Image Analysis (FMAN20) Lecture F11, 2018

MAGNUS OSKARSSON

Image Analysis - Motivation

Overview – Semantic Segmentation

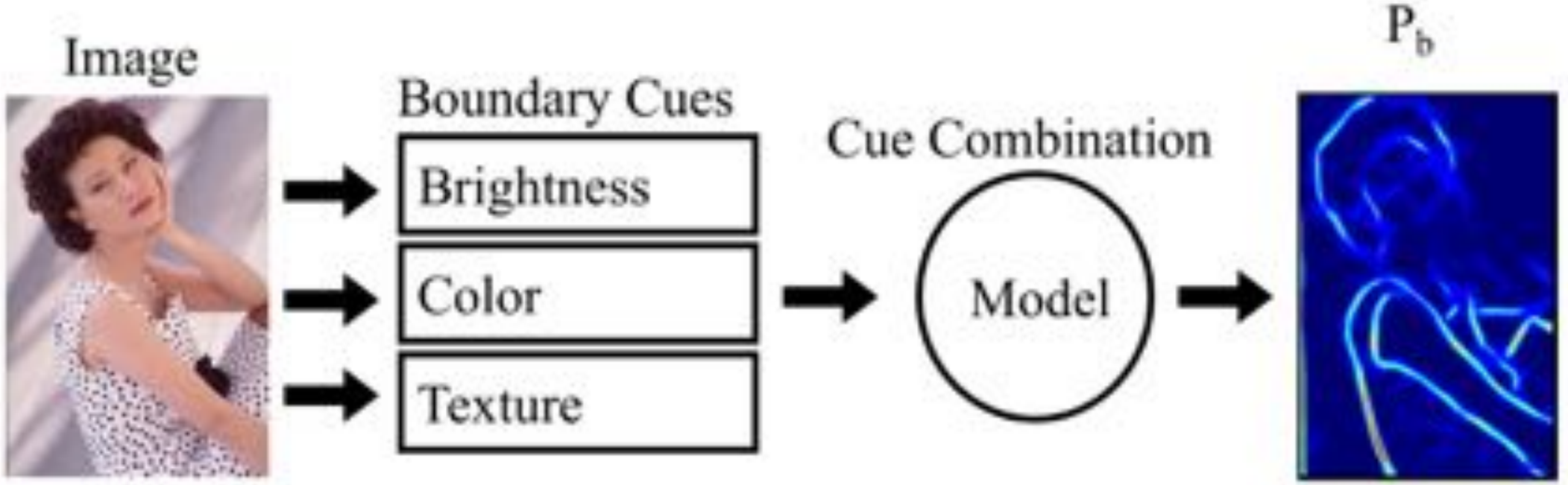
1. **Edge detectors based on machine learning**
2. Segmentation is an ill-posed problem
3. Generating a pool of possible segments (CPMC)
4. Rating segments in the pool
5. Visual and Semantic Processing
6. Second Order Pooling

Segmentation

- **Image segmentation**: breaking the pixels or tokens of an image into regions (groups) that share some property

- **Semantic segmentation**: attach category labels to groups

Modern Edge Detection



1. Collect data set of human segmented images
2. Learn local boundary model by combining brightness, color and texture
3. Global framework to capture closure, continuity

Edge Detection
Current state-of-the-art methods

1. Treat boundary detection as a machine learning problem: **learn a classifier that imitates boundary detection by humans.**

2. Consider both **local cues** (e.g. color, brightness, texture) and **global cues** (e.g. segmentation, probabilistic graphical models) in order to represent structure beyond just local contrast

Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues, D. Martin et al., in PAMI 2004

"Contour Detection and Hierarchical Image Segmentation, P. Arbelaez et al., in PAMI 2011

Occlusion boundary detection and figure/ground assignment from optical flow, Sundberg et al., in CVPR 2011

Efficient closed form Solution to Generalized Boundary Detection, Leordeanu et al., ECCV12

Structured Forests for Fast Edge Detection , P. Dollár and C. Zitnick, ICCV 2013

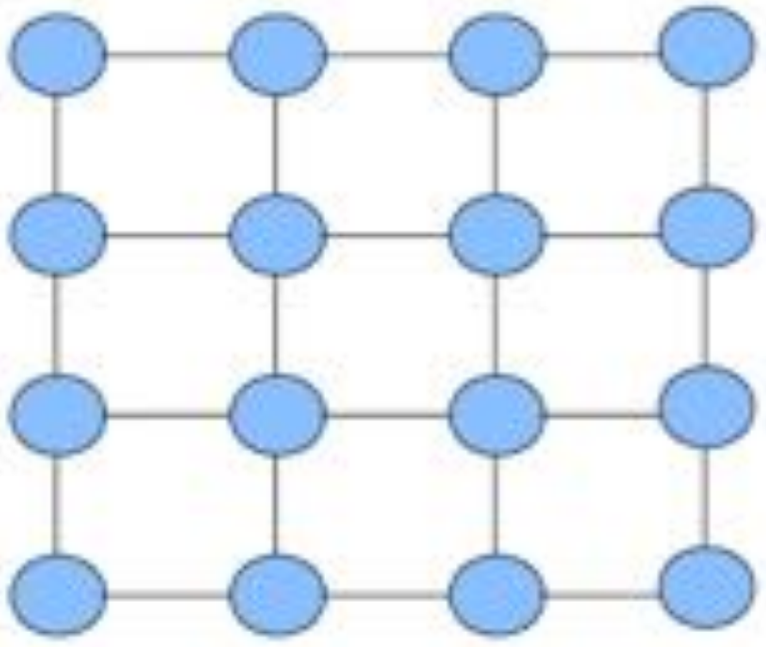
Going to Object-class Segmentation Directly?

Slide by L. Ladický

Pairwise CRF over pixels

Input image

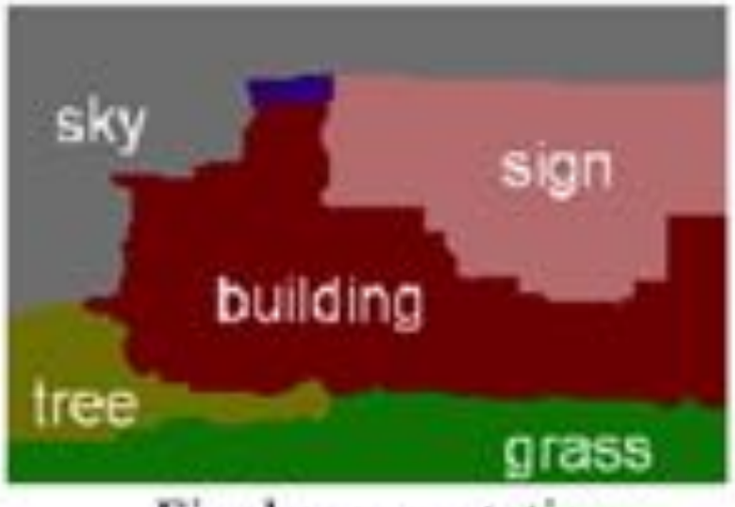
CRF construction



Training of Potentials

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

MAP

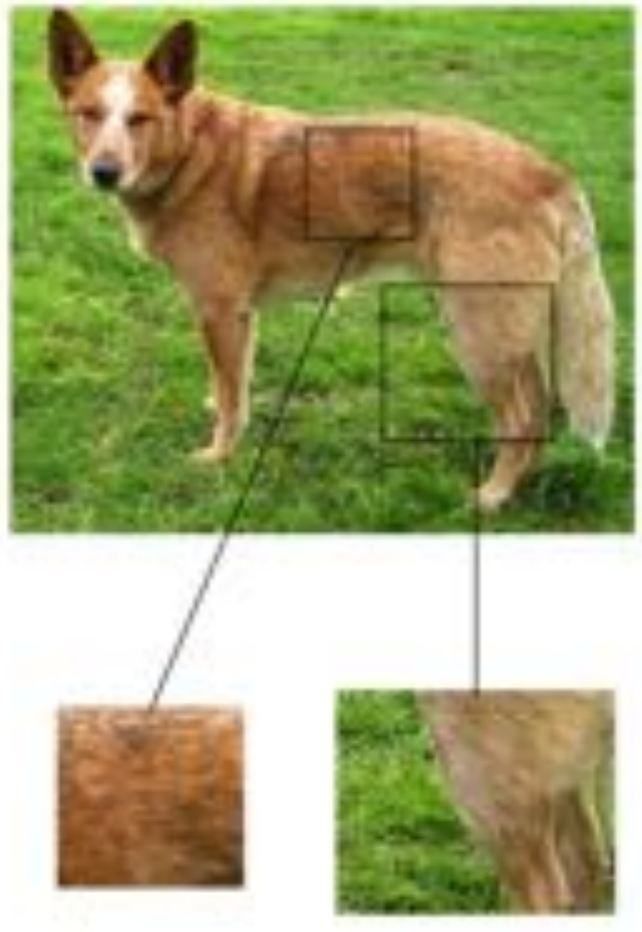


Final segmentation

Shotton et al. ECCV06

Drawbacks of Pixel Classification

- Individual objects are not identified
- Operate over full patches (no figure/ground)
- Difficult to distinguish many object categories locally



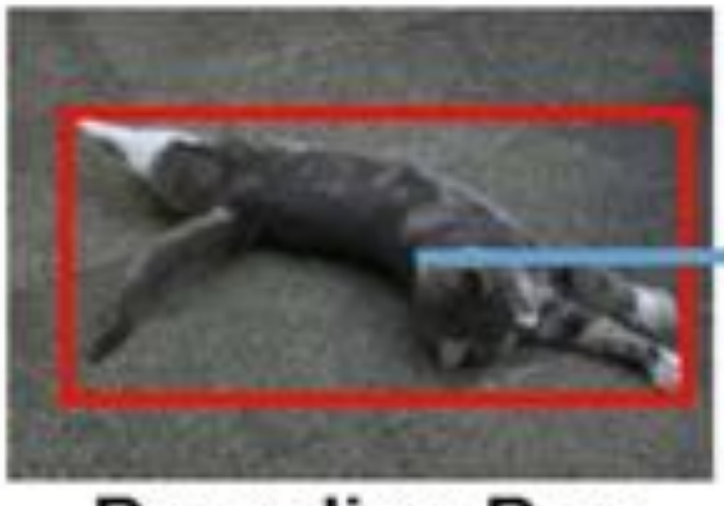
Drawbacks of Pixel Classification

May not segment objects previously unseen

Does spatial support matter?

Ground-Truth Segment

VS.



Bounding Box

Malisiewicz & Efros, BMVC07

Overview – Semantic Segmentation

1. Edge detectors based on machine learning
2. **Segmentation is an ill-posed problem**
3. Generating a pool of possible segments (CPMC)
4. Rating segments in the pool
5. Visual and Semantic Processing
6. Second Order Pooling

Image Segmentation Issues

Mechanism: `jumping' segments

- **Multiple figure-ground segmentations** generated by searching for breakpoints of constrained min-cut energies, at multiple locations and spatial scales on image grid (CPMC)

- **Plausible object segments** are selected after ranking and diversifying the low-level segmentations based on mid-level, class-independent, visual cues

- **Recognition stage** detects objects from the multiple categories and sequentially resolves inconsistencies

Segmentation - ill-posed
What is the right segmentation?

