Applications of Graphical Models

Dr. Michael Yang

September 12, 2016

Brief CV

- Since 2016, Assistant Prof., EOS-ITC
- 2015-2016, Senior Scientist, CVLD, TU Dresden
- 2012-2015, Postdoc, TNT, Leibniz University Hannover
- 2008-2011, Ph.D, Inst. Photogrammetry, Bonn University
- 2016-2020, Co-Chair ISPRS WG Dynamic Scene Analysis
- Main Research Areas: Photogrammetry, Computer Vision

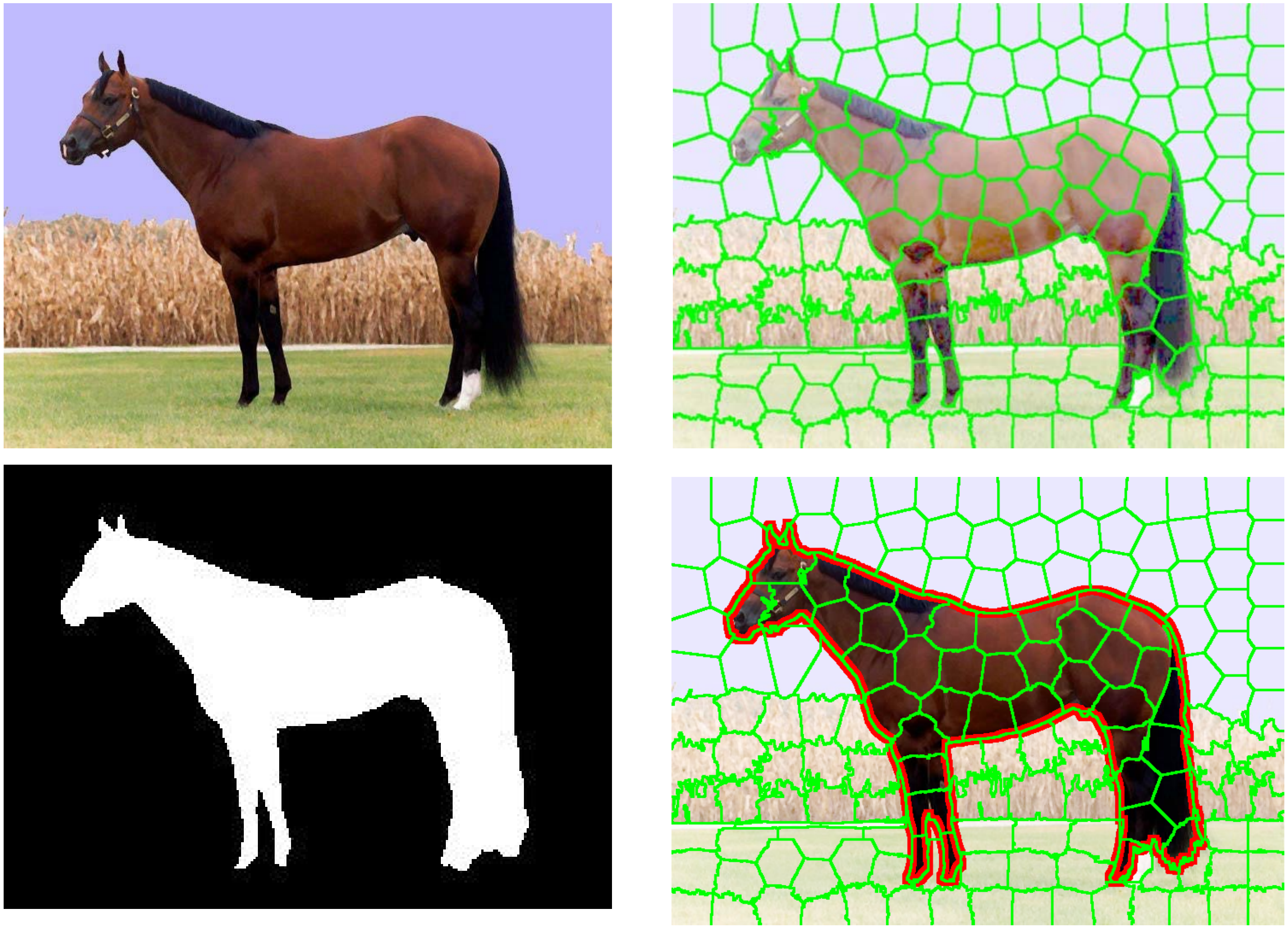
Outline

- **Introduction**
- **Random Fields**
- **Future**

Applications

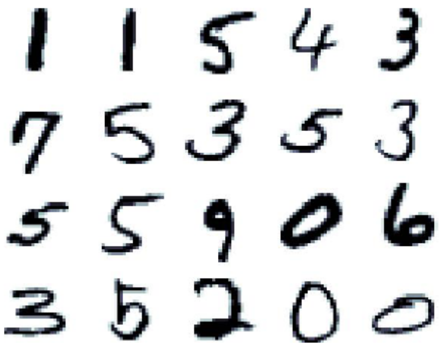
- Medical diagnosis
- Social network models
- Speech recognition
- Robot localization
- Remote sensing
- Natural language processing
- Computer vision
  - Image segmentation
  - Tracking
  - Scene understanding
- Photogrammetry
  - Image classification
  - 3D reconstruction
  - 3D urban modeling
- ....

Segmentation

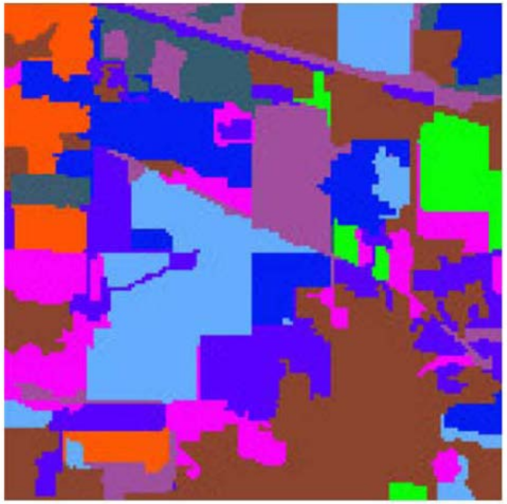


Yang, Rosenhahn, 2016

Classification



(MNIST benchmark data)

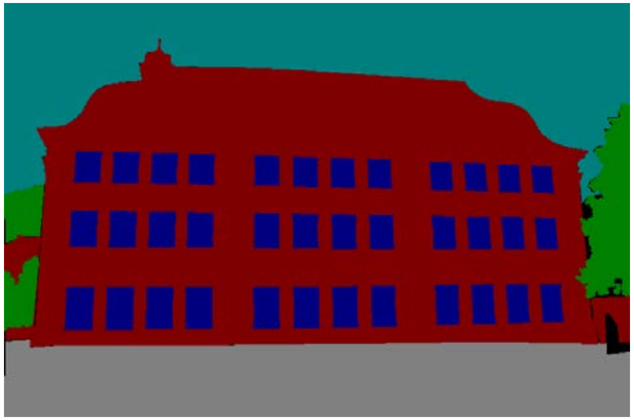


Zhong & Wang 2011

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

Interpretation

Chai et al., 2013

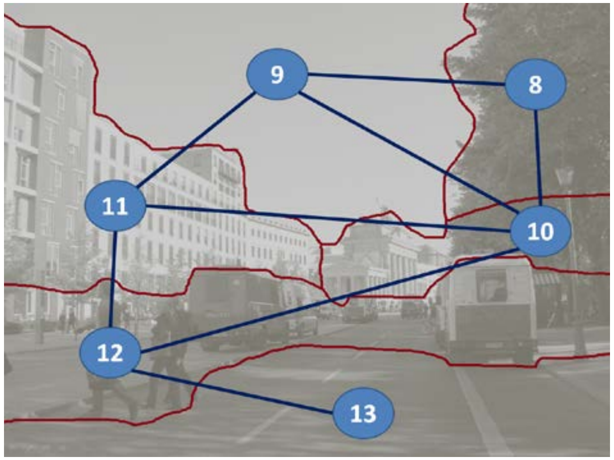


Yang & Förstner, 2011

- Building and road extraction
- Facade interpretation
- Traffic scene interpretation
- Holistic scene analysis

Interpretation

Barth et al., 2010



Yang, 2015

- Building and road extraction
- Facade interpretation
- Traffic scene interpretation
- Holistic scene analysis

Probabilistic Graphical Models

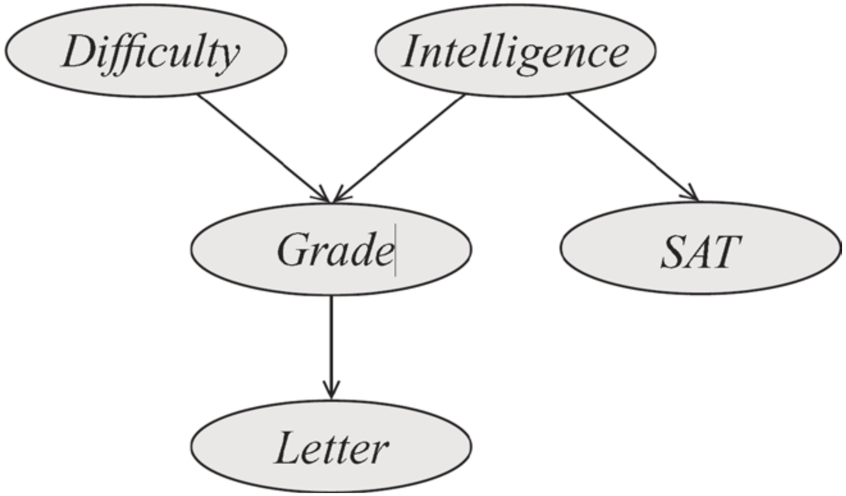
are a marriage between

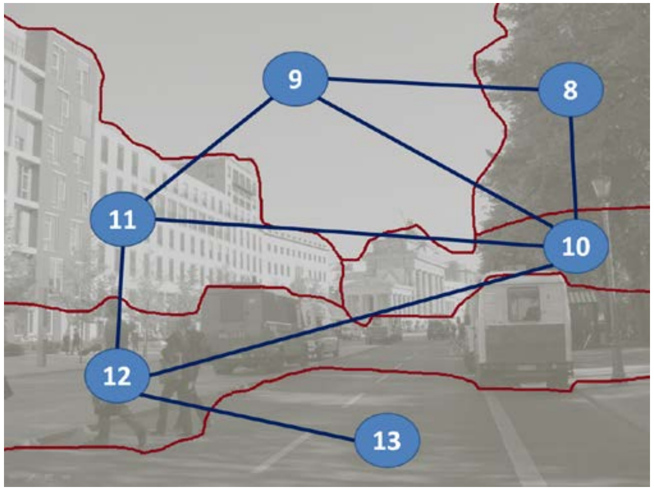
probability theory & graph theory

Graphical Models

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 $\mathcal{G}$

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

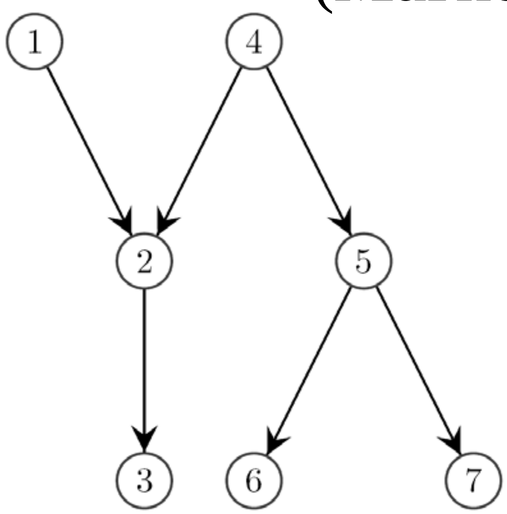
- Nodes $i \in \mathcal{V}$ represent random variables $\underline{\mathbf{x}}_i$
- Edges represent mutual relationships
  - ➢ Undirected edges $\{i, j\}$ model joint probabilities

