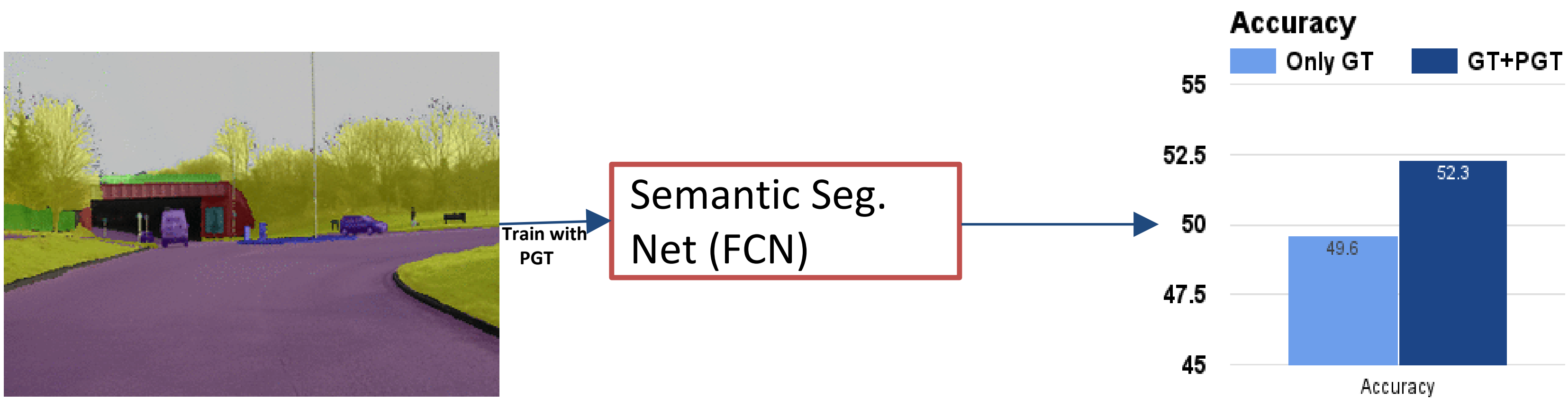
- Training CNN requires large amount of ground-truth data

- Dense labeling requires extensive human effort
- Labeling one image from CityScapes ~ 1.5 hours



Semantic Video Segmentation

- Use video to propagate labels. *Pseudo Ground Truth (PGT)*

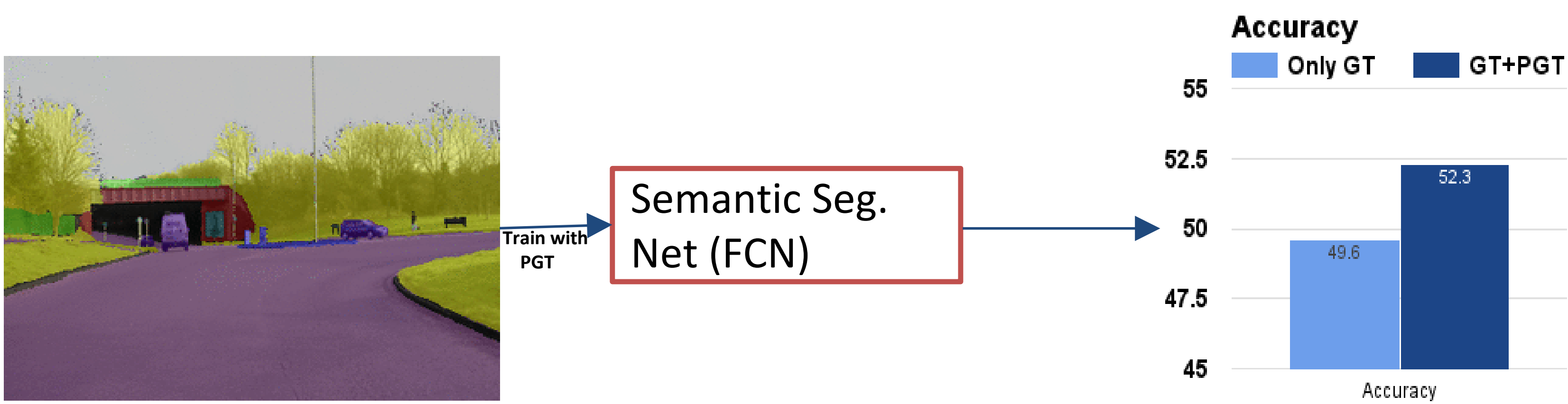


Mustikovela, <u>Yang</u>, Rother, ECCV Workshop 2016
Can Ground Truth Label Propagation from Video help Semantic Segmentation?