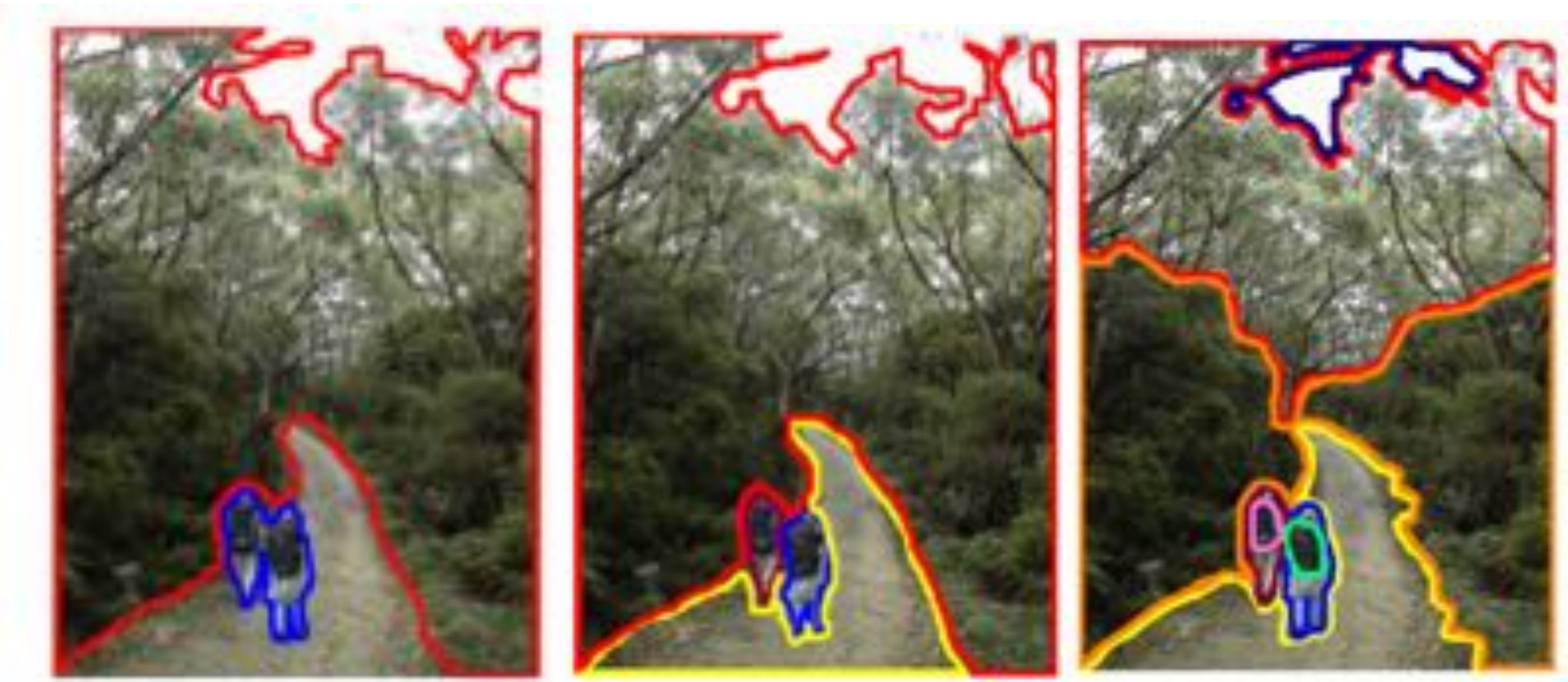
LUND UNIVERSITY

Segmentation - ill-posed
What is the right segmentation?

"A woman with a backpack and a man, also wearing a backpack, are walking on a road. On the sides of the road high trees as well as lower vegetation can be seen. Above, a white sky is peeking through the treetops."

LUND UNIVERSITY

Segmentation - ill-posed
What is the right segmentation?



Visual and semantic processing

Input:

"a man with a white, black and red football uniform is standing behind a trunk with a koala on it"

LUND UNIVERSITY

Visual and semantic processing

Output:

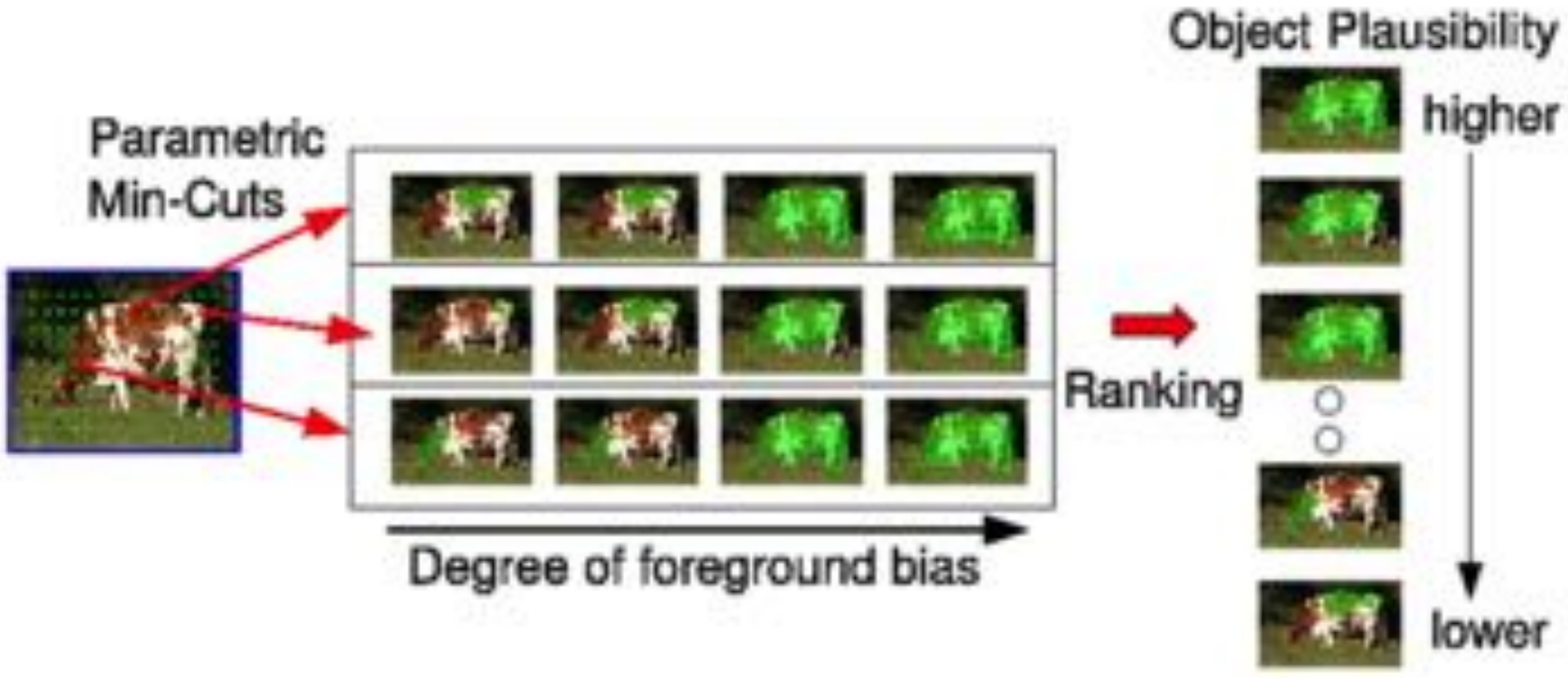
"a man with a white, black and red football uniform is standing behind a fence with a koala on it"

LUND UNIVERSITY

Overview – Semantic Segmentation

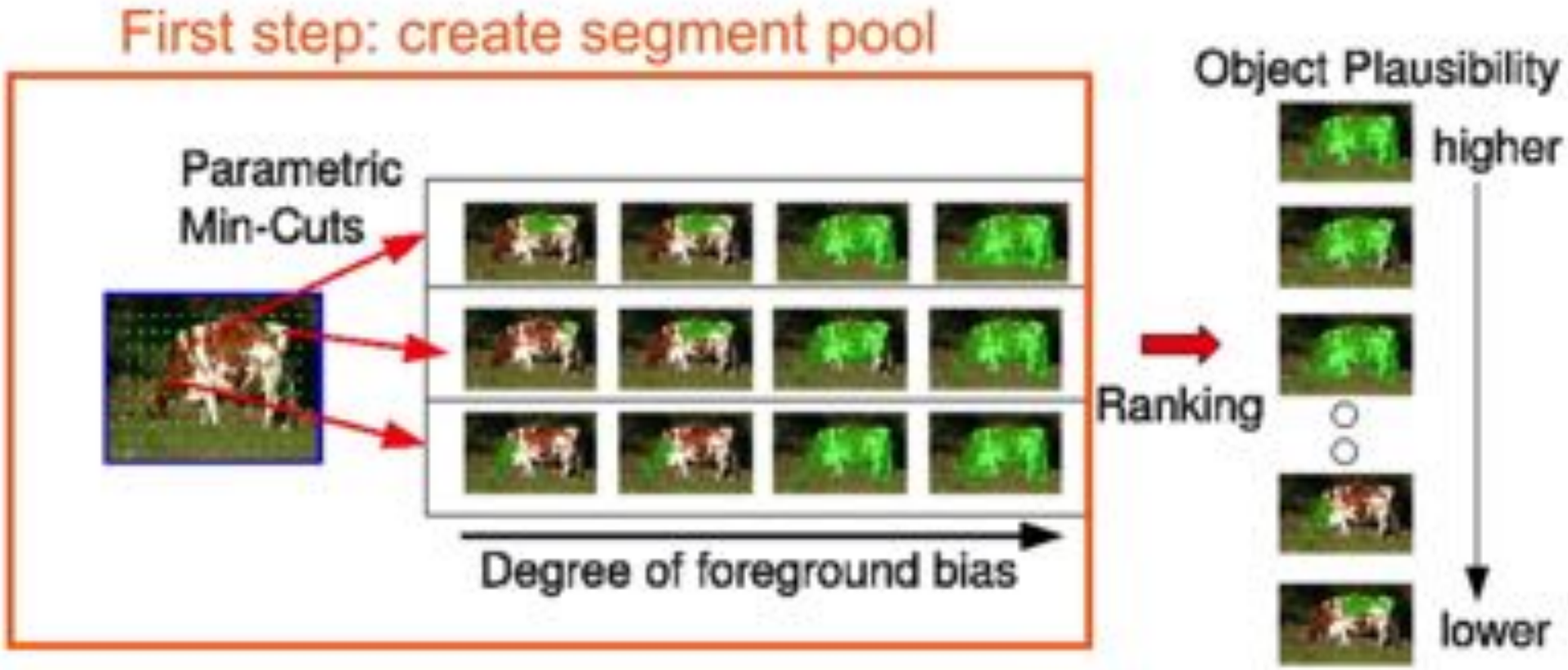
1. Edge detectors based on machine learning
2. Segmentation is an ill-posed problem
3. **Generating a pool of possible segments (CPMC)**
4. Rating segments in the pool
5. Visual and Semantic Processing
6. Second Order Pooling

CPMC: Constrained Parametric Min-Cuts for Automatic Object Segmentation



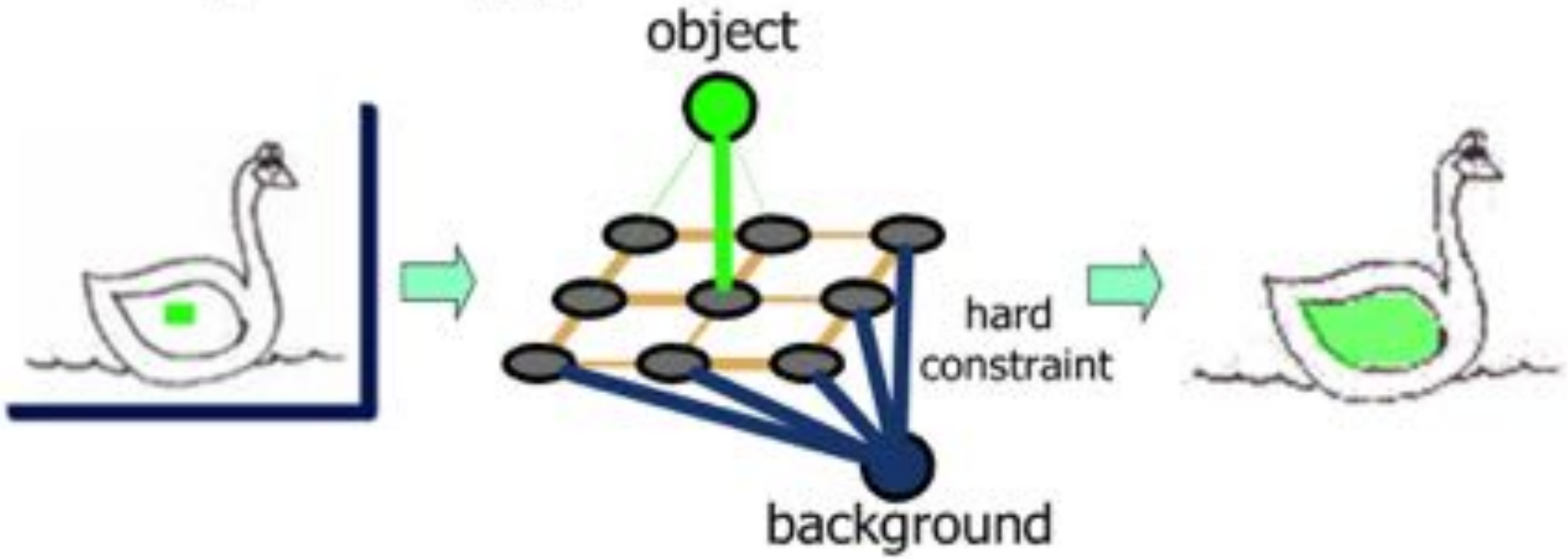
Carreira and Sminchisescu. CVPR 10, PAMI 11

CPMC: Constrained Parametric Min-Cuts for Automatic Object Segmentation



Generating a segment pool: constrained min-cut

$$E_\lambda(x) = \sum_{u \in V} D(x_u, \lambda) + \sum_{(u,v) \in E} V_{uv}(x_u, x_v)$$