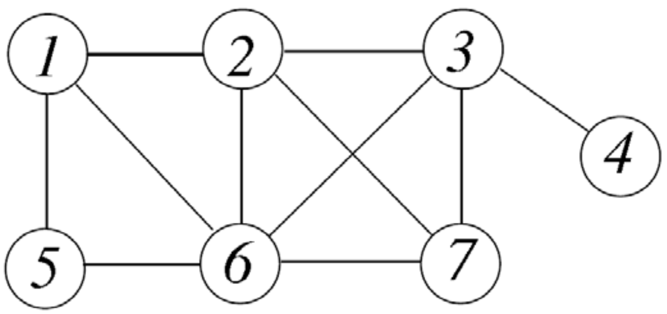
$$P(\mathbf{x}_i, \mathbf{x}_j)$$

  - ➢ 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
- **Conditional/Markov Random Fields**
- Future

Applications

Photogrammetry/CV:

- 2D/3D Image Segmentation
- Object Recognition
- 3D Reconstruction
- Stereo / Optical Flow
- Image Denoising
- Texture Synthesis
- Pose Estimation
- Panoramic Stitching
- …

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

- **Joint distribution** *(Hammersley & Cliord, 1971)*

  If positive distribution and Markov property:

  Markov random field $\iff$ Gibbs random field

$$P(\mathbf{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \phi_c(\mathbf{x}_c)$$

Potential functions referring to maximal cliques $c \in \mathcal{C}$

$$\phi_c(\mathbf{x}_c) > 0$$

Partition function, normalization constant

$$Z = \sum_{\mathbf{x}} \prod_{c \in \mathcal{C}} \phi_c(\mathbf{x}_c)$$

Sum over *all states* the complete Markov field!

MRFs

- **Equivalent representation of distribution in MRF**

  If positive distribution and Markov property:

  Markov random field $\iff$ Gibbs random field

$$P(\mathbf{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \phi_c(\mathbf{x}_c) = \frac{1}{Z} \exp(-E(\mathbf{x}))$$

Energy

$$E(\mathbf{x}) = \sum_{c \in \mathcal{C}} \varphi_c(\mathbf{x}_c)$$

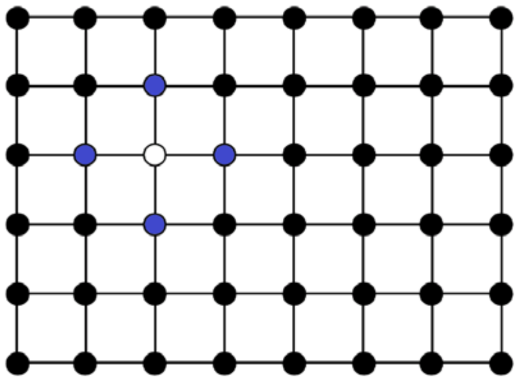
- **Choice of potential functions**

  Need not be probabilities

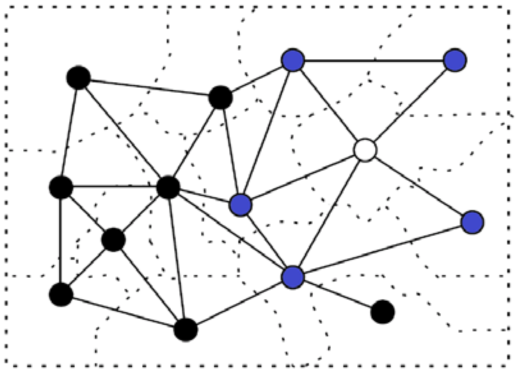
MRFs

- **Structure of MRFs**

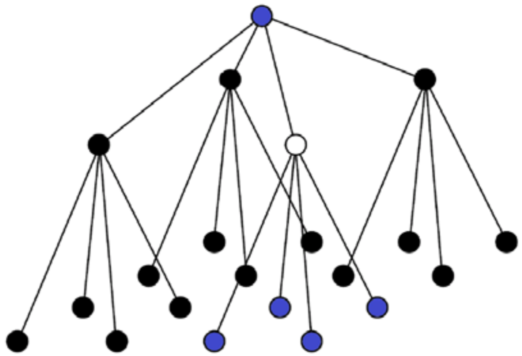
Typical graph structures



rectangular grid



irregular graph



pyramid structure

Figure courtesy of P. Perez

MRFs

- **Pairwise MRFs**

  popular

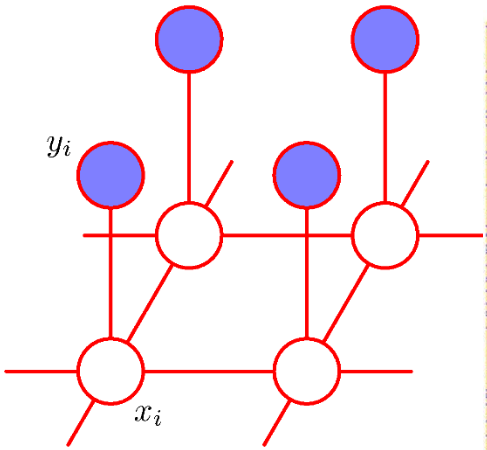
$$P(\mathbf{x}) = \frac{1}{Z}\exp(-E(\mathbf{x}))$$

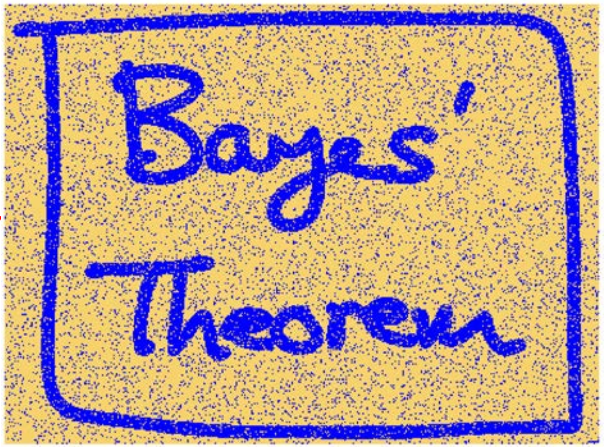
with energy function

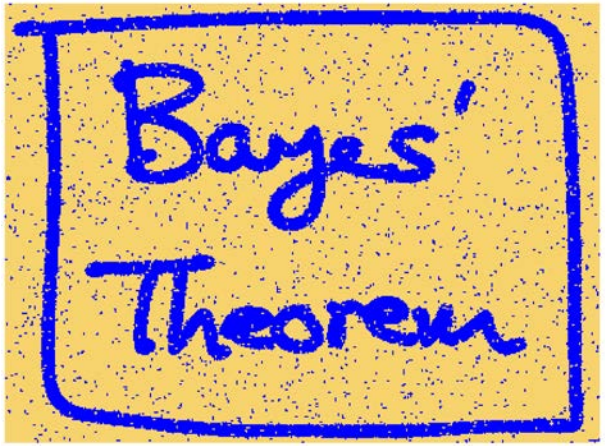
$$E = \sum_{i\in\mathcal{V}} \underbrace{E_1(x_i)}_{Unary} + \alpha \sum_{\{i,j\}\in\mathcal{N}} \underbrace{E_2(x_i, x_j)}_{Pairwise}$$

MRFs

- **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

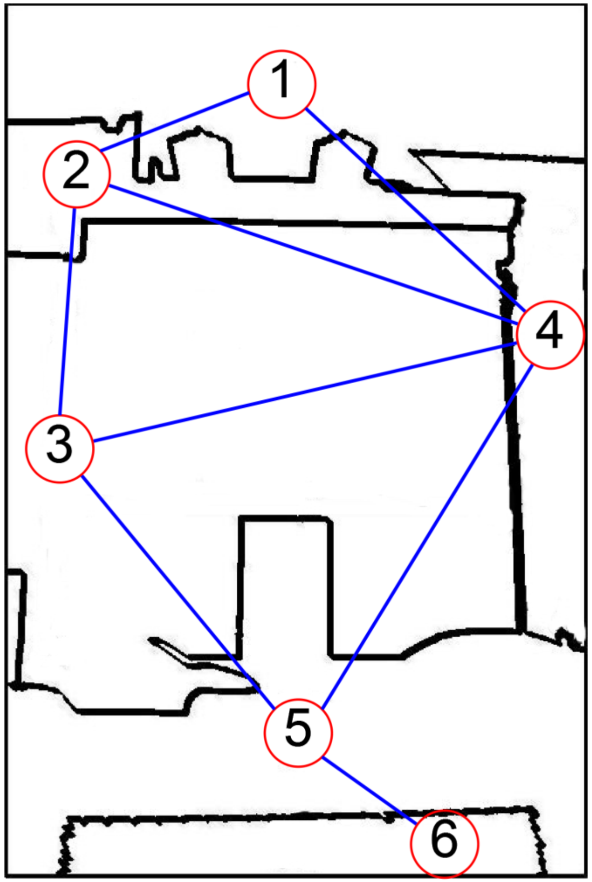
Conditional distribution

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

CRFs

Yang & Förstner, 2011

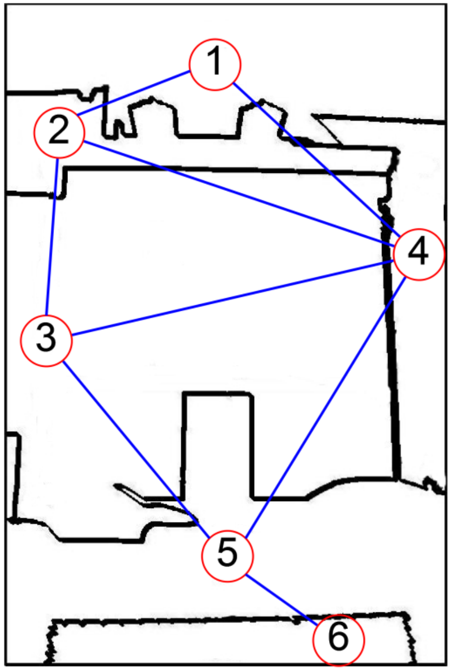
Building facade image



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*)

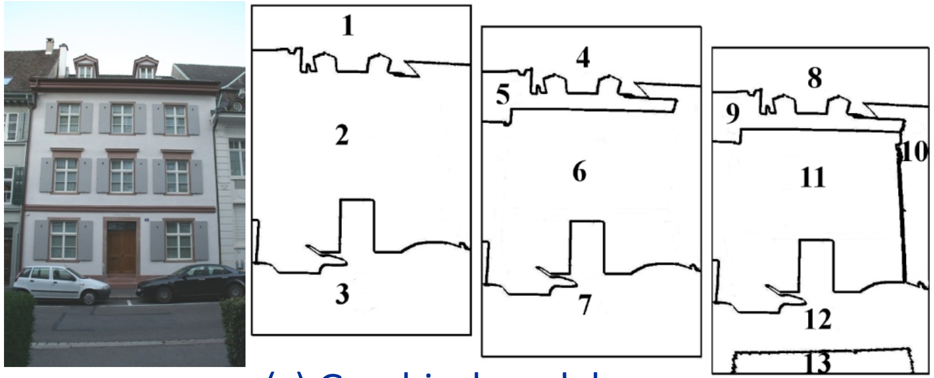
$$E = \sum_{i \in \mathcal{V}} \underbrace{E_1(x_i)}_{Unary} + \alpha \sum_{\{i,j\} \in \mathcal{N}} \underbrace{E_2(x_i, x_j)}_{Pairwise}$$