Semantic Video Segmentation

- Use video to propagate labels. *Pseudo Ground Truth (PGT)*
- Weakly-Supervised Learning CNN+CRF

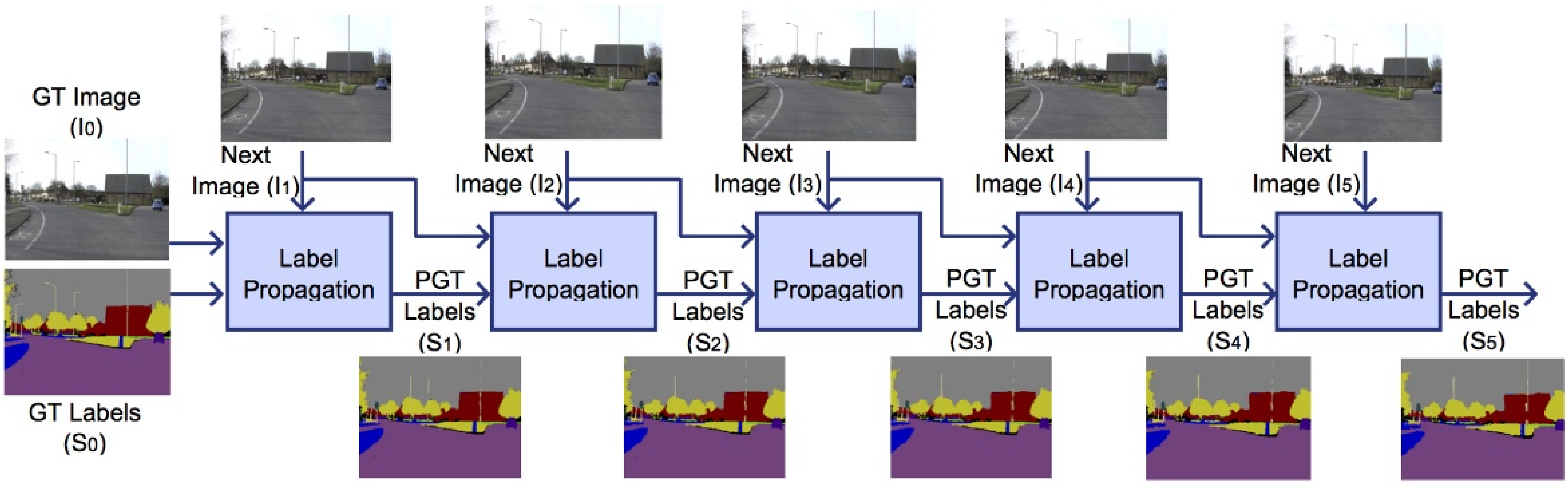
  ➢ Basic idea: given a few videos with limited labeled frames, we first estimate pseudo noisy ground truth for each frame in training set. Then we use all the labeled frames to train a CNN.