Constrained Parametric Min-Cuts

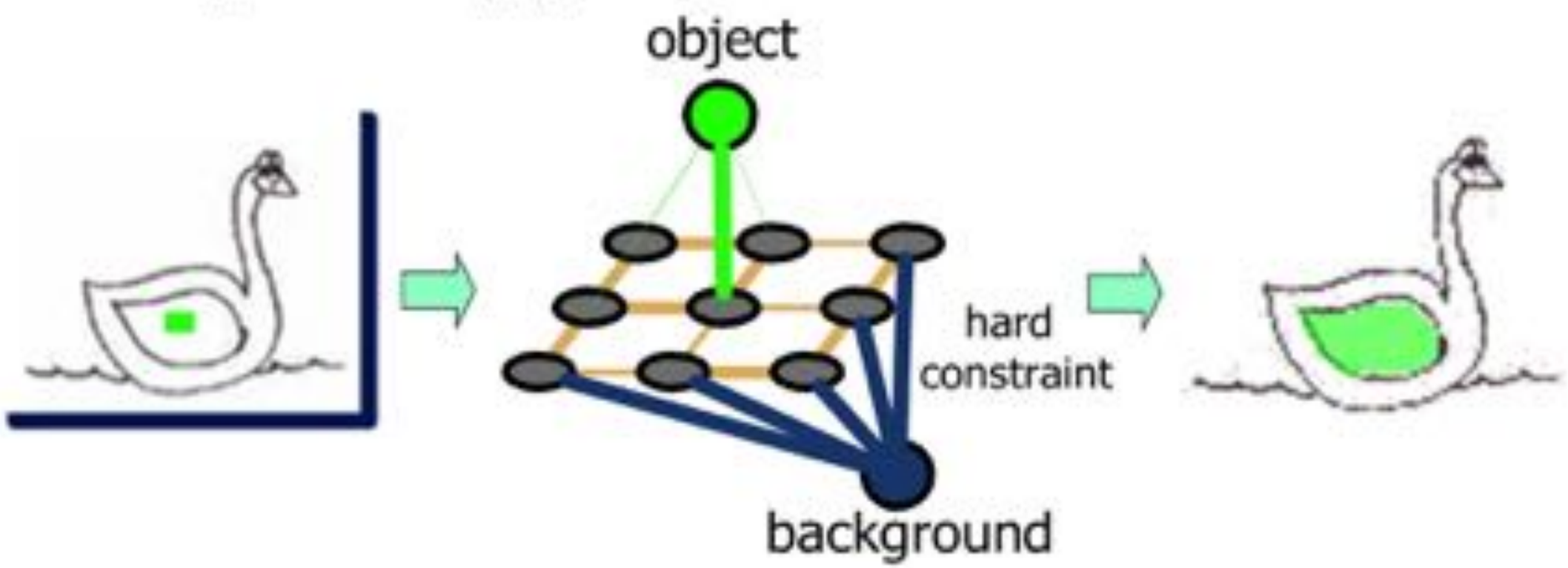
$$E_\lambda(x) = \sum_{u \in V} D(x_u, \lambda) + \sum_{(u,v) \in E} V_{uv}(x_u, x_v)$$

$$D(x_u, \lambda) = \begin{cases} 0, & \text{if } x_u = 1, x_u \notin X_b \\ \infty, & \text{if } x_u = 1, x_u \in X_b \\ \infty, & \text{if } x_u = 0, x_u \in X_f \\ \ln \dfrac{p_f(x_u)}{p_b(x_u)} + \lambda, & \text{if } x_u = 0, x_u \notin X_f \end{cases}$$

$$V(x_u, x_v) = \begin{cases} 0, & \text{if } x_u = x_v \\ \exp\left[- \dfrac{\max(gPb(u), gPb(v))}{\sigma^2} \right], & \text{if } x_u \neq x_v \end{cases}$$

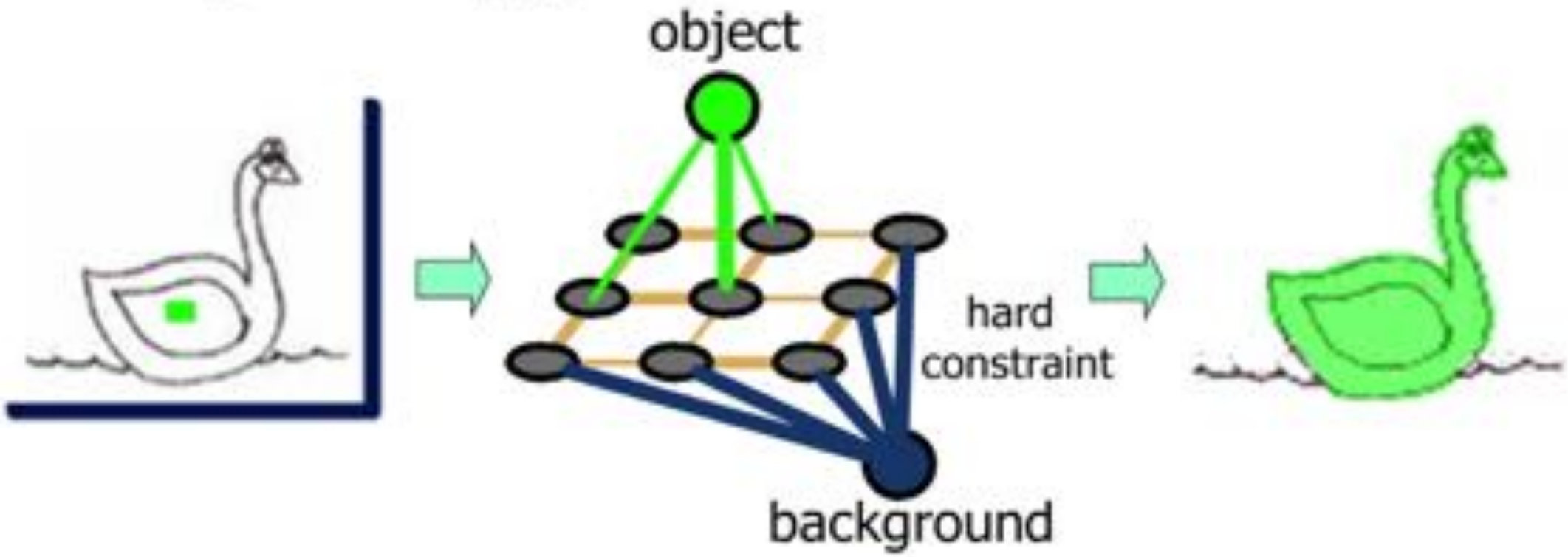
Generating a segment pool: constrained min-cut

$$E_\lambda(x) = \sum_{u \in V} D(x_u, \lambda) + \sum_{(u,v) \in E} V_{uv}(x_u, x_v)$$

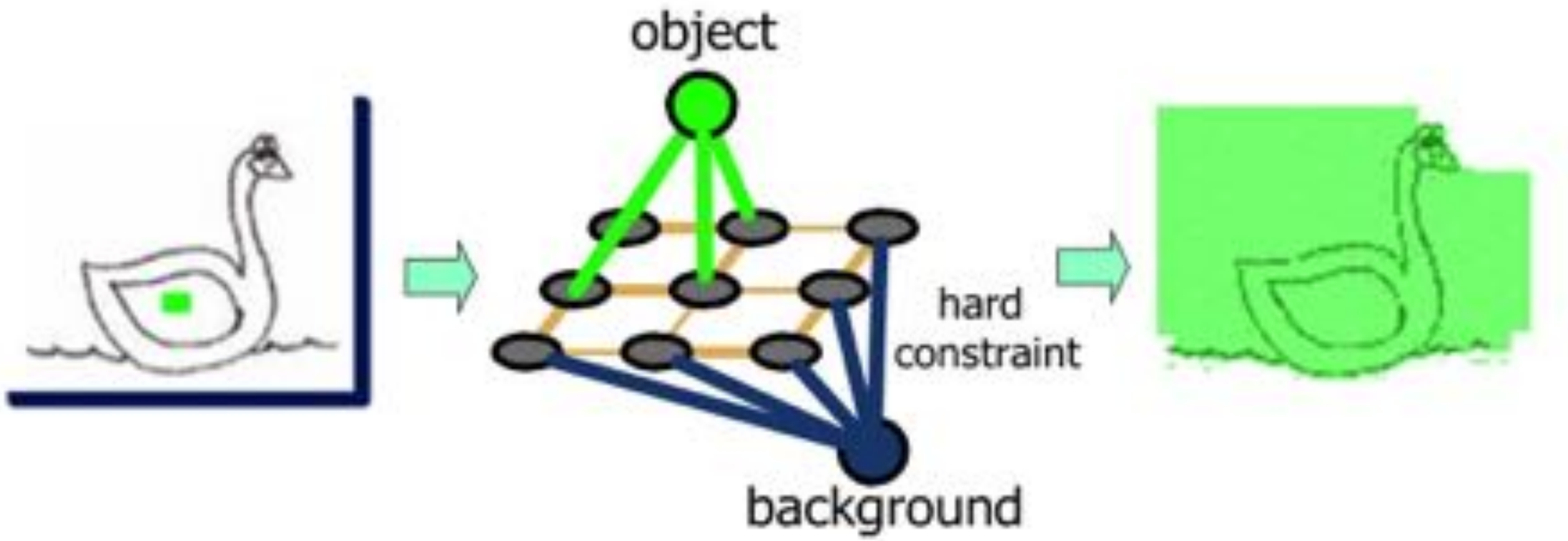


Generating a segment pool: constrained *parametric* min-cuts

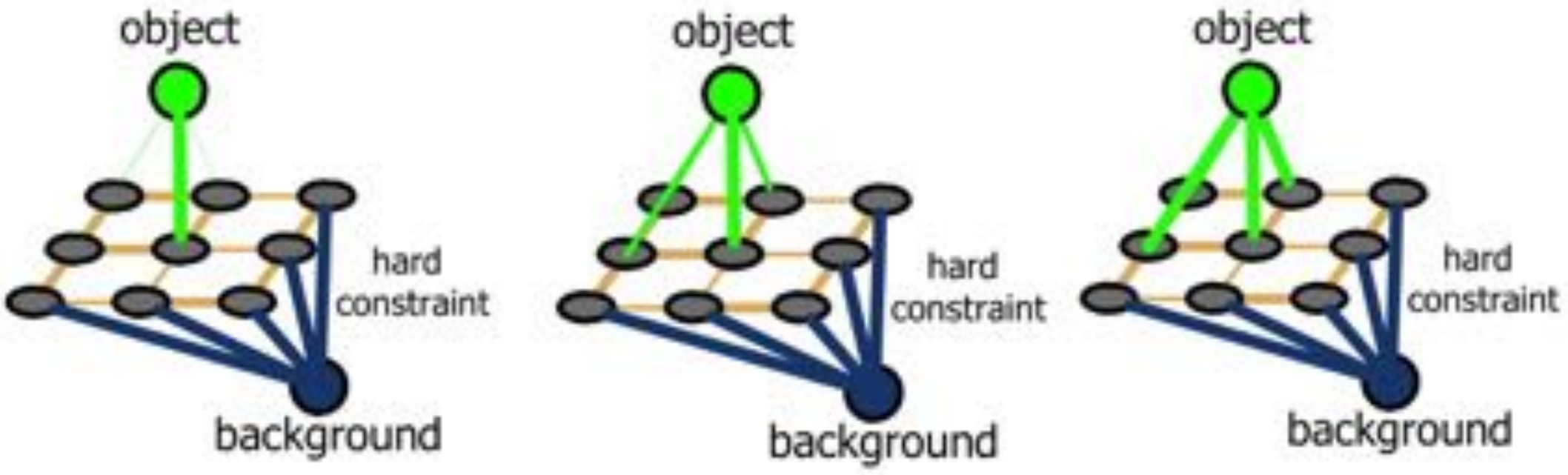
$$E_\lambda(x) = \sum_{u \in V} D(x_u, \lambda) + \sum_{(u,v) \in E} V_{uv}(x_u, x_v)$$



Generating a segment pool: constrained *parametric* min-cuts



Generating a segment pool: constrained *parametric* min-cuts



Can solve for all values of object bias in the same time complexity of solving a single min-cut using a **parametric max-flow solver**

RGB-D Edges

RGB and Depth channels provide complementary information.
Max operator helps identify both intensity and depth discontinuities.