Hierarchical CRFs

Yang & Förstner, 2011

(a) Test image

(b) Multi-scale segmentation



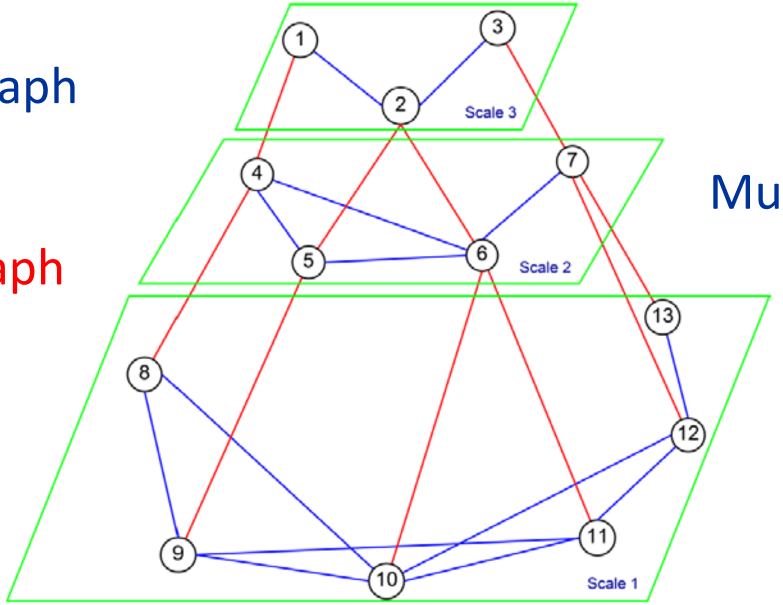
(c) Graphical model

Region adjacency graph

Blue edges

Region hierarchy graph

Red edges



Multi-layer CRF

Hierarchical CRFs

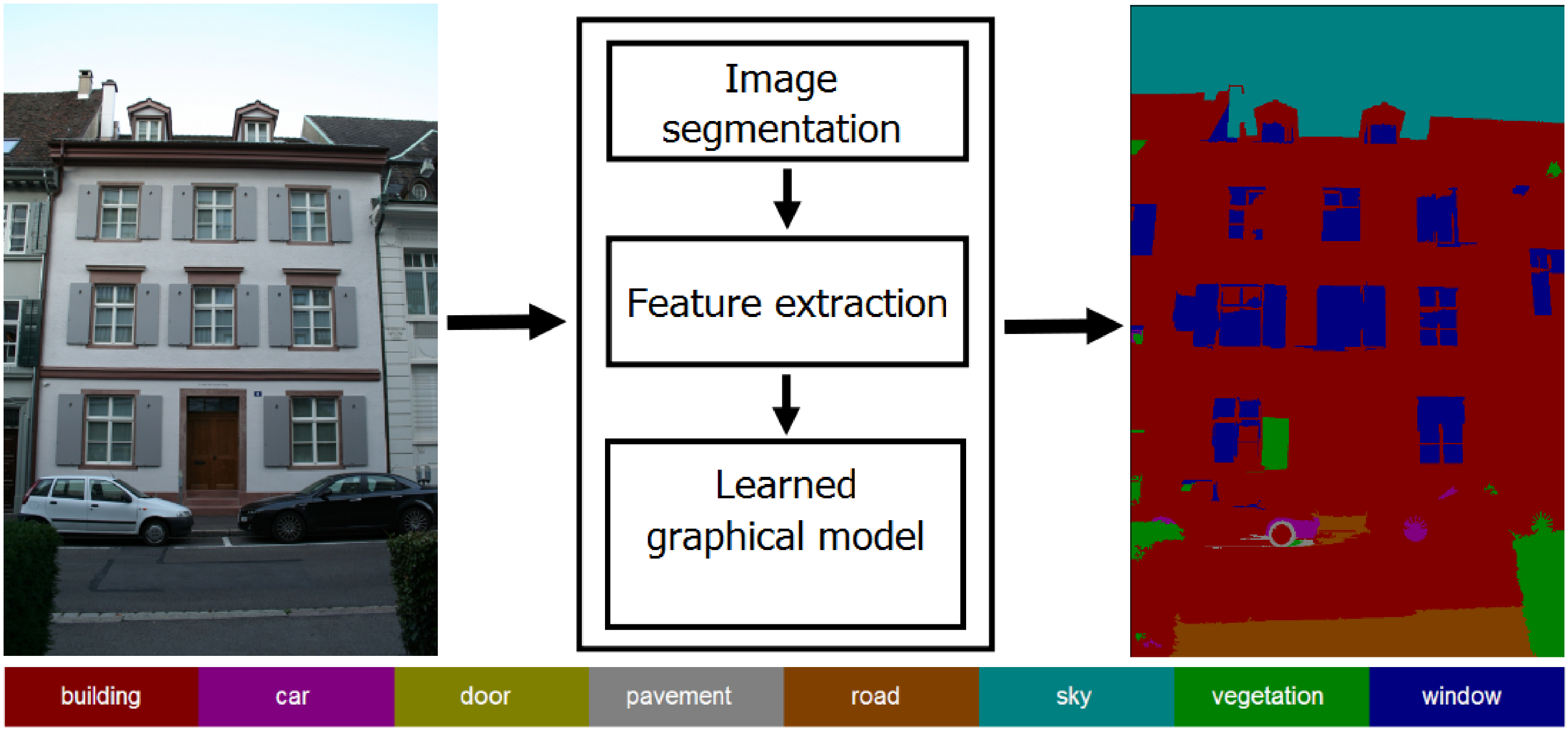
Energy function

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

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

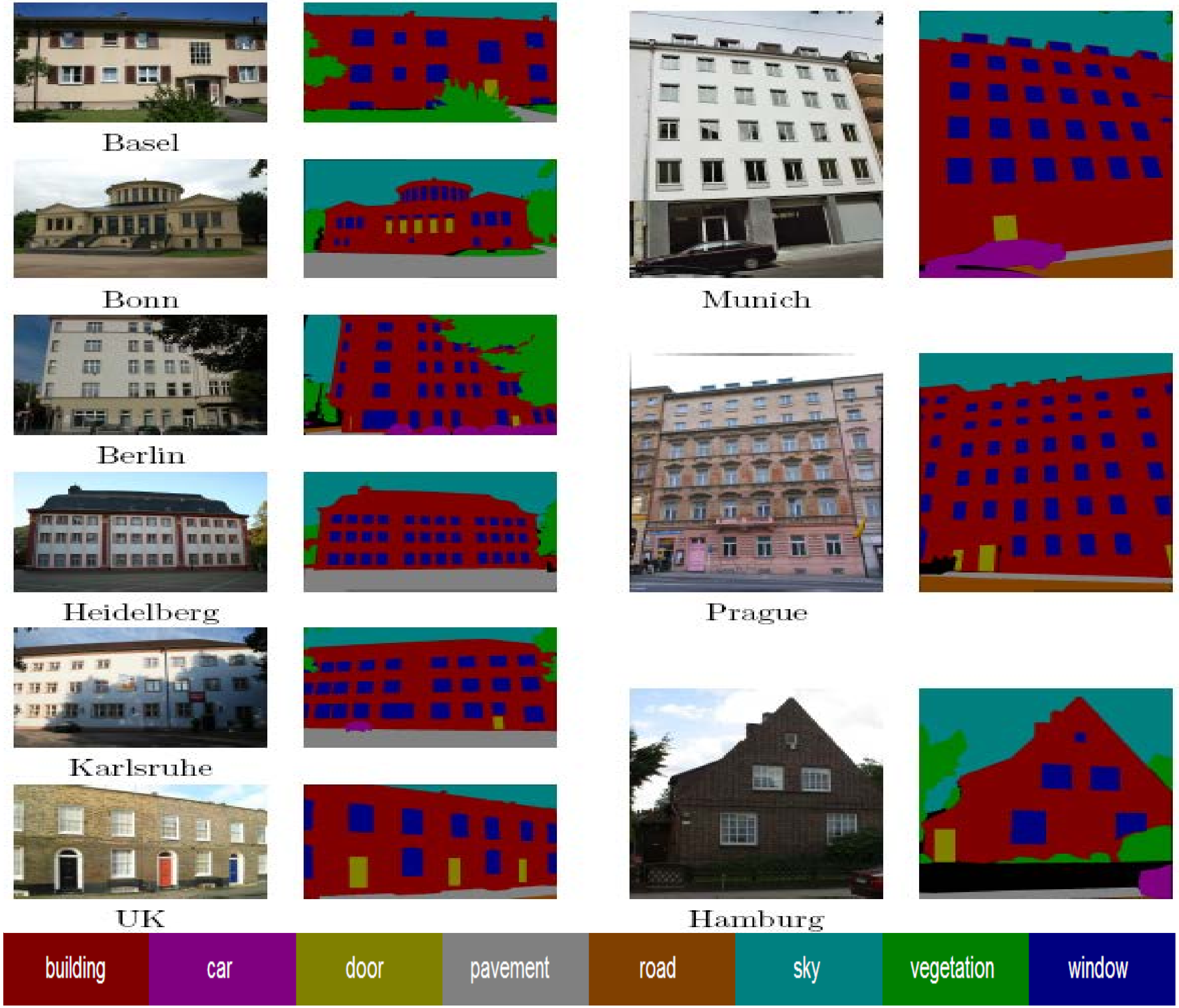
Scene Interpretation

Framework



Workflow for image interpretation of man-made scenes

ETRIMS Database



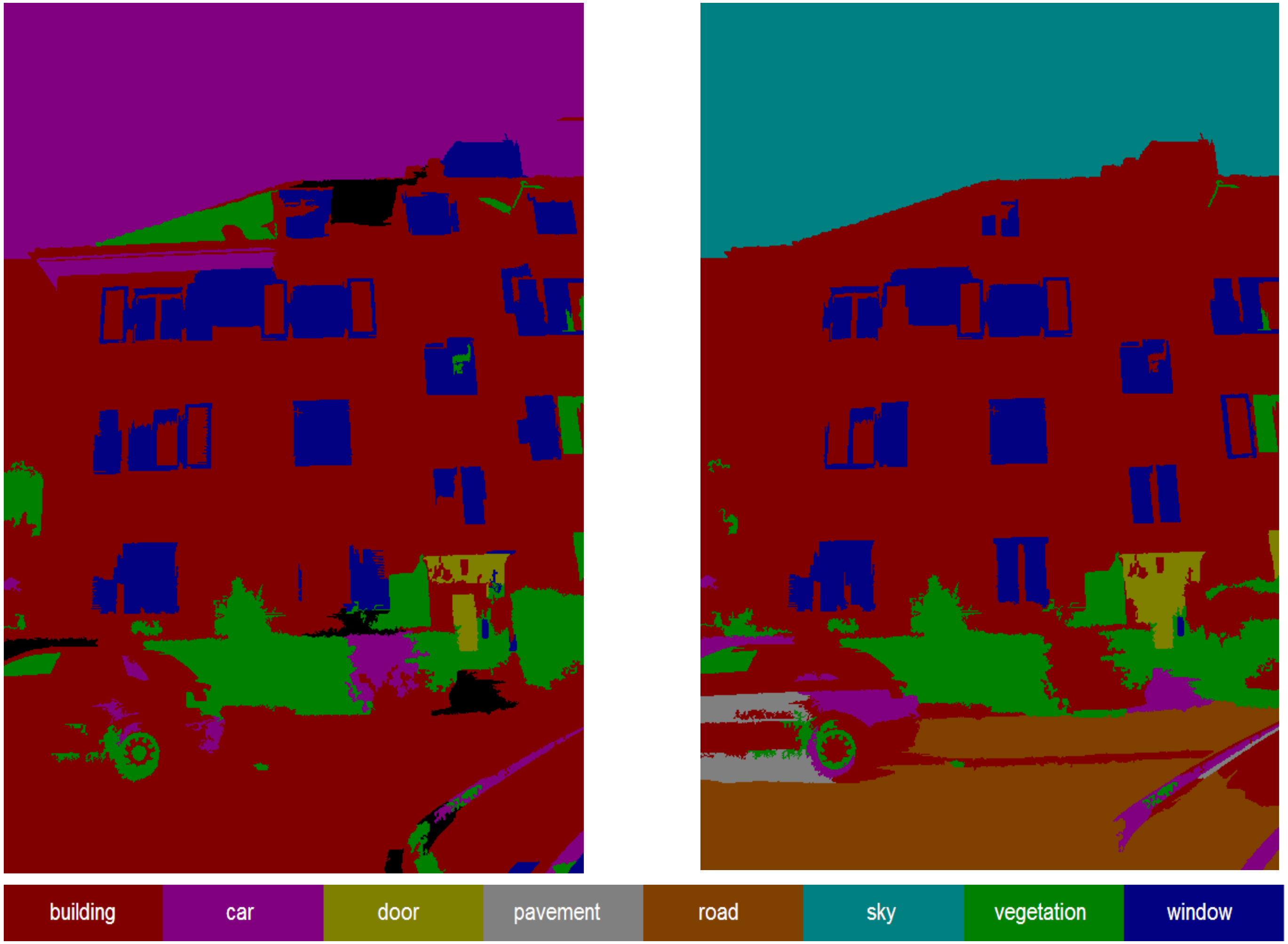
Example Image

building | car | door | pavement | road | sky | vegetation | window

One example image

Ground truth labeling

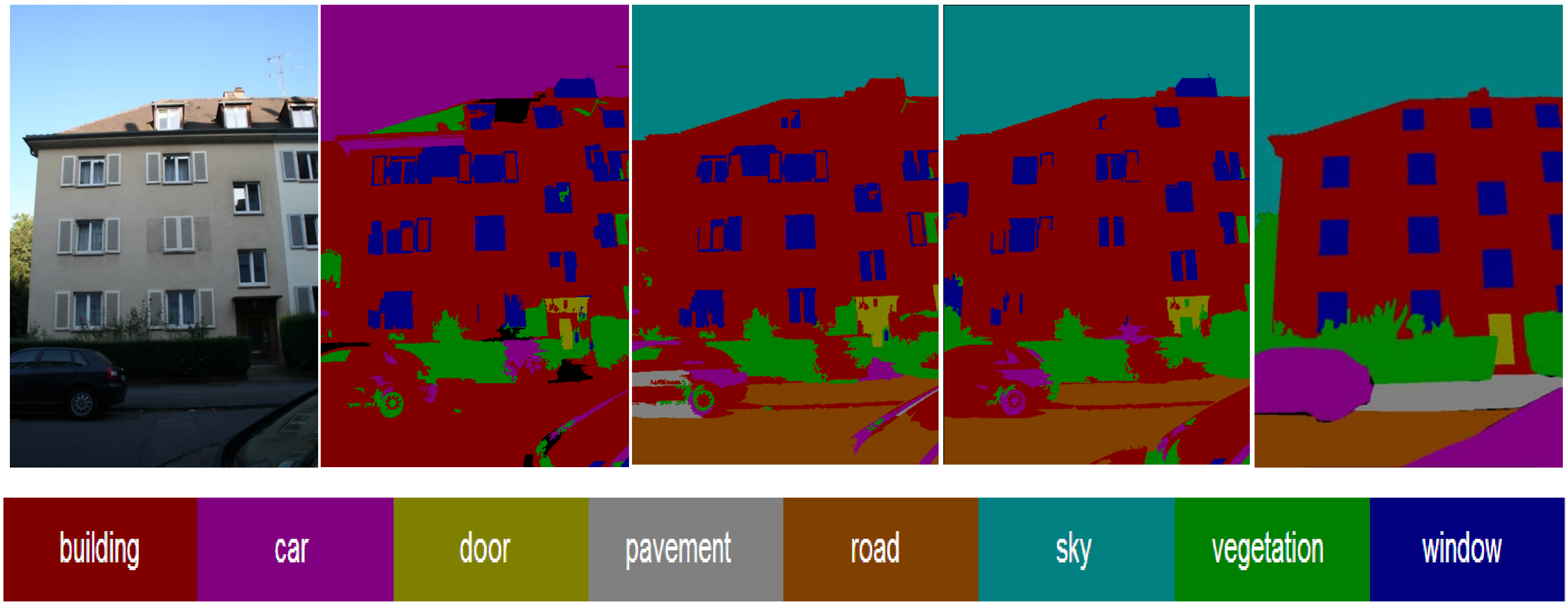
Classification Results



Region classifier (RDF)   Pairwise CRF

HCRF Results

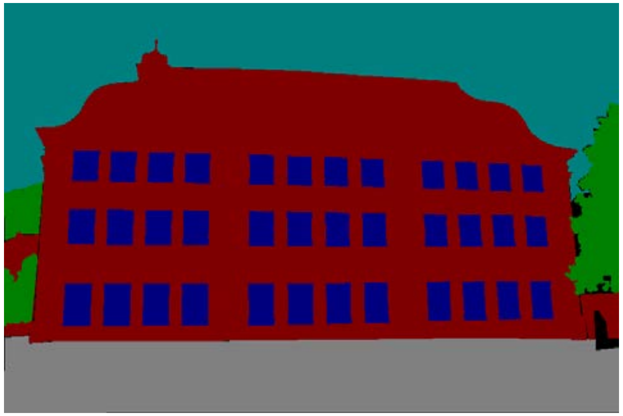
Image | RDF | CRF | HCRF | GT



HCRF Results

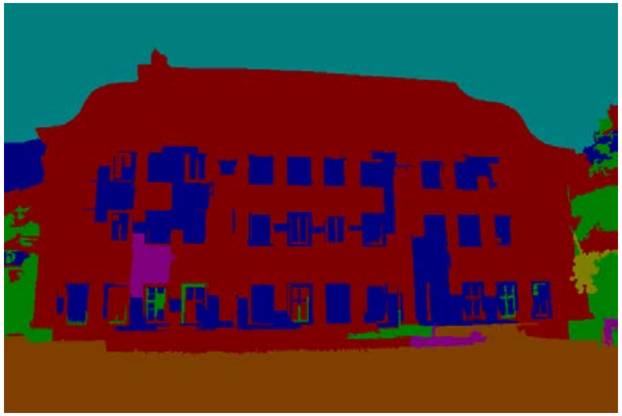
Image

GT

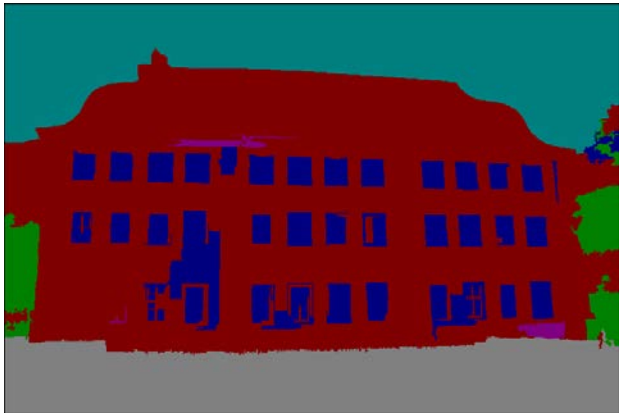


RDF

CRF



HCRF