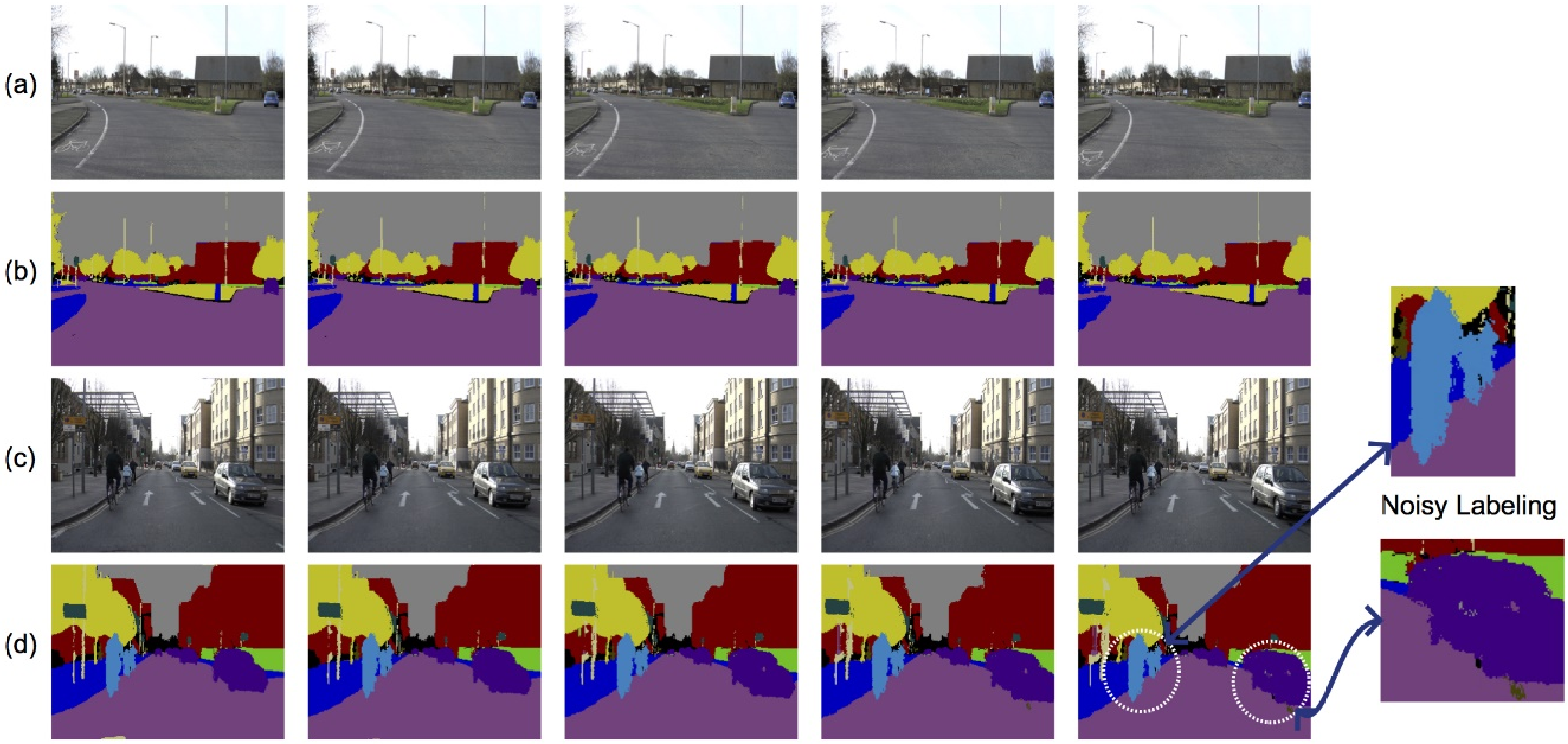
Mustikovela, Yang, Rother, ECCV Workshop 2016
Can Ground Truth Label Propagation from Video help Semantic Segmentation?