Parametric Proposals (CPMC-RGBD)

- Use **RGB and depth** to improve the segment-pool
- Combine RGB and depth channels in the pairwise term of the energy model

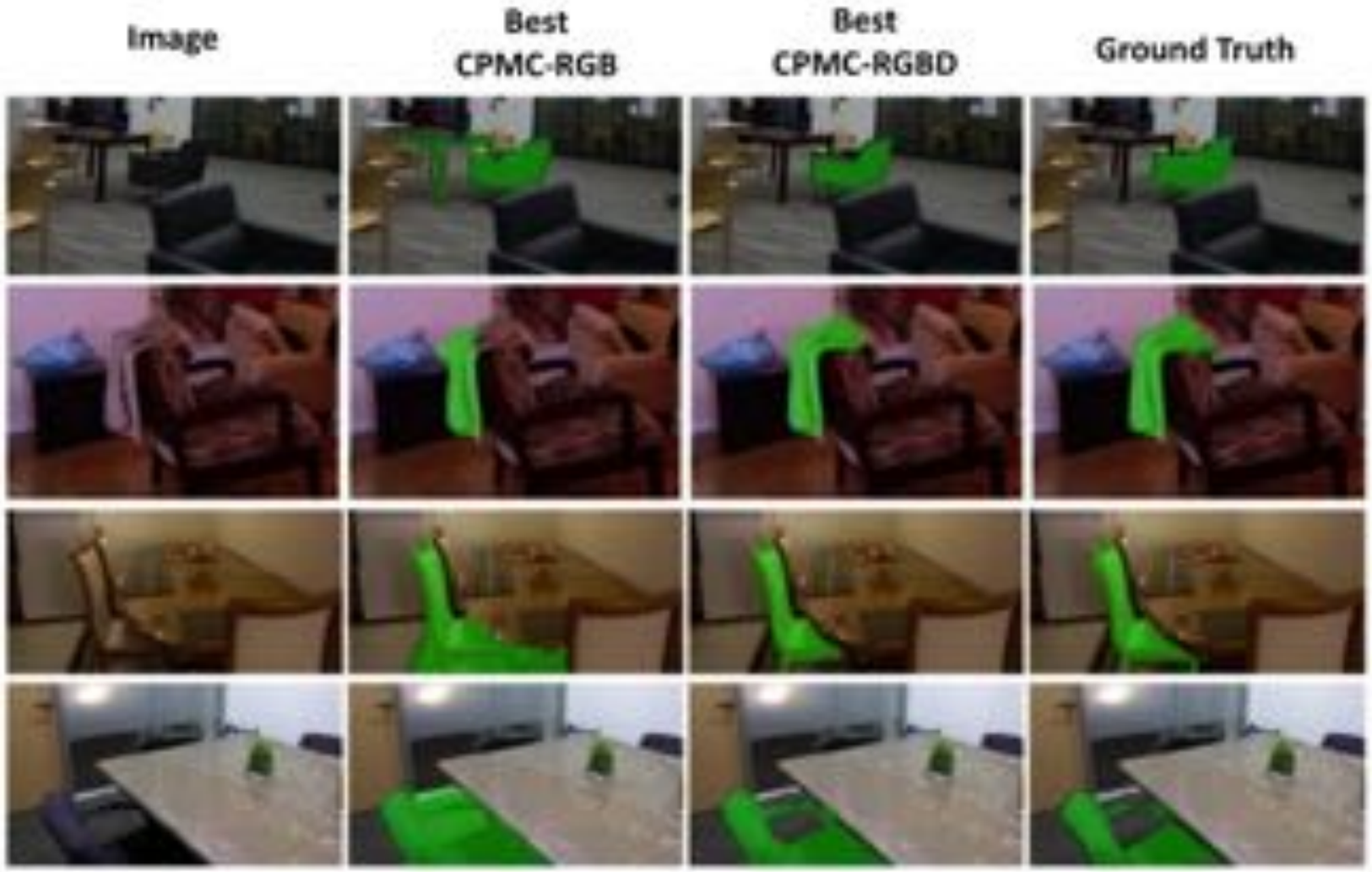
$$E_\lambda(x) = \sum_{u \in V} D(x_u, \lambda) + \sum_{(u,v) \in E} V_{uv}(x_u, x_v)$$

$$V_{uv}(x_u, x_v) = \exp\left[-\frac{\max\left(B_{RGB}(u), B_{RGB}(v), B_{Depth}(u), B_{Depth}(v)\right)}{\sigma^2} \right]$$

N.B. Motion information (e.g. optical flow) can also be incorporated

D. Banica, A. Agape, A. Ion, and C. Sminchisescu. Video Object Segmentation by Salient Segment Chain Composition. In International Conference on Computer Vision, ICCV 2013

Parametric Proposals (CPMC-RGBD)

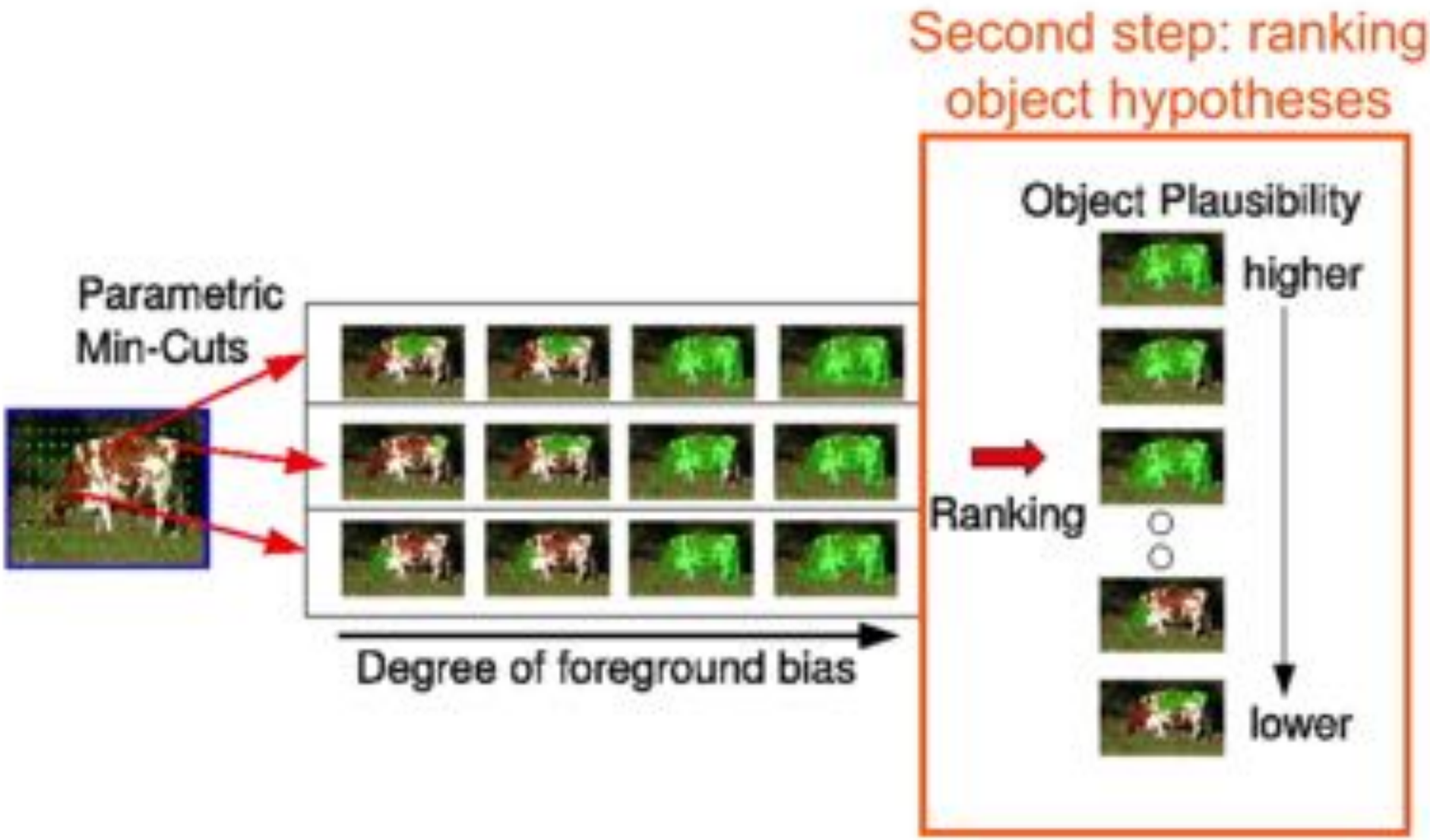


- Captures subtle details (thin structures, handles)
- Fewer segments across depth discontinuities. Improvement may not be well captured by current metrics, but may have great functional importance

Overview – Semantic Segmentation

1. Edge detectors based on machine learning
2. Segmentation is an ill-posed problem
3. Generating a pool of possible segments (CPMC)
4. **Rating segments in the pool**
5. Visual and Semantic Processing
6. Second Order Pooling

CPMC: Constrained Parametric Min-Cuts for Automatic Object Segmentation

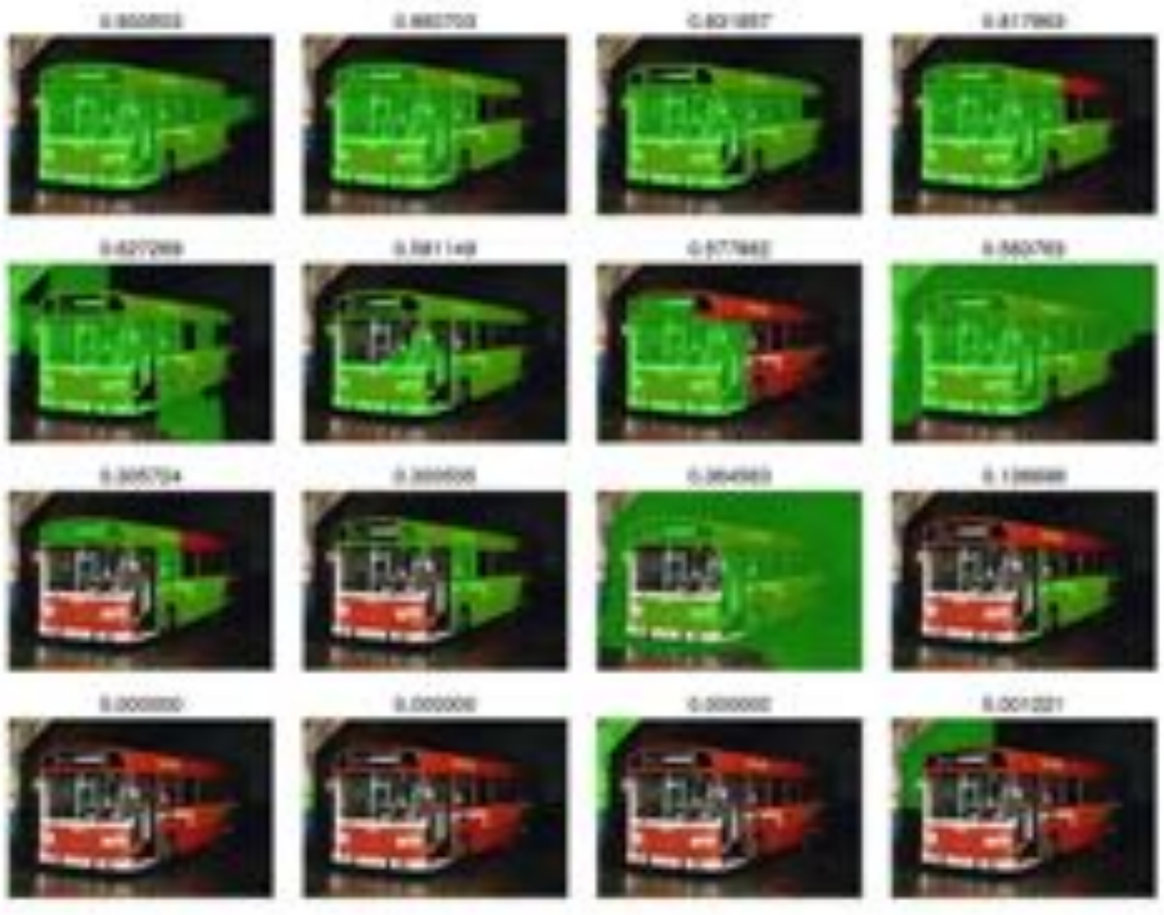


How to model segment quality ?

Best **overlap** with a ground truth object computed by intersection-over-union.