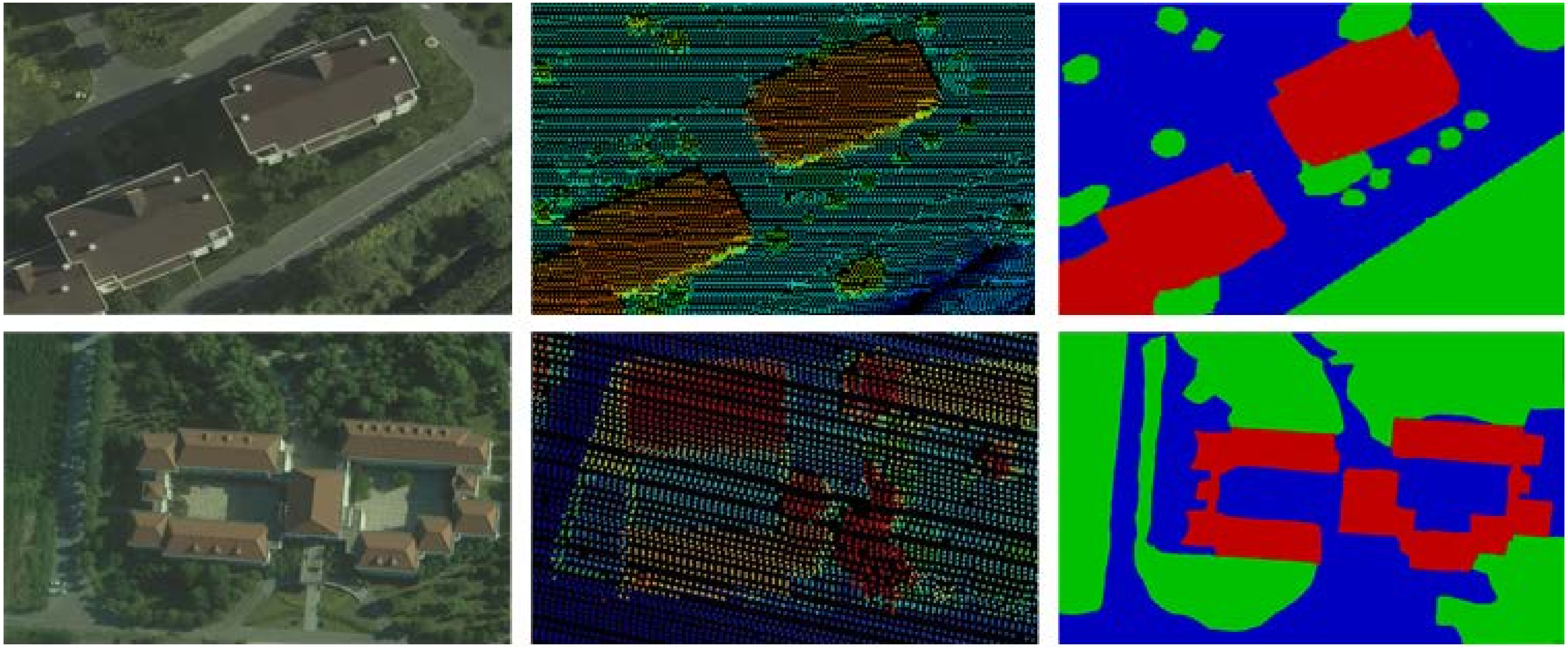


HCRF Results

Pixelwise accuracy comparison

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

CRF for Sensor Fusion



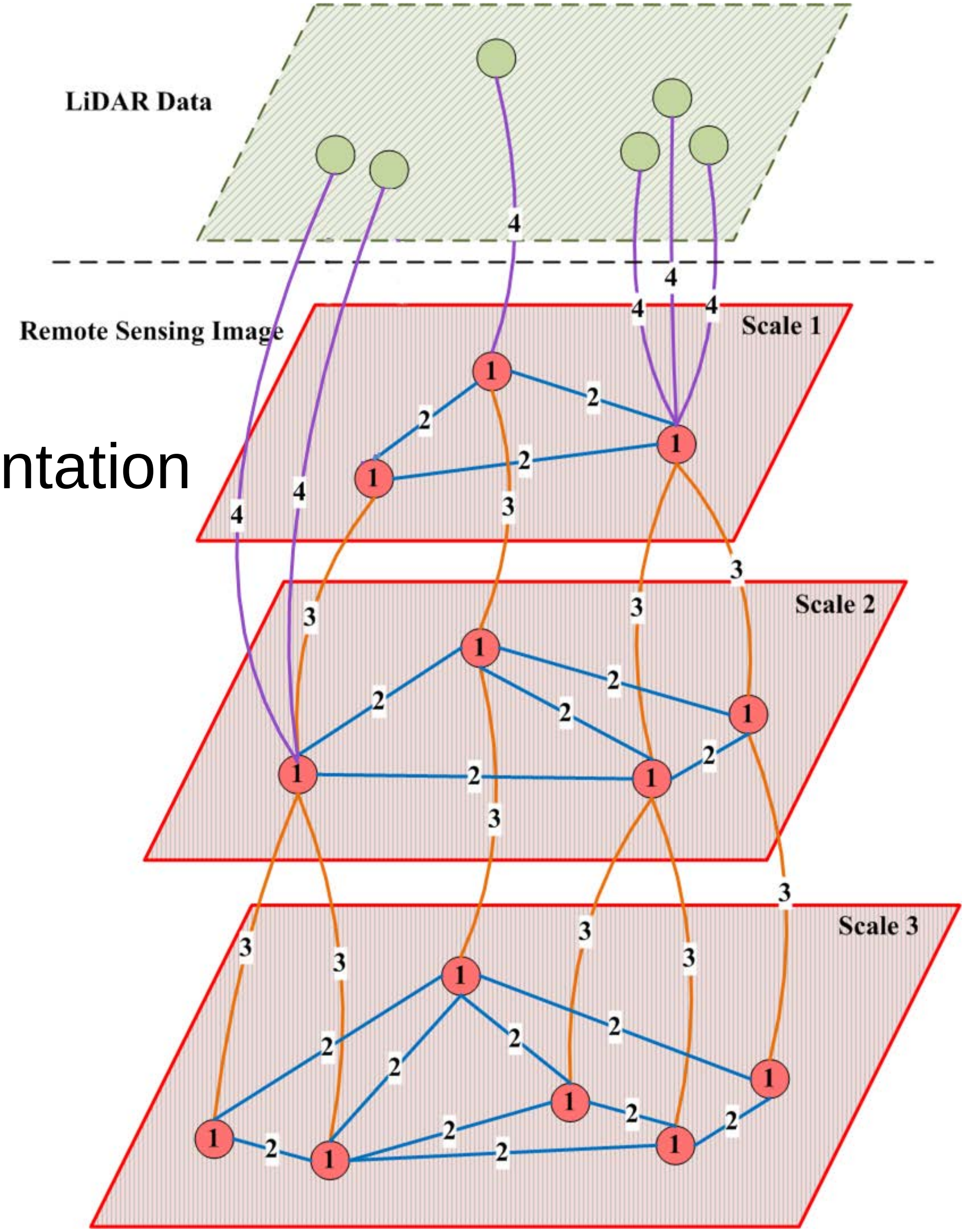
Zhang, Yang, Zhou, 2015

Multi-sensor fusion

- Optical image
- Lidar data

Graph Construction

- ➢ Lidar level
- ➢ Image level
  - ➢ Multi-scale segmentation



MSMSHCRF Model

The conditional probability of the class labels x given an image d and Lidar data L

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

MSMSHCRF Model

Energy function

$$E = \sum_{i \in \mathcal{V}} \underbrace{E_1(x_i)}_{Unary} + \alpha \sum_{\{i,j\} \in \mathcal{N}} \underbrace{E_2(x_i, x_j)}_{Pairwise}$$
$$+ \beta \sum_{\{i,k\} \in \mathcal{S}} \underbrace{E_3(x_i, x_k)}_{scaleHierar.} + \gamma \sum_{\{i,t\} \in \mathcal{M}} \underbrace{E_4(x_i, x_t)}_{sourceHierar.}$$

MSMSHCRF Model

Energy function

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

$$+ \beta \sum_{\{i,k\} \in \mathcal{S}} \underbrace{E_3(x_i, x_k)}_{scaleHierar.} + \gamma \sum_{\{i,t\} \in \mathcal{M}} \underbrace{E_4(x_i, x_t)}_{sourceHierar.}$$