Generating Pseudo Ground Truth Data

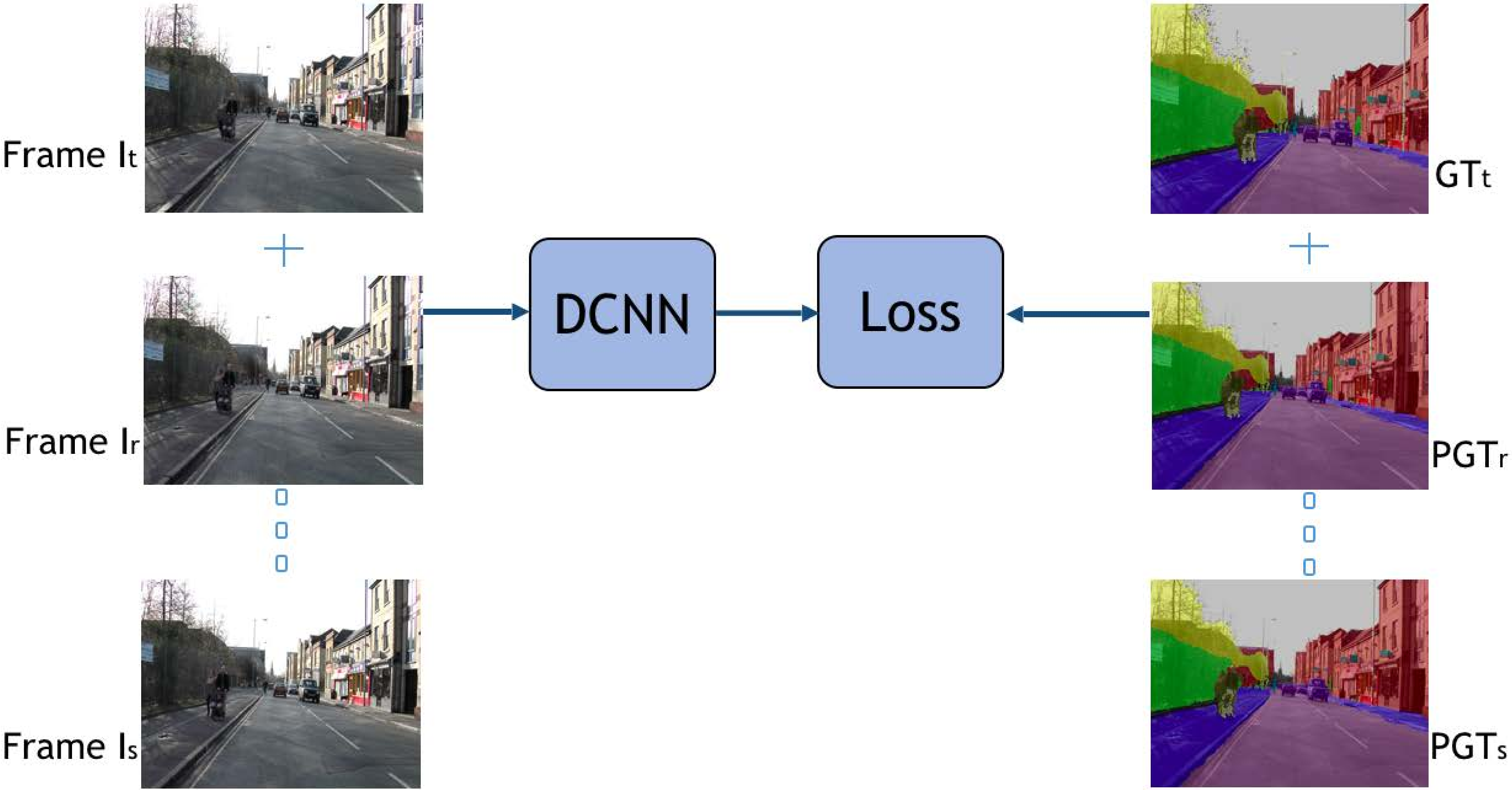
CRF for Label Propagation



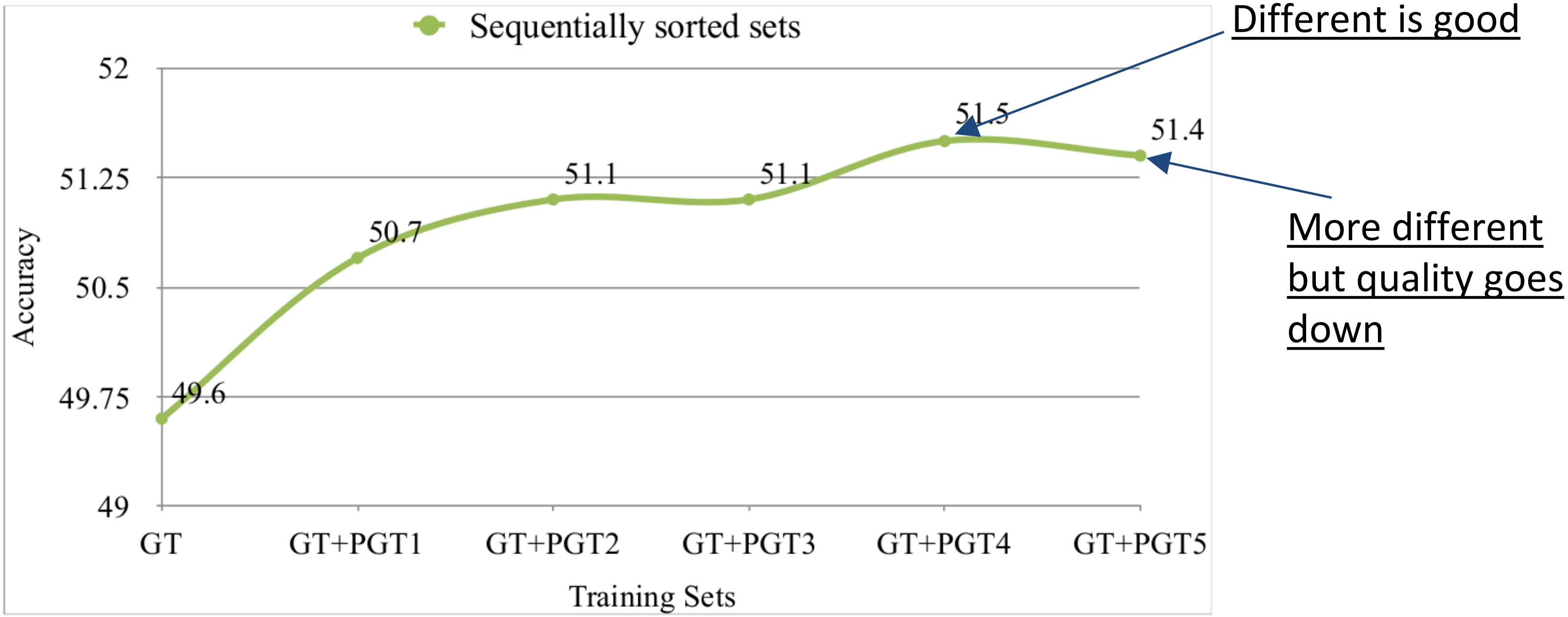
Quality of Pseudo Ground Truth Data



CNN Training



CamVid Results



- Image 4 is more different to GT than Image 1
- Quality of labeling of Image 5 might go down