Ranking figure-ground hypotheses

- Supervised learning framework
- Hypothesized segments ranked using regression
- Ranking is class–independent (mid-level)



Learning to rank object hypotheses

Input object hypotheses

Feature Vectors

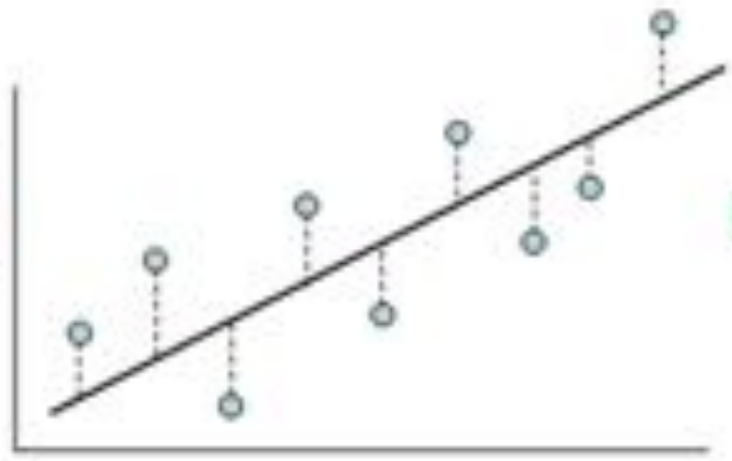
Parameter Optimization

Desired Prediction

$$\begin{bmatrix} 0.1 \\ -0.3 \\ 0.5 \\ \vdots \\ 0.4 \end{bmatrix}$$

0.63

$$\begin{bmatrix} 0.5 \\ 0.3 \\ 0.9 \\ \vdots \\ 0.4 \end{bmatrix}$$



0.93

$$\begin{bmatrix} -0.3 \\ 0 \\ 0.1 \\ \vdots \\ 0.9 \end{bmatrix}$$

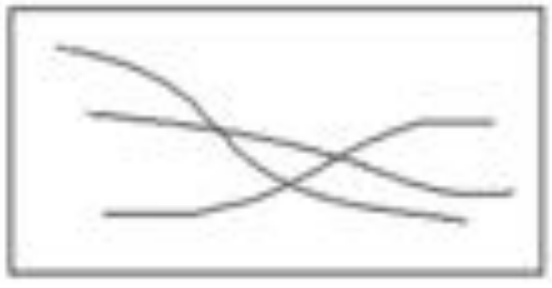
0

Gestalts

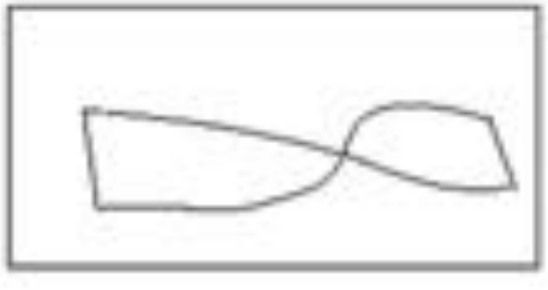
Gestalt psychology identifies several properties that result in grouping/segmentation:

Parallelism

Symmetry



Continuity



Closure

Ranking Object Hypotheses

- Aims to handle full object segments and fragments
- Modeled as regression on overlap
- Features
  - *Boundary* – normalized boundary energy
  - *Region* – location, perimeter, area, Euler number, orientation, contrast with background
  - *Gestalt* – convexity, smoothness

Good

High boundary energy

Smooth

Euler number = 0

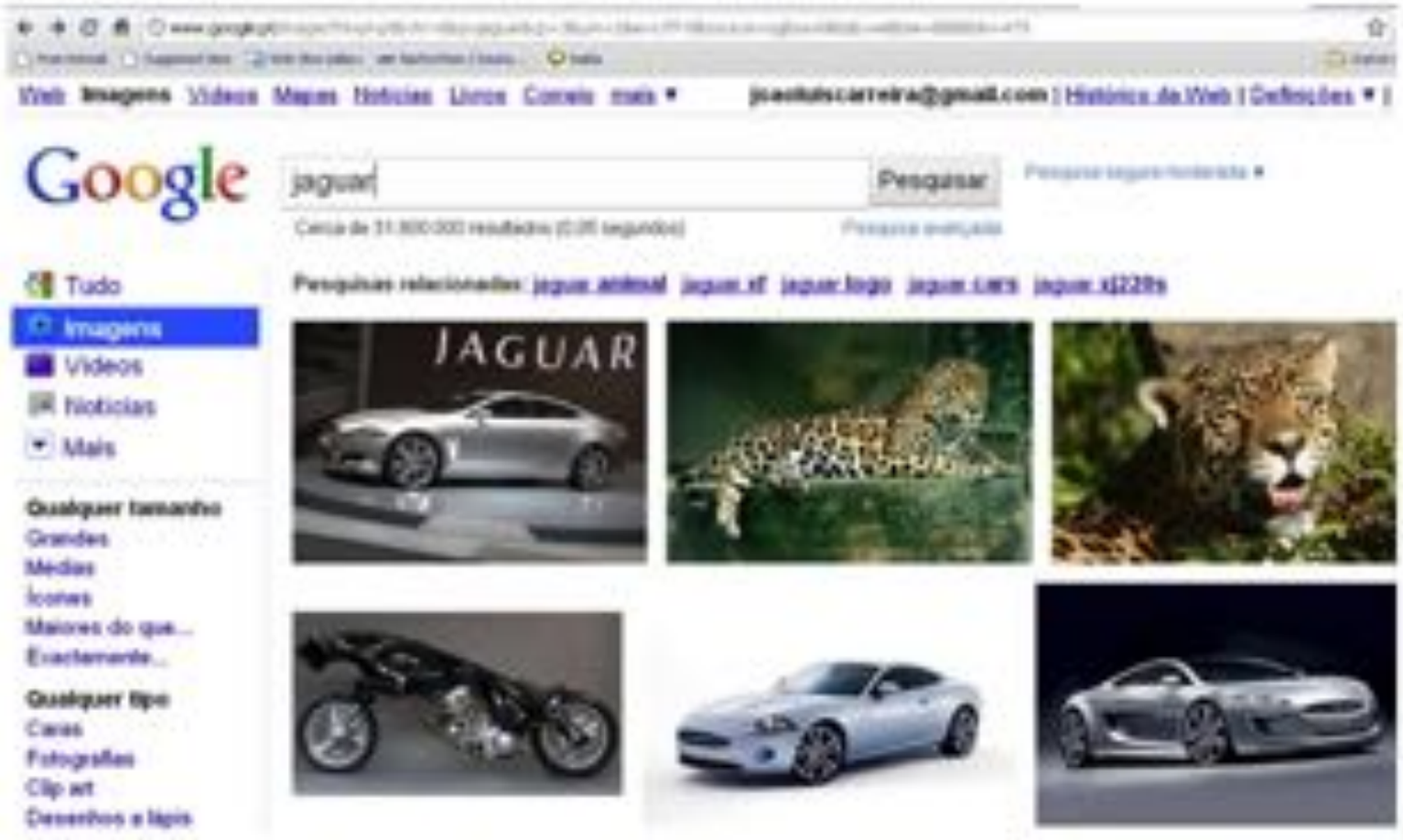
Bad

Low boundary energy

Non smooth

High Euler number

Ranking Diversification



Object hypotheses are re-ranked using the Maximum Marginal Relevance algorithm

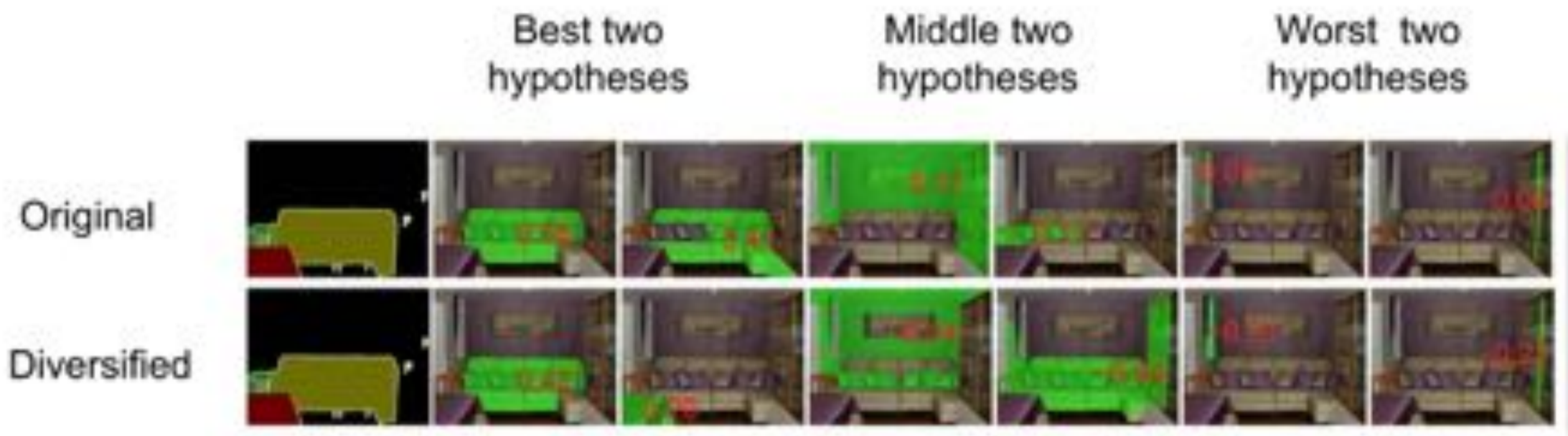
Diversifying Ranking

Rank segments based on ***M**aximum **M**arginal **R**elevance*

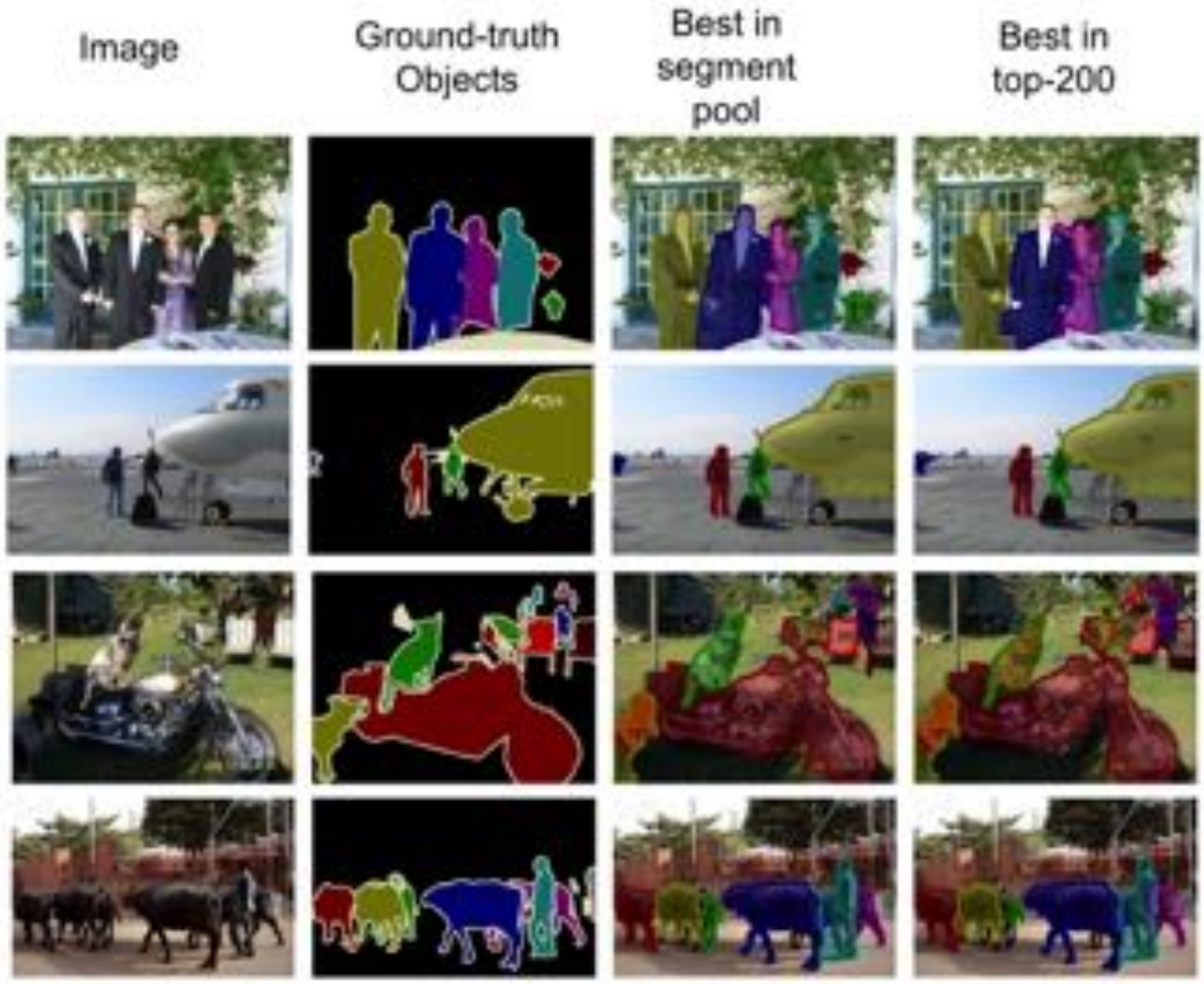
$$MMR = \underset{H_i \in H \setminus H_p}{\operatorname{argmax}} \left[\theta \cdot s(H_i) - (1-\theta) \cdot \max_{H_j \in H_p} o(H_i, H_j) \right]$$