- E1: Unary potentials
  relation between class labels and image
- E2: Pairwise potentials
  relation between class labels of neighboring regions within each scale

MSMSHCRF Model

Energy function

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

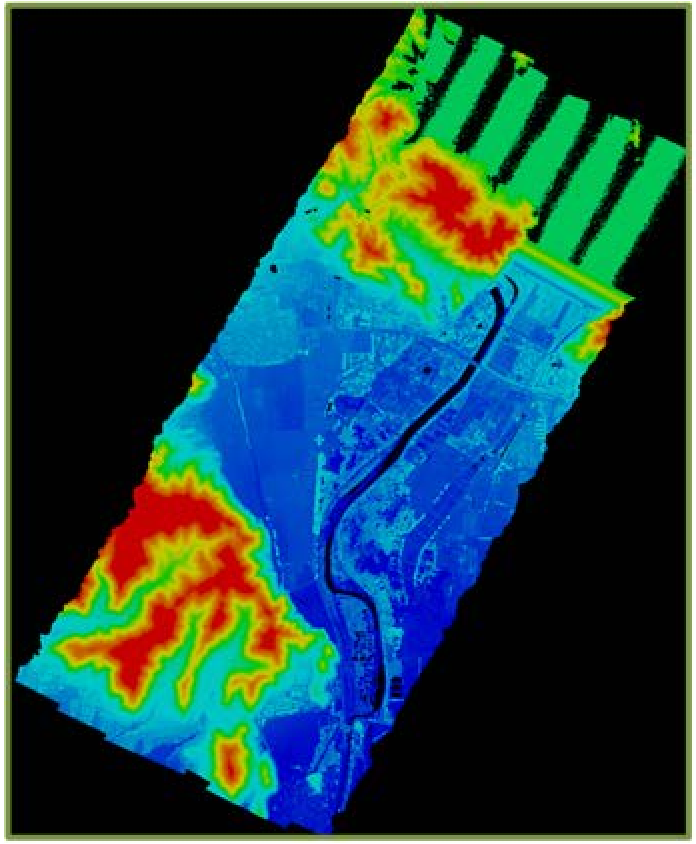
$$+ \beta \sum_{\{i,k\} \in \mathcal{S}} \underbrace{E_3(x_i, x_k)}_{scale Hierar.} + \gamma \sum_{\{i,t\} \in \mathcal{M}} \underbrace{E_4(x_i, x_t)}_{source Hierar.}$$

- E3: Multi-Scale hierarchical pairwise potential
  relation between regions in neighboring scales of images

- E4: Multi-Source hierarchical pairwise potential
  relation between image and Lidar data

Results

Dataset: Beijing Airborne Data



3 classes: {Building, Road, Vegetation}

50 images for training / 50 images for testing

Results

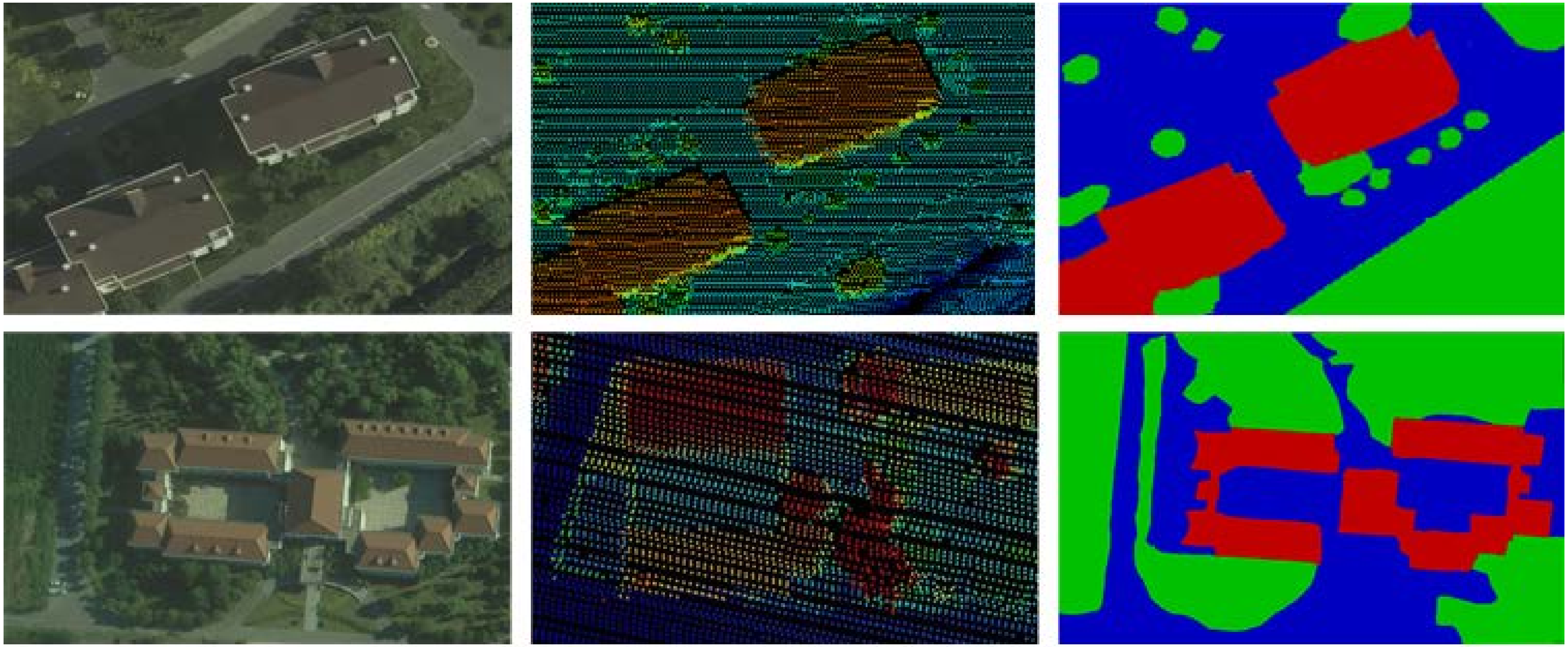


Image Lidar Classification result
(red - building, blue - road, green – vegetation)

Results

Comparison

Method	Accuracy (%)
Standard CRF	64.2
Hierarchical CRF	70.3
Multi-Source CRF	73.6
MSMSCRF	83.7