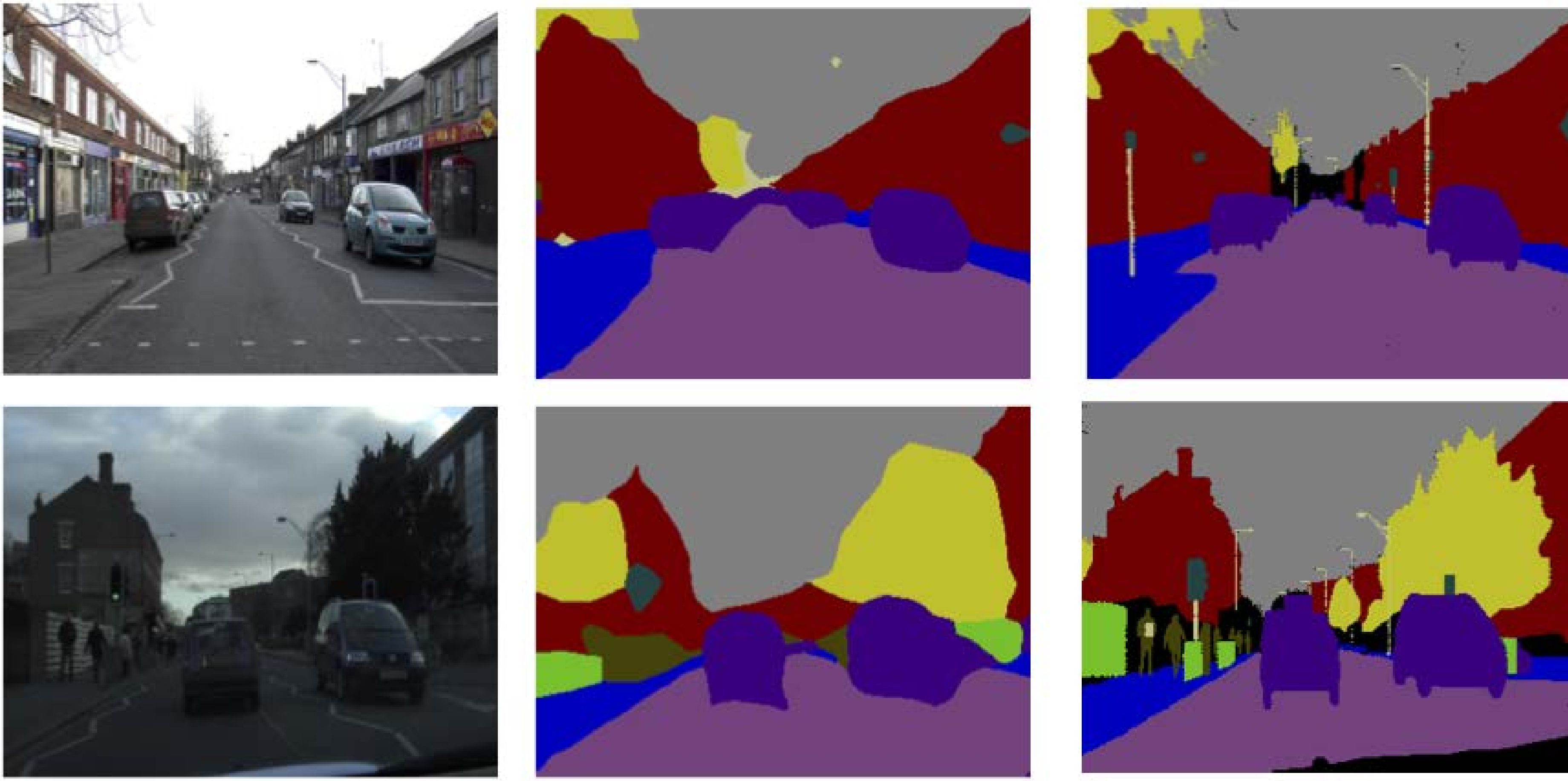
CamVid Results

- Model trained with GT + 4th images performs the best
- Performs better in 10/11 classes

Approach	Build ing	Tree	Sky	Car	Sign	Road	Pedes trian	Fence	Pole	Side walk	Bicy cle	Avg IoU
FCN (Only GT)	70.5	63.1	84.8	61.9	19.1	89.8	19.8	30.9	6.5	70.1	29.3	49.6
Ours (GT+PGT_S4 $t_f = 0.9$)	**72**	**65.6**	84.6	**64.6**	**20.8**	**90.6**	**24.9**	**38.8**	**8.0**	**71.8**	**33.9**	**52.3**