Segment score

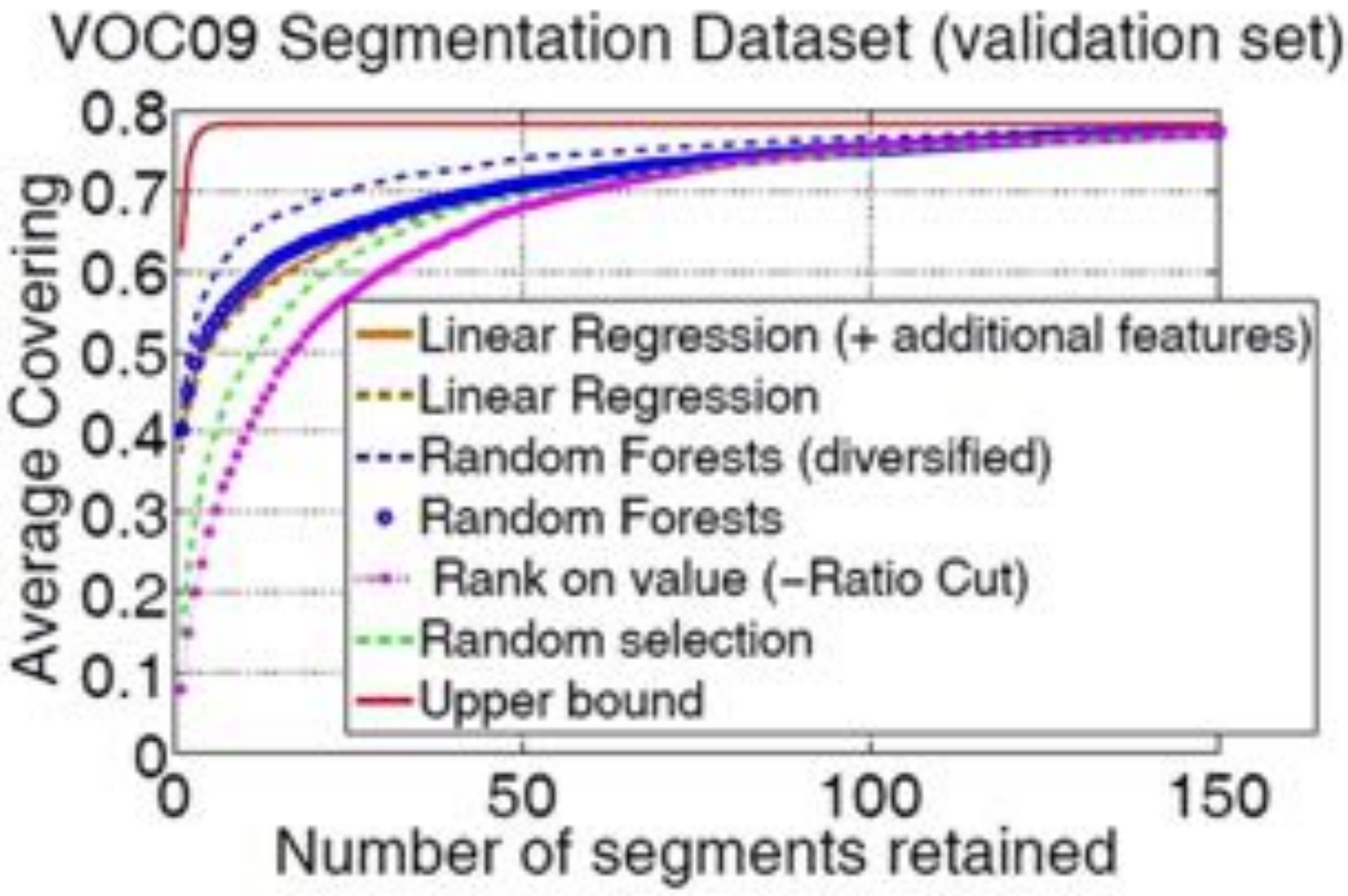
Pairwise segment overlap



Segmentation Examples



Ranking

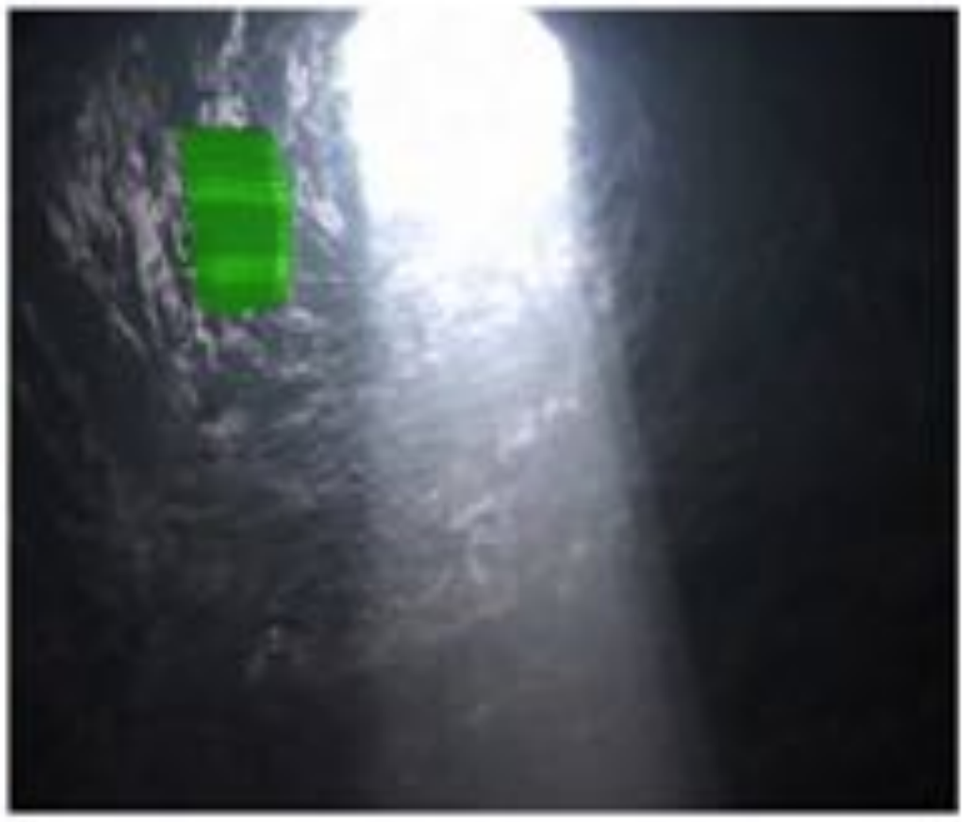

VOC09 Segmentation Dataset (validation set)
Average Covering
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
Number of segments retained
Linear Regression (+ additional features)
Linear Regression
Random Forests (diversified)
Random Forests
Rank on value (−Ratio Cut)
Random selection
Upper bound

CPMC Segment Generation

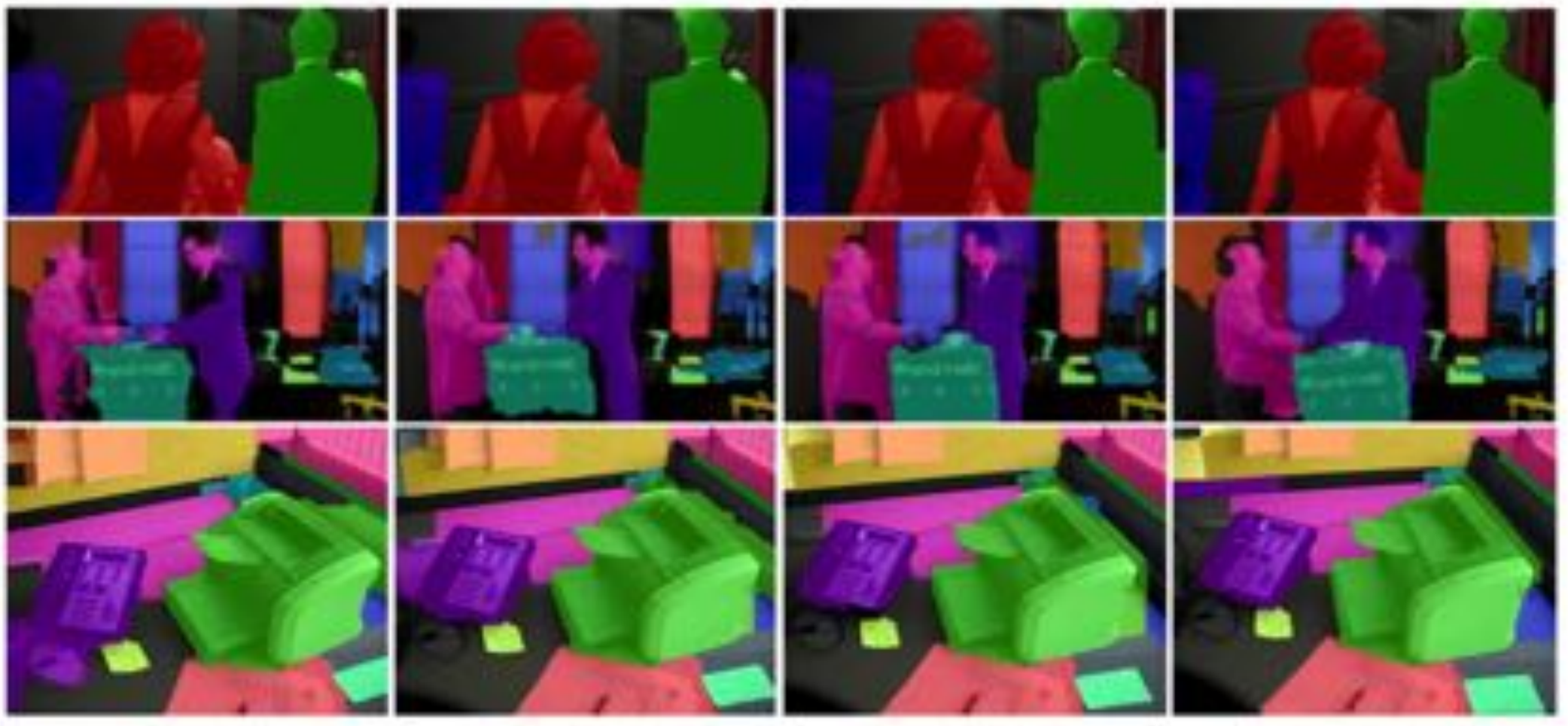
CPMC Segment Generation

Video Segmentation by Figure-Ground Composition

D. Banica, A. Agape, A. Ion, and C. Sminchisescu. **Video Object Segmentation by Salient Segment Chain Composition**. In International Conference on Computer Vision, ICCVW 2013



Video Segmentation



D. Banica, A. Agape, A. Ion, and C. Sminchisescu. **Video Object Segmentation by Salient Segment Chain Composition**. In International Conference on Computer Vision, ICCVW 2013

Overview – Semantic Segmentation

1. Edge detectors based on machine learning
2. Segmentation is an ill-posed problem
3. Generating a pool of possible segments (CPMC)
4. Rating segments in the pool
5. **Visual and Semantic Processing**
6. Second Order Pooling

Visual and Semantic Processing

"People walking in the woods”

“A woman with a backpack and a man, also wearing a backpack, are walking on a road. On the sides of the road high trees as well as lower vegetation can be seen. Above, a white sky is peeking through the treetops."

Problem Formulation

- We investigate the problem of segmenting images using the information in text annotations.
- In contrast to the general image understanding problem, this type of annotation guided segmentation is less ill-posed.
- We present a system based on a combined visual and semantic pipeline.