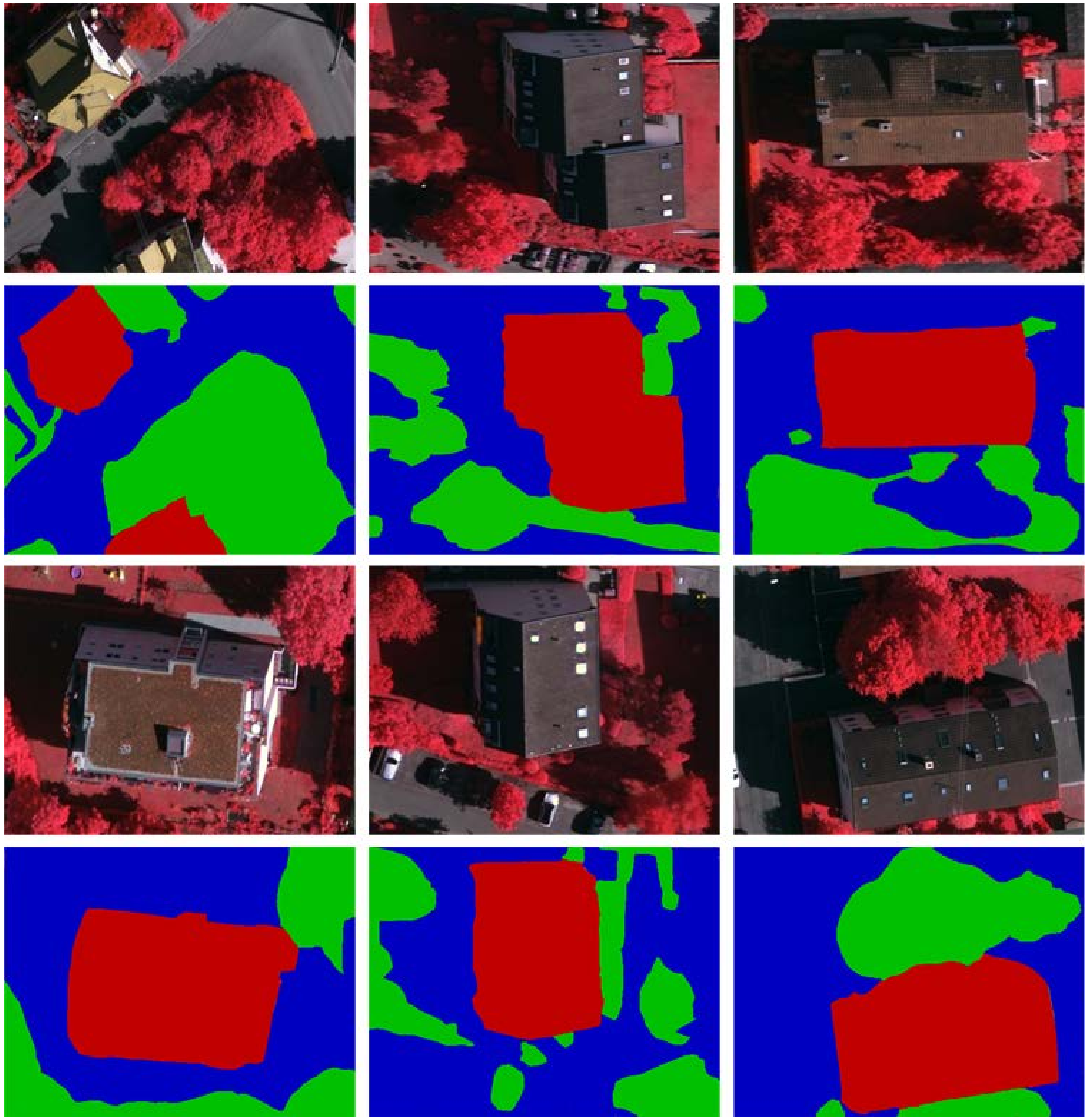
Results

Confusion Matrix

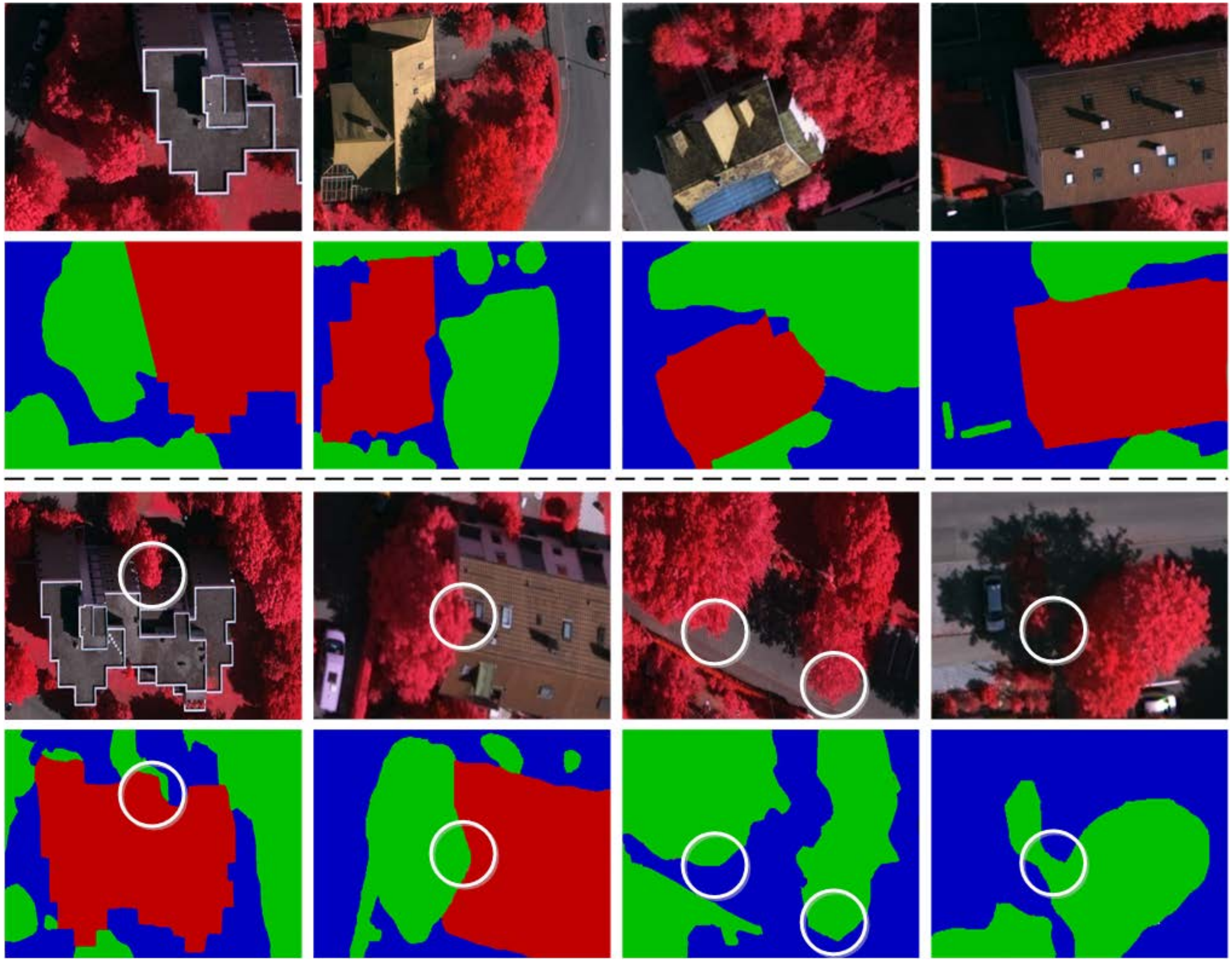
	building	road	vegetation
building	78.3	11.9	9.8
road	9.5	85.9	4.6
vegetation	9.7	8.7	81.6

Results

Dataset: ISPRS Benchmark

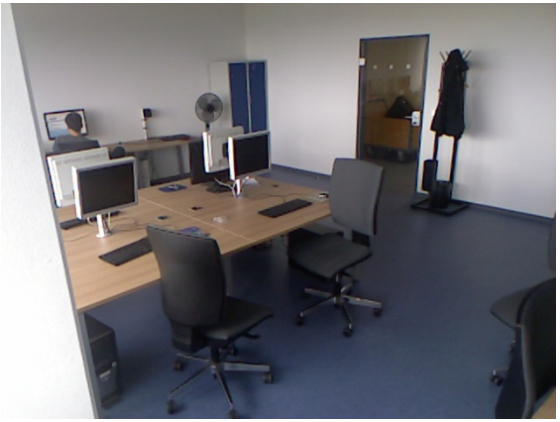


Dataset: ISPRS Benchmark

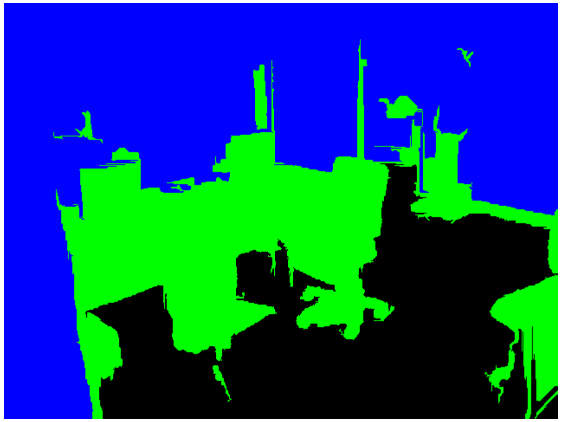


Layout Estimation

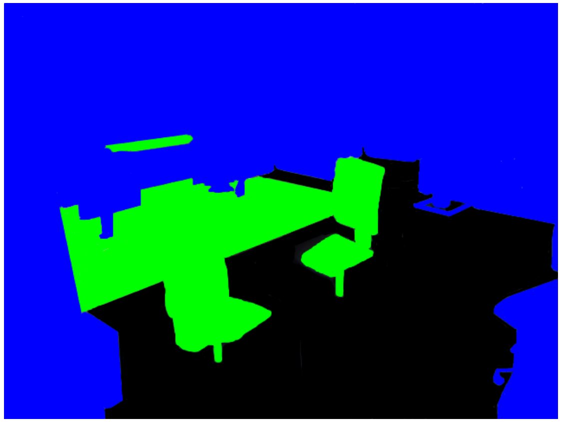
- CRF: fuse RGB and depth



Image+depth



Object/Layout



Ground truth

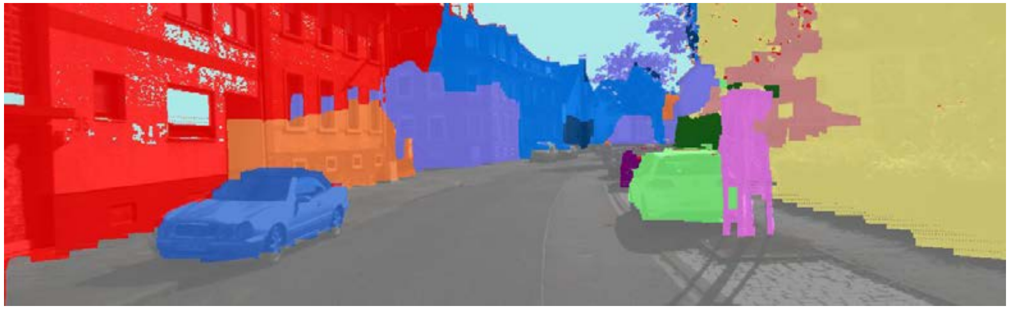
{sitting place, ground floor, background}