Semantic Video Segmentation

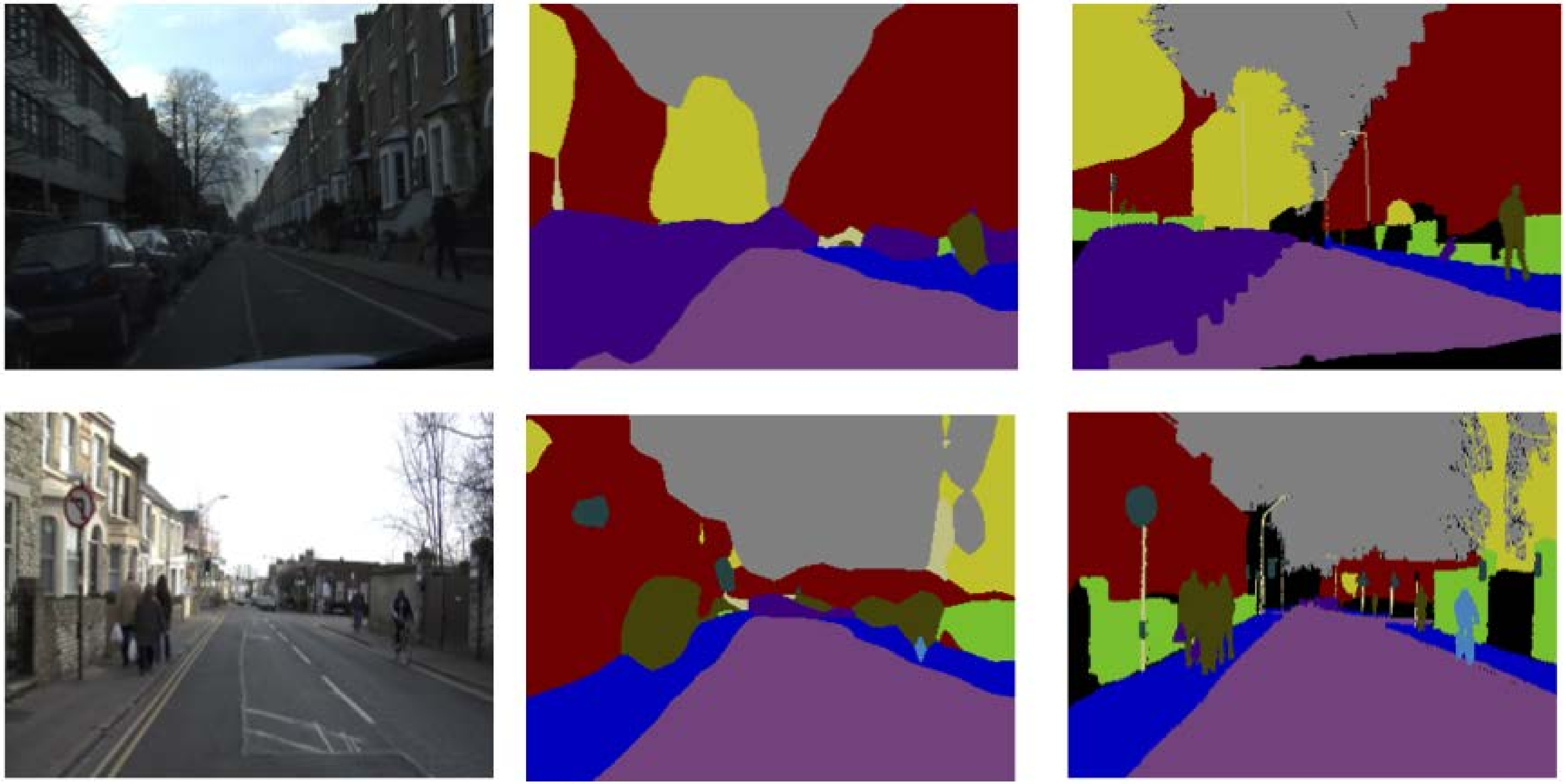
Results



Video frame | Our CNN result | ground-truth

Semantic Video Segmentation

Results



Video frame | Our CNN result | ground-truth

Outline

- **Introduction**

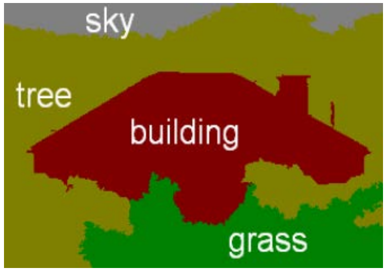
- **Random Fields**
  - **Man-made Object Segmentation**
  - **Semantic Video Segmentation**

Thank you!

ITC
University of Twente, NL

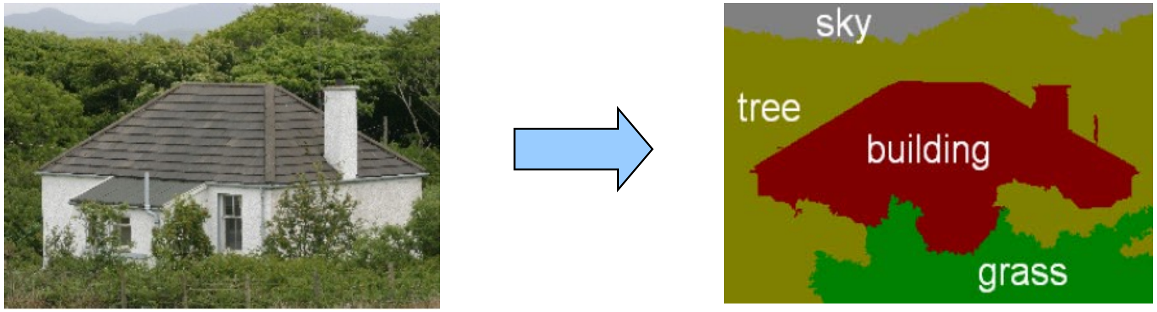
Image Labeling Problems

- Labelings highly structured
- Labels highly correlated with very complex dependencies



- Neighbouring pixels tend to take the same label
- Low number of connected components
- Classes present may be seen in one image
- Geometric / Location consistency
- Planarity in depth estimation
- … many others (task dependent)

Object-class Segmentation



$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j)$$

Unary term **Pairwise term**

Unary term

Discriminatively trained classifier (RF, Boosting, etc.)

Pairwise term

$$\psi_{ij}(x_i, x_j) = K_{ij}\delta(x_i \neq x_j)$$

where $K_{ij} = \lambda_1 + \lambda_2 \exp(-\beta(I_i - I_j)^2)$