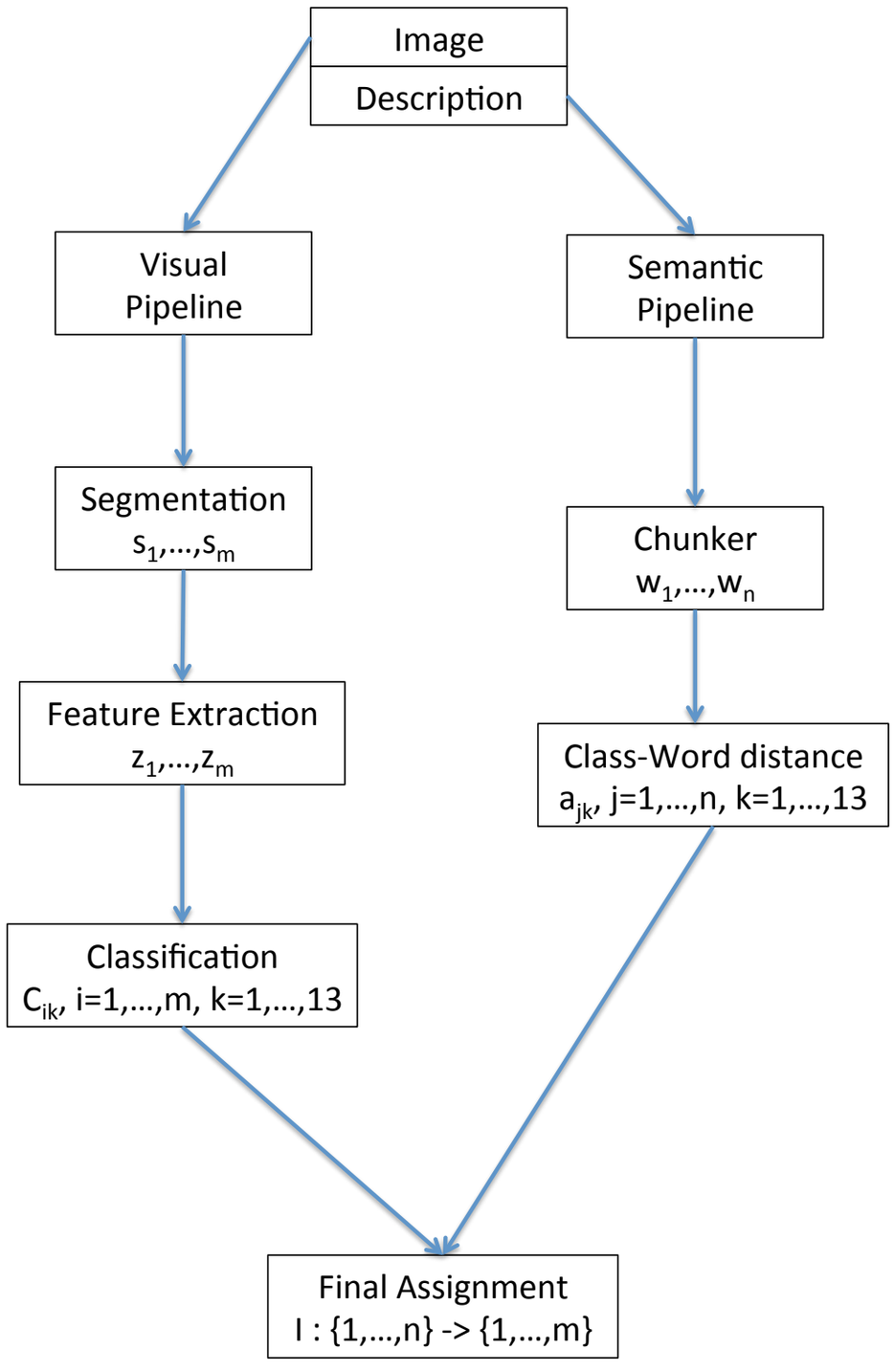
Visual and Semantic Processing

Visual Parsing

- Image segmented using CPMC. Usually 500-1000 segments.
- 27 features.
- Classification into 13 visual categories

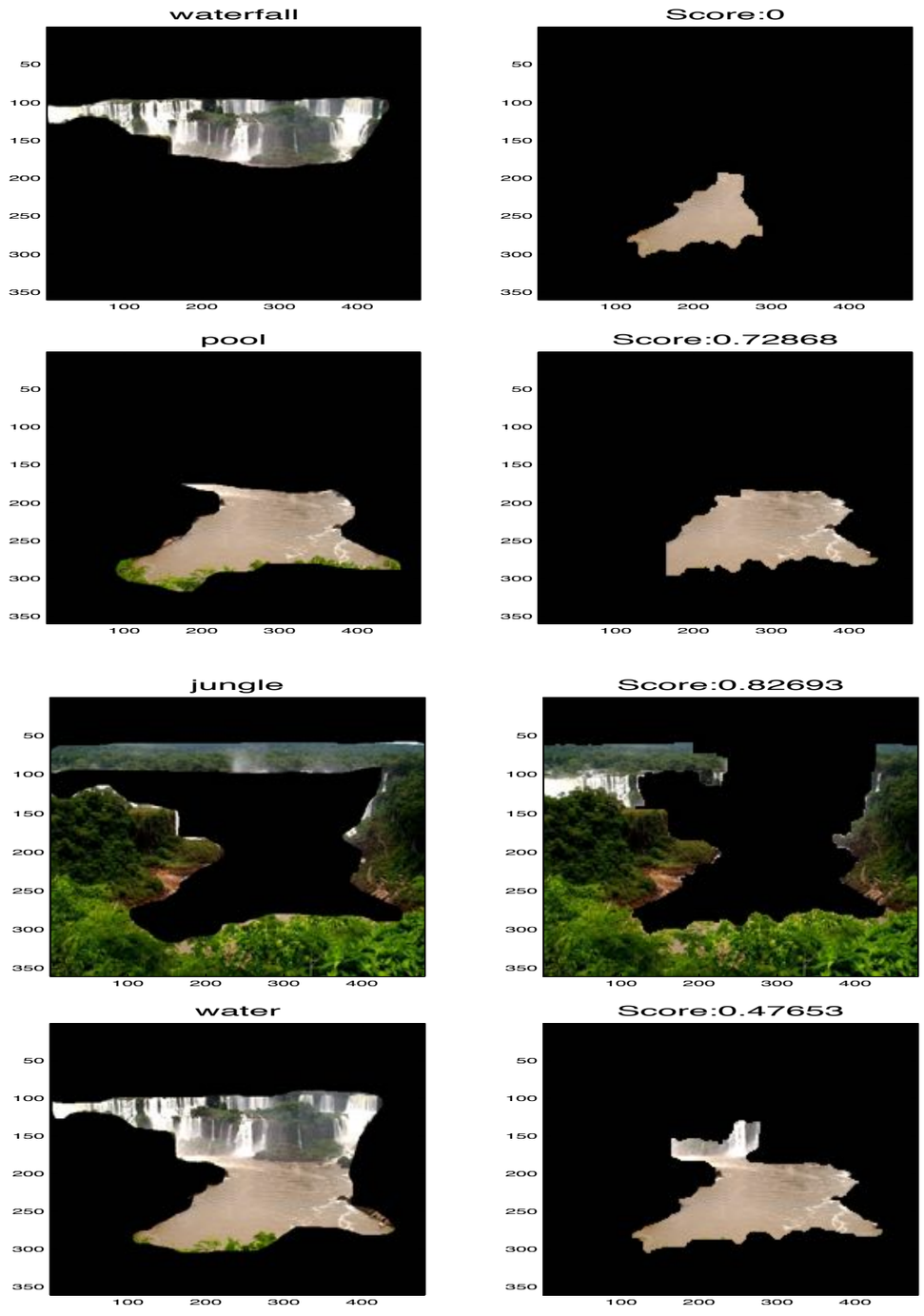
Semantic Parsing

- Chunking of text to produce key-nouns in text. Usually 3-10 key-nouns per annotation
- Calculation of semantic distance between each key-noun and each visual category.
- Final assignment using combinatorial optimization of segment for each key-noun



Visual and Semantic Processing

"A cascading **waterfall** in the middle of the **jungle**; front view with **pool** of dirty **water** in the foreground"



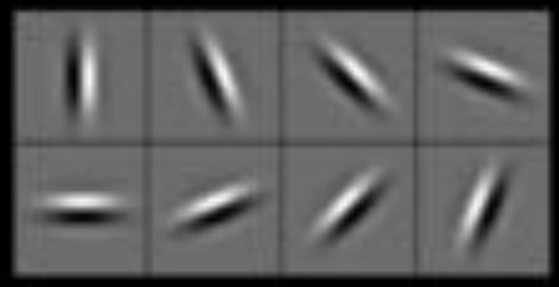
Overview – Semantic Segmentation

1. Edge detectors based on machine learning
2. Segmentation is an ill-posed problem
3. Generating a pool of possible segments (CPMC)
4. Rating segments in the pool
5. Visual and Semantic Processing
6. **Second Order Pooling**

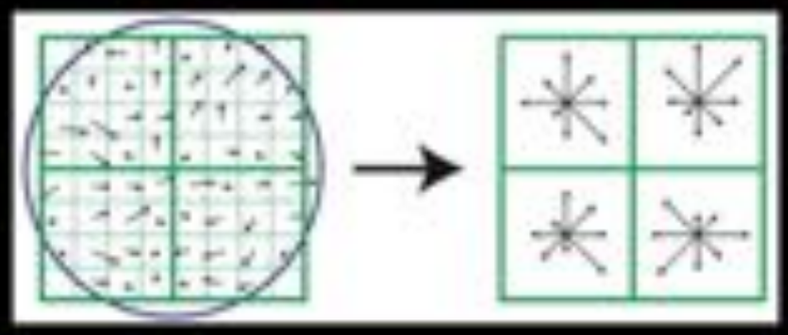
Local Features: SIFT Descriptor

Image Pixels

Apply Gabor filters



Spatial pool (Sum)



Normalize to unit length

Feature Vector

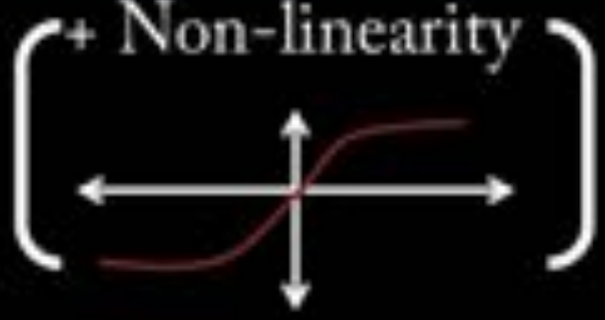
Slide from Fergus

Example Feature Learning Architectures

Pixels / Features

Filter with Dictionary (patch/tiled/ convolutional)

+ Non-linearity



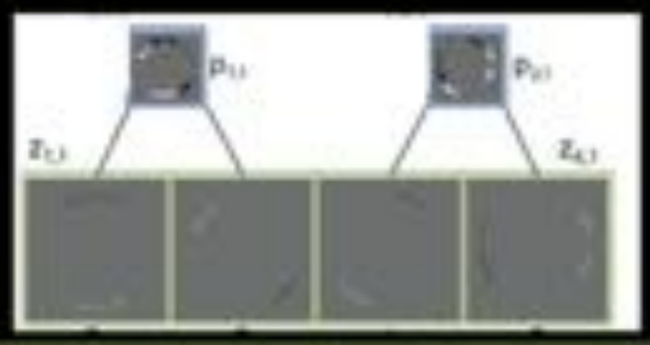
Normalization between feature responses

(Group) ℓ_1 Sparsity

Max / Softmax

Local Contrast Normalization (Subtractive & Divisive)

Spatial/Feature (Sum or Max)