- Object/Layout

Shoaib, Yang, Rosenhahn, Ostermann, 2014

Object Segmentation


Single Image

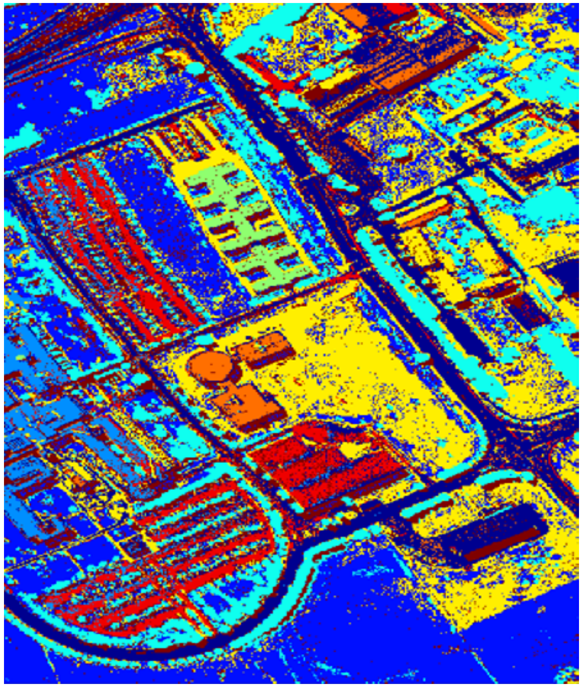

Object Class


Depth Upsampling

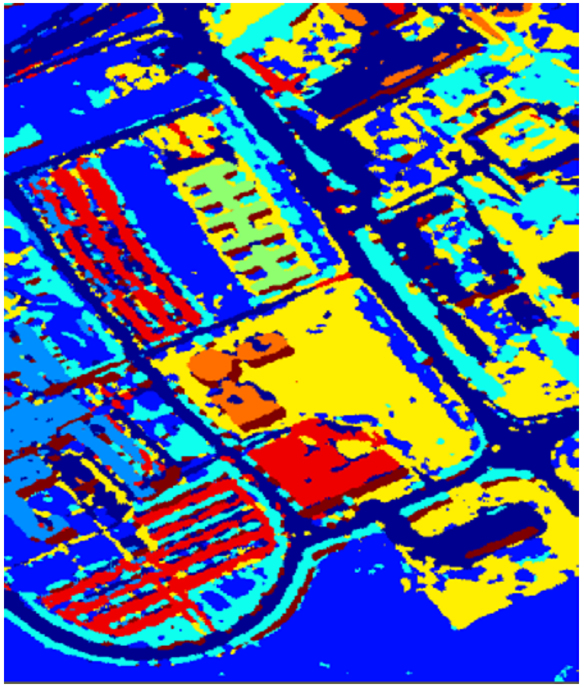
Huang, Gong, Yang, 2015

Hyperspectral Image Classification

Image



GP result



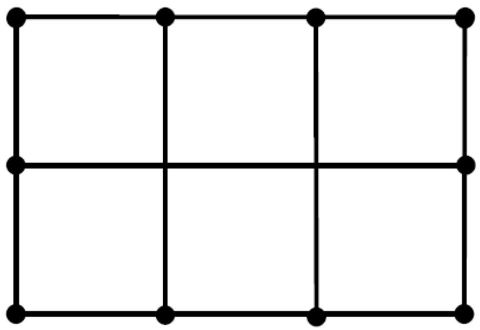
GP-MRF result

Liao, Tang, Rosenhahn, Yang, 2015

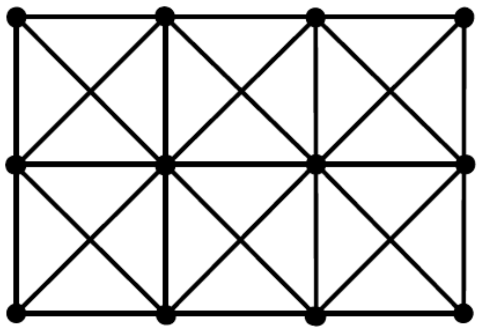
Outline

- Introduction
- Random Fields
- **Future**

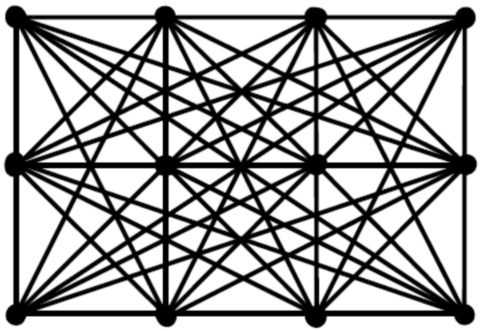
Fully Connected CRF



4-connected CRF



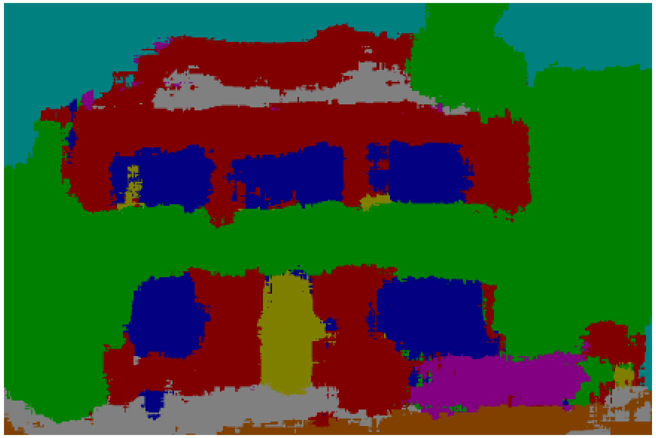
8-connected CRF



Fully-connected CRF

Fully Connected CRF

Image



Unary

Final

Li, Yang, 2016

Fully Connected CRF

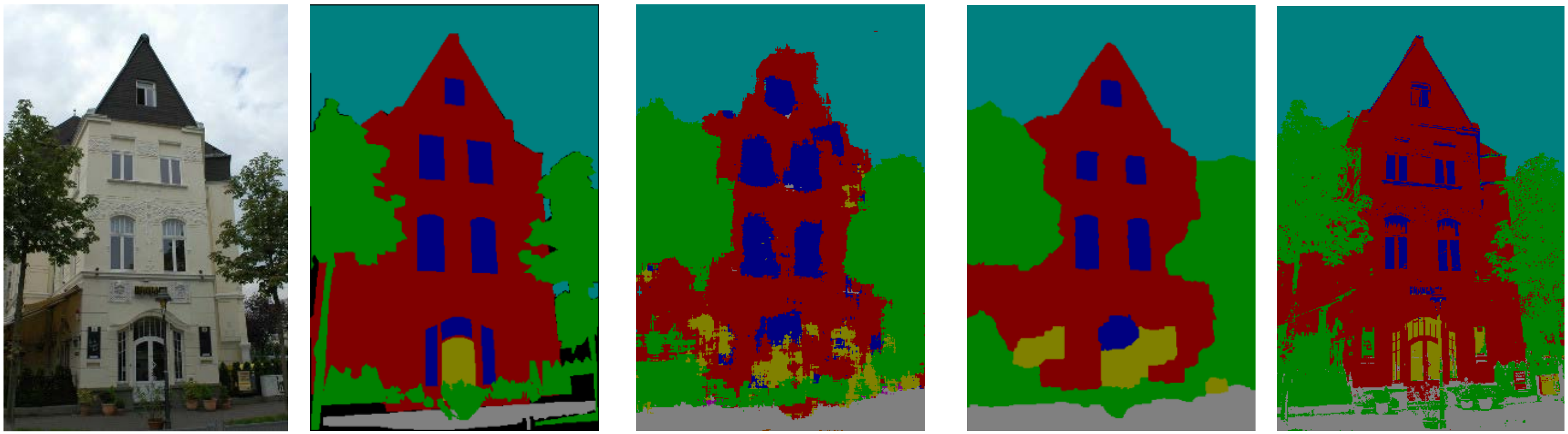
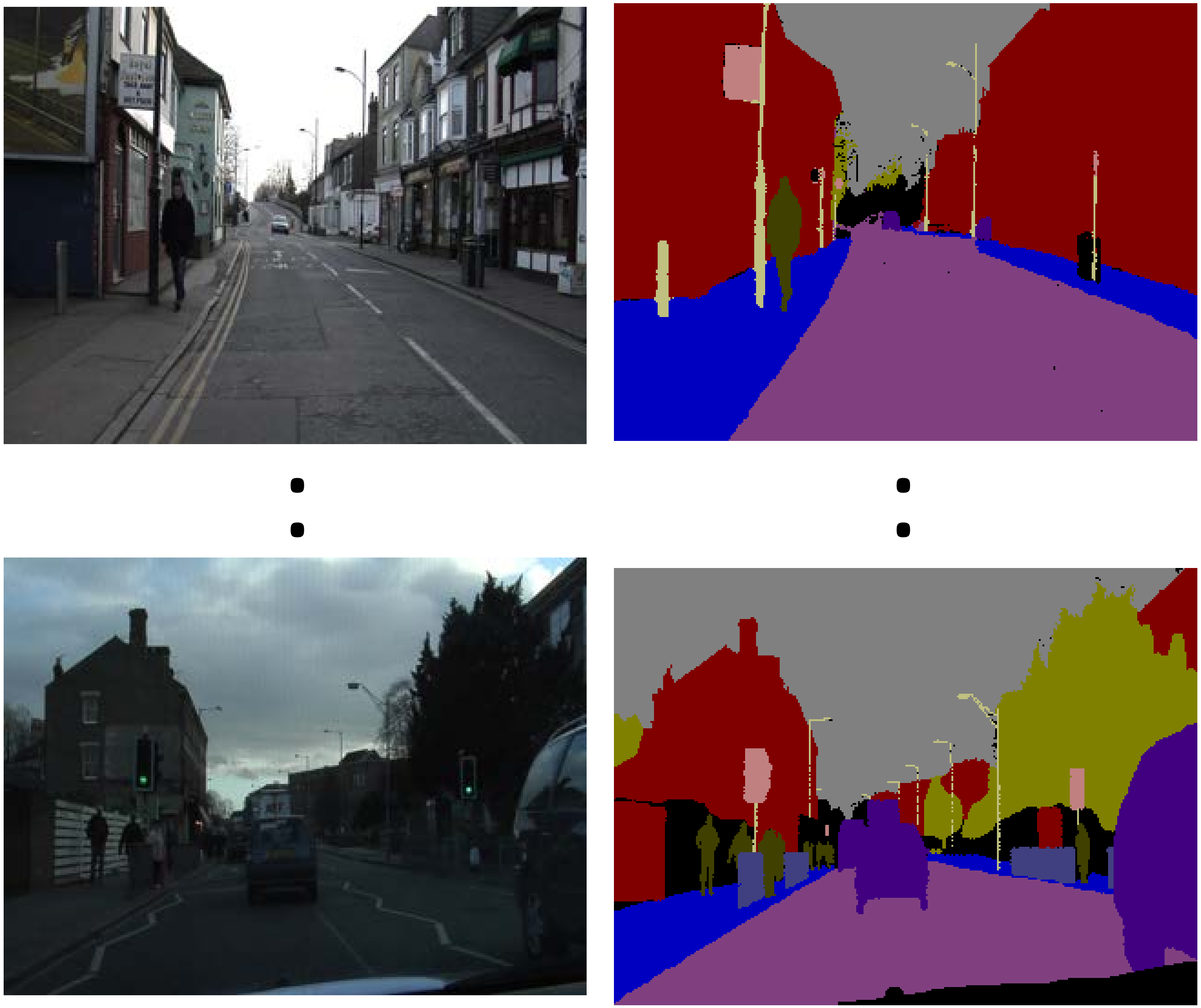
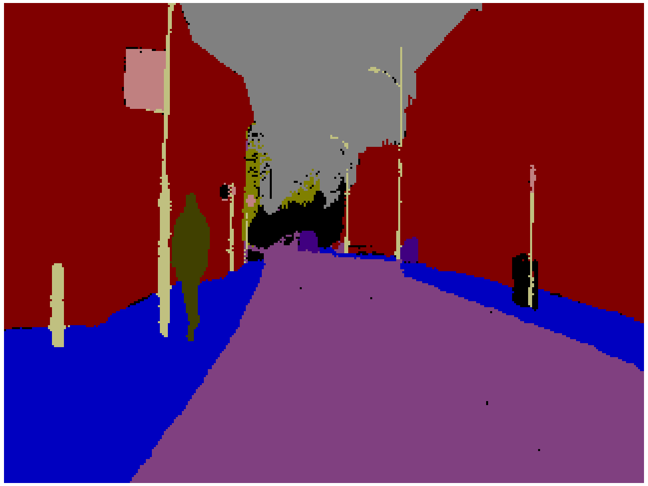