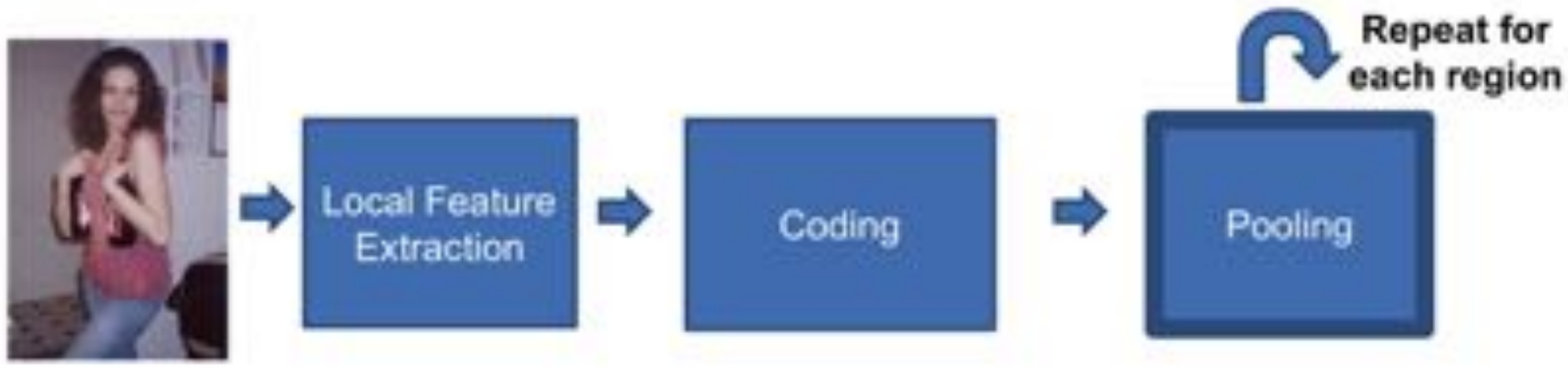


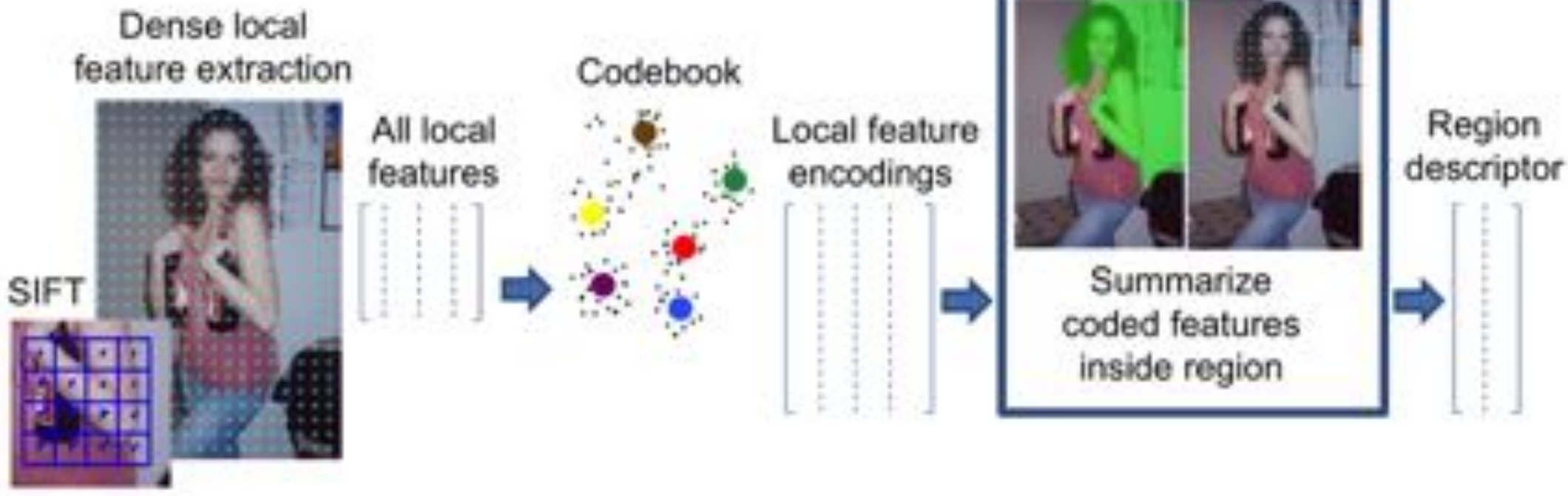
Features

Not an exact separation

Fergus, 2013

Aggregation-based Descriptors





Second Order Pooling (O2P)

Can we pursue richer statistics for pooling ?

Capture correlations

$$\boldsymbol{g}_{avg} = \frac{1}{N}\sum_i \boldsymbol{x}_i \quad \Longrightarrow \quad \boldsymbol{G}_{avg} = \frac{1}{N}\sum_i \boldsymbol{x}_i \cdot \boldsymbol{x}_i^{\mathrm{T}}$$

Dimensionality = (local descriptor size)^2

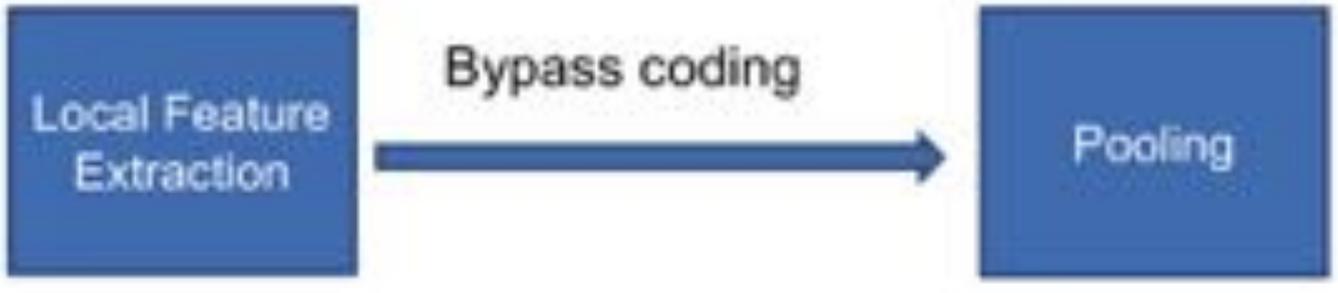
Second Order Pooling (O2P)

Can we pursue richer statistics for pooling ?

Capture correlations

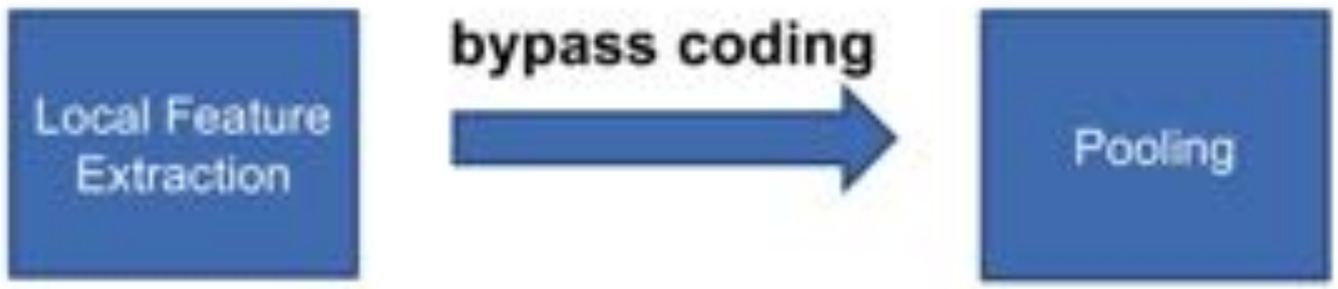
$$\boldsymbol{g}_{avg} = \frac{1}{N}\sum_{i}^{N} \boldsymbol{x}_i \quad \Rightarrow \quad \boldsymbol{G}_{avg} = \frac{1}{N}\sum_{i}^{N} \boldsymbol{x}_i \cdot \boldsymbol{x}_i^{\mathrm{T}}$$

Dimensionality = (local descriptor size)^2



Second Order Pooling (O2P)

Can we pursue higher-order statistics for pooling ?



Capture correlations

$$g_{avg} = \frac{1}{N}\sum_{i}^{N} x_i$$

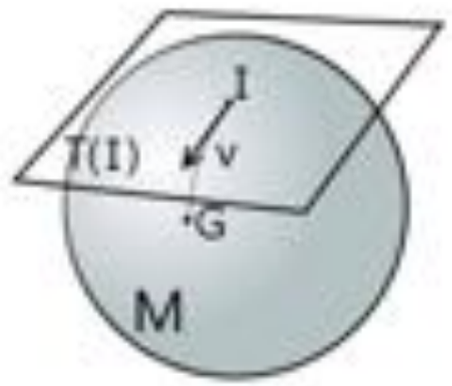
$$G_{avg} = \frac{1}{N}\sum_{i}^{N} x_i \cdot x_i^T$$

Use correct metric

$$G_{log} = \log(G_{avg})$$

Using **Log-Euclidean metric** we can directly embed entire manifold of SPD matrices

Dimensionality = (local descriptor size)^2



Carreira, Caseiro, Batista, Sminchisescu, ECCV12

Caltech 101

Important testbed for coding and pooling




	SIFT-O_2P	eSIFT-O_2P	SPM[1]	LLC[2]	EMK[3]	MP[4]
Accuracy	**79.2**	**80.8**	64.4	73.4	74.5	77.3

1. *Lazebnik et al. '06*
2. *Wang et al. '10*
3. *Bo & Sminchisescu '10*
4. *Boureau et al. '11*

Semantic Segmentation in the Wild Pascal VOC 2011

	comp6		comp5				
	O_2P	Berkeley	BONN-FGT	BONN-SVR	BROOKES	NUS-C	NUS-S
Mean Score	**47.6**	40.8	41.4	43.3	31.3	35.1	37.7
N classes best	**13**	1	2	4	0	0	1

Linear

Exp-Chi² kernels

	Feature Extraction	Prediction	Learning
Exp-Chi²	7.8s / image	87s / image	59h / class
O_2P	4.4s / image	0.004s / image	26m / class
		20,000x faster	**130x faster**

Semantic Segmentation Second Order Pooling

https://www.youtube.com/watch?v=u5Ee0HFboLA

LUND UNIVERSITY

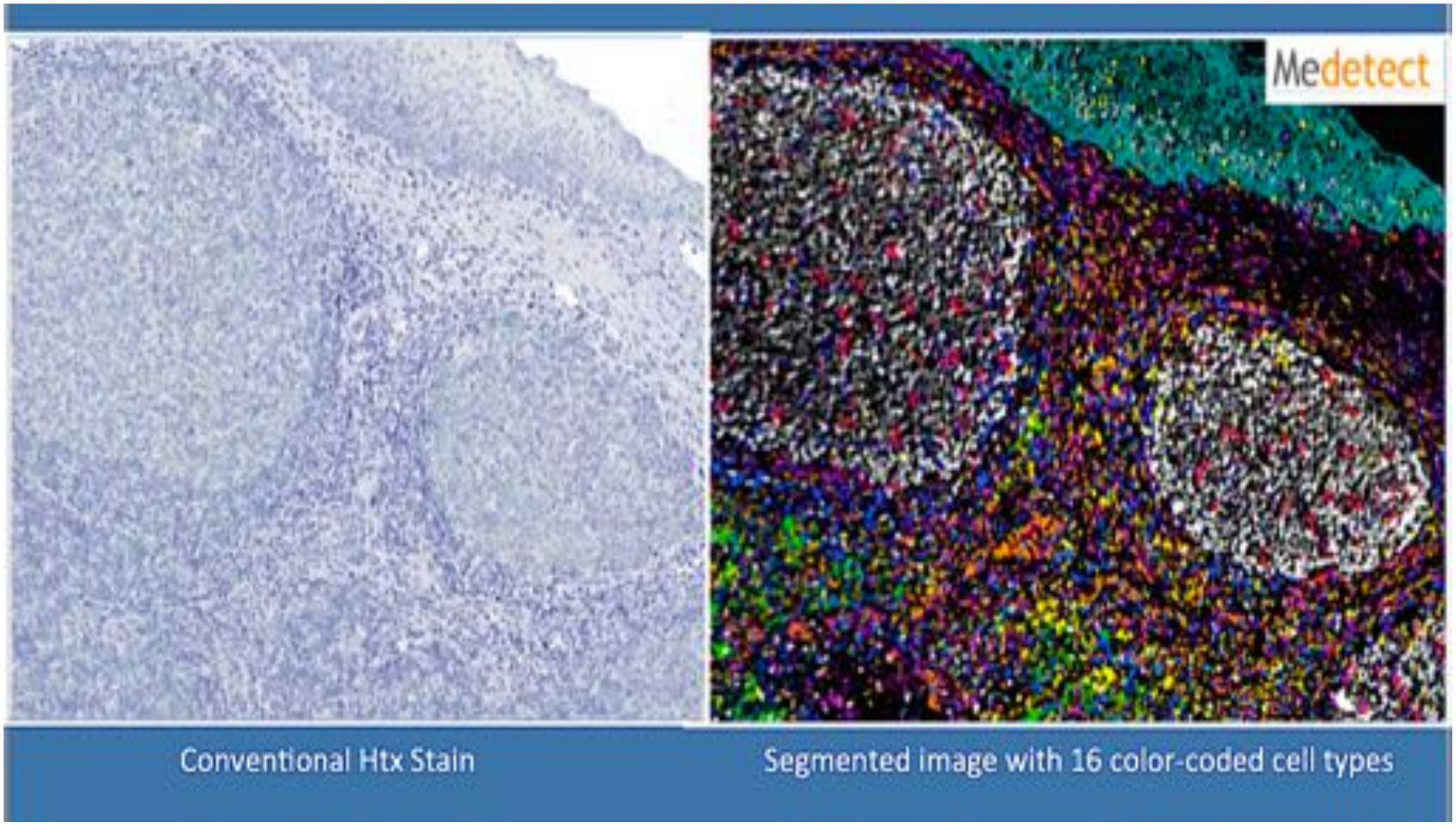
Conclusions

- Significant advances in image and semantic segmentation as well as boundary detection over the past 10 years
- Mature boundary detection methods
- Well-developed figure-ground region proposal generation methods
- Second-order region pooling descriptors available
  - Better description compared to histograms, better performance
  - Fast. Support linear learning methods (no need for expensive exp-chi2 kernels as in histograms/bag-of-words models)

Overview – Semantic Segmentation

1. Edge detectors based on machine learning
2. Segmentation is an ill-posed problem
3. Generating a pool of possible segments (CPMC)
4. Rating segments in the pool
5. Visual and Semantic Processing
6. Second Order Pooling

Master's thesis suggestion
Analysis of AMLC images




LUND
UNIVERSITY