Image GT Texonboost CRF FC-CRF

Semantic Video Segmentation

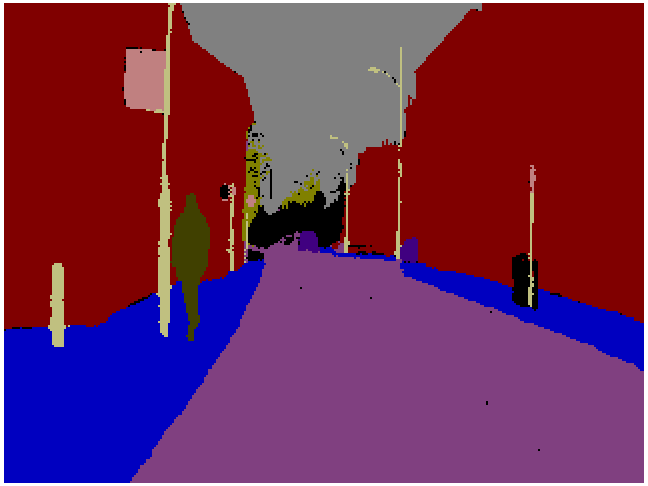


Semantic Video Segmentation



- Spatial-Temporal Deep Structured Models

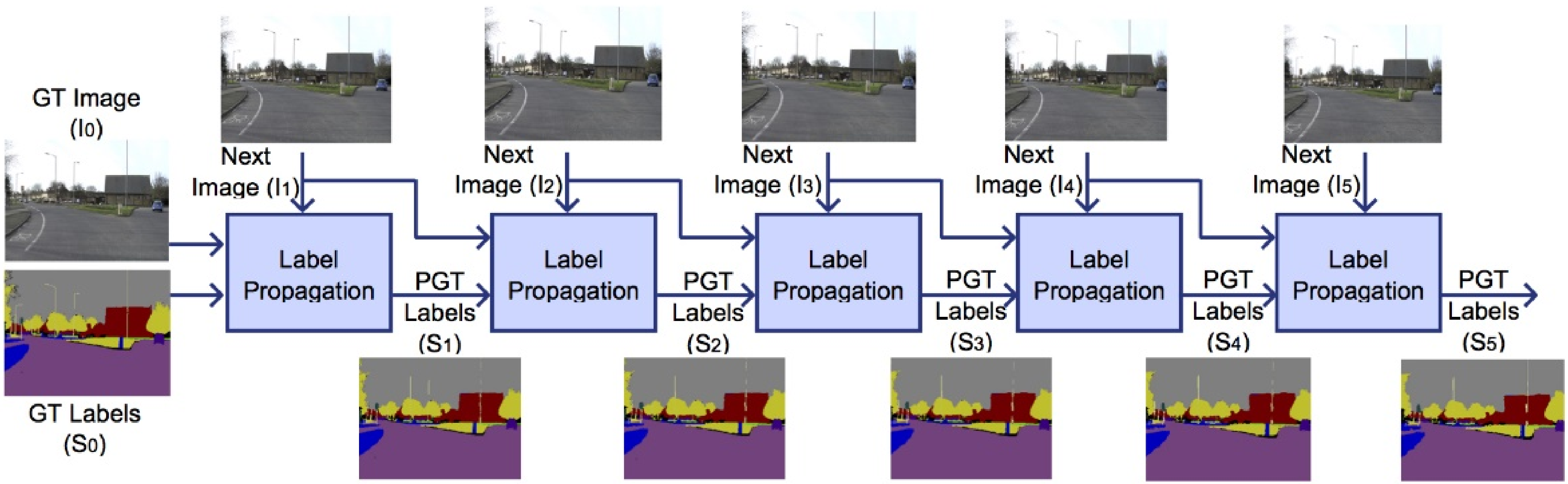
Semantic Video Segmentation

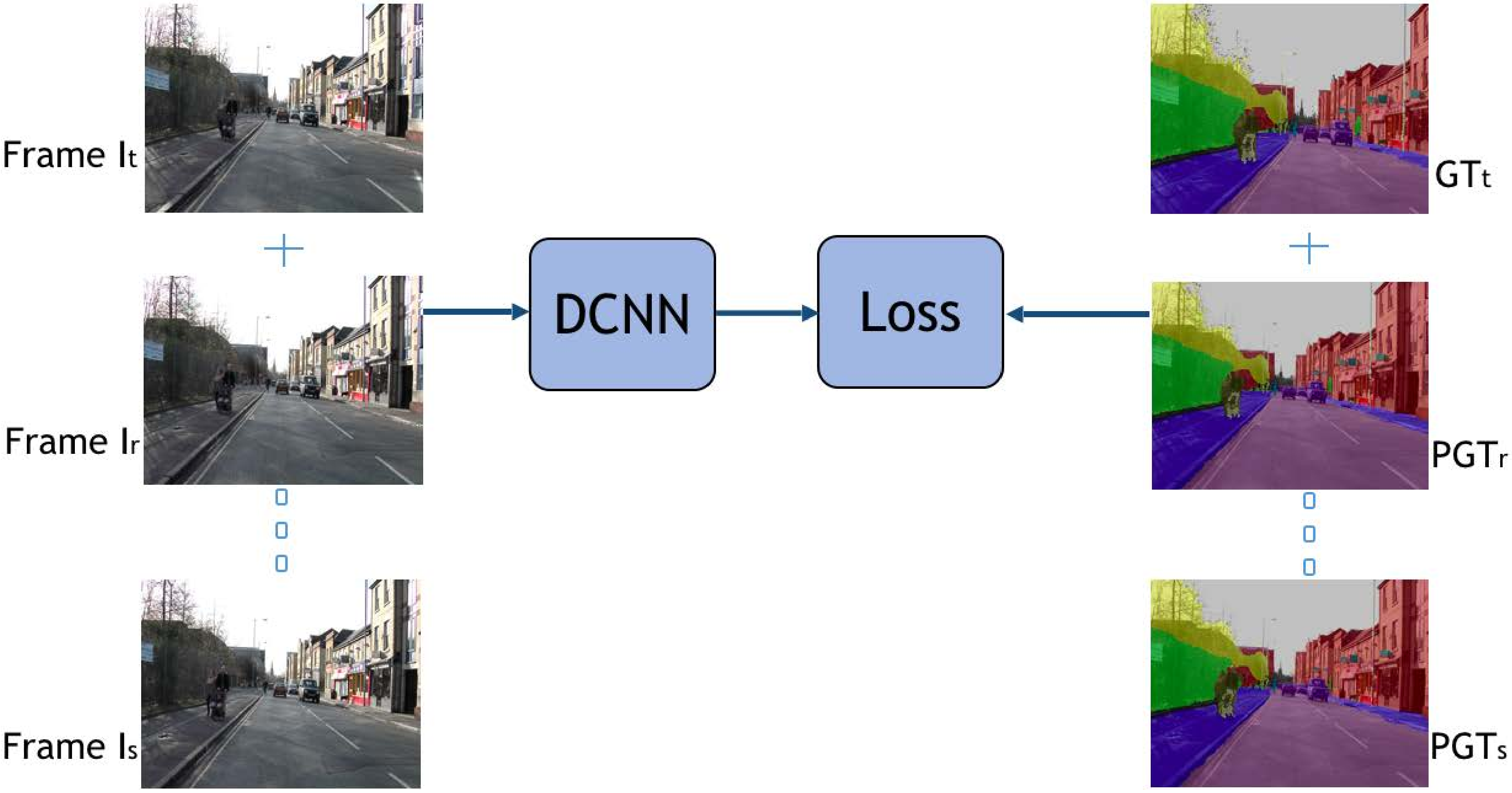
- Spatial-Temporal Deep Structured Models
- Weakly-Supervised Learning CNN+CRF

  - ➢ 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.

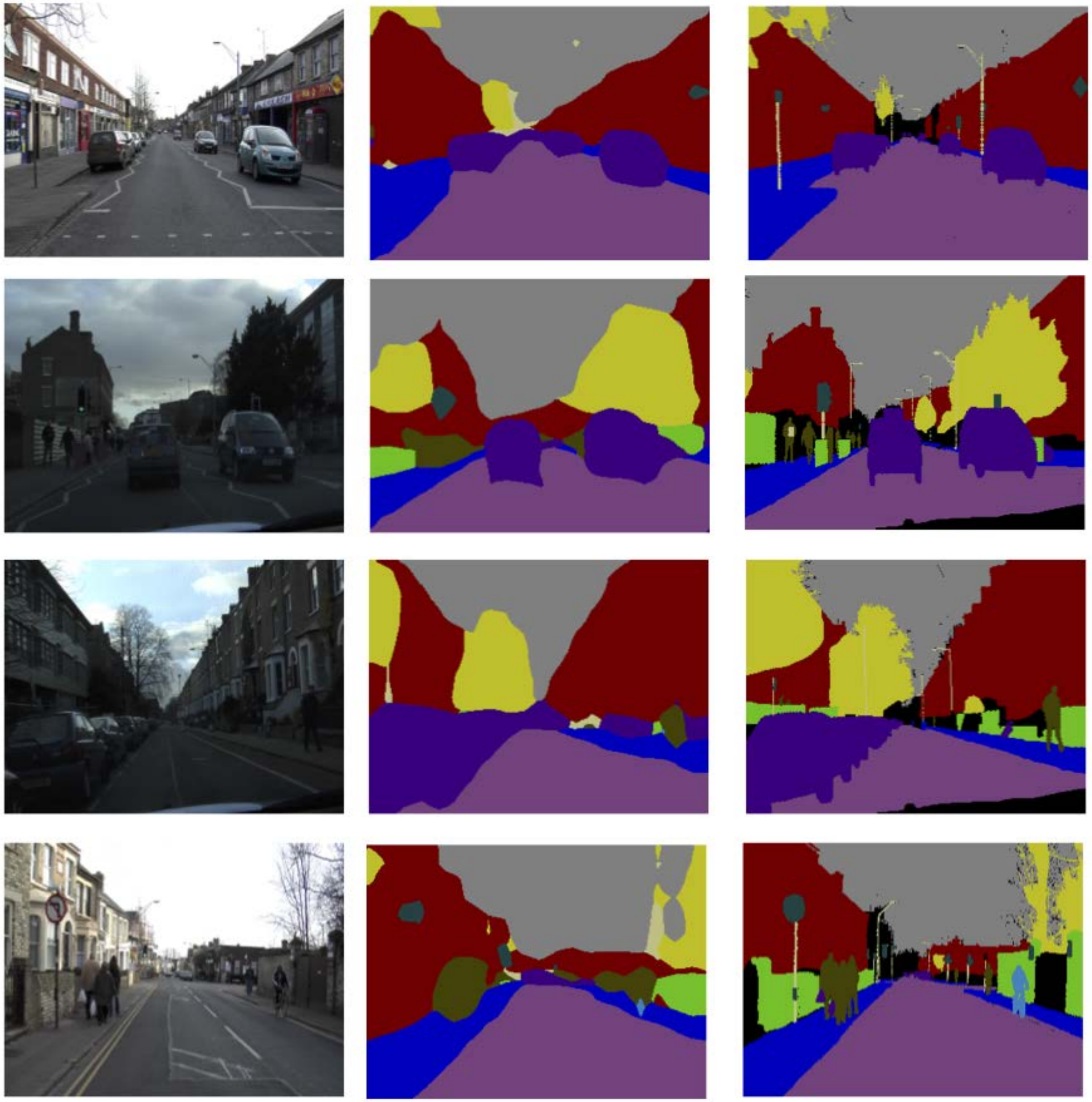
Generating Pseudo Ground Truth Data
CRF for Label Propagation



CNN Training



Results



(a) (b) (c)

Acknowledgement

Collaborators

Rosenhahn

Rother

Förstner

Funding

Deutsche Forschungsgemeinschaft DFG

Leibniz Universität Hannover

Thank you!

ITC
University of Twente, NL