POTTICS - THE POTTS TOPIC MODEL FOR SEMANTIC IMAGE SEGMENTATION

TECHNISCHE UNIVERSITÄT DARMSTADT

MAX-PLANCK-GESELLSCHAFT

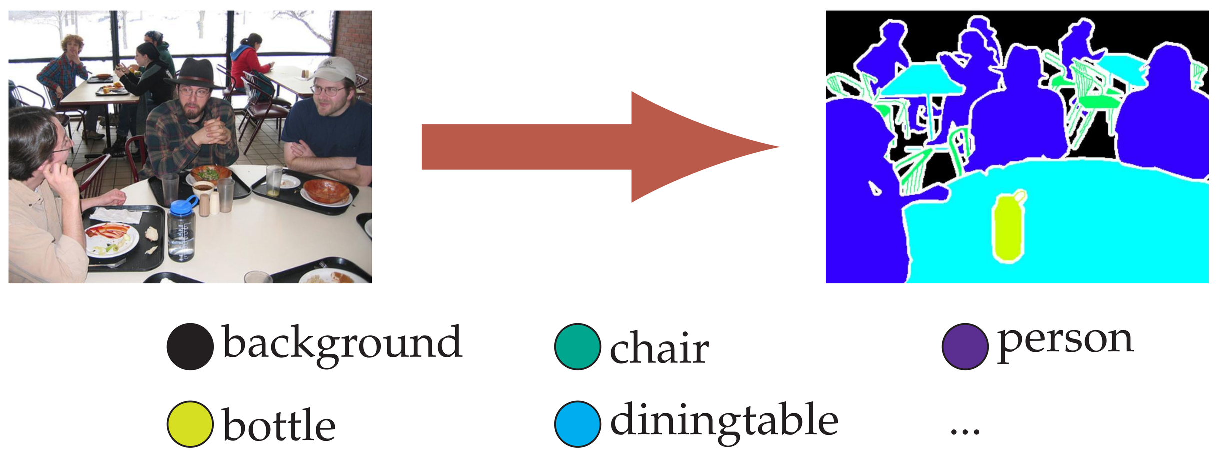
Microsoft Research

Christoph Dann[1], Peter Gehler[2], Stefan Roth[1], Sebastian Nowozin[3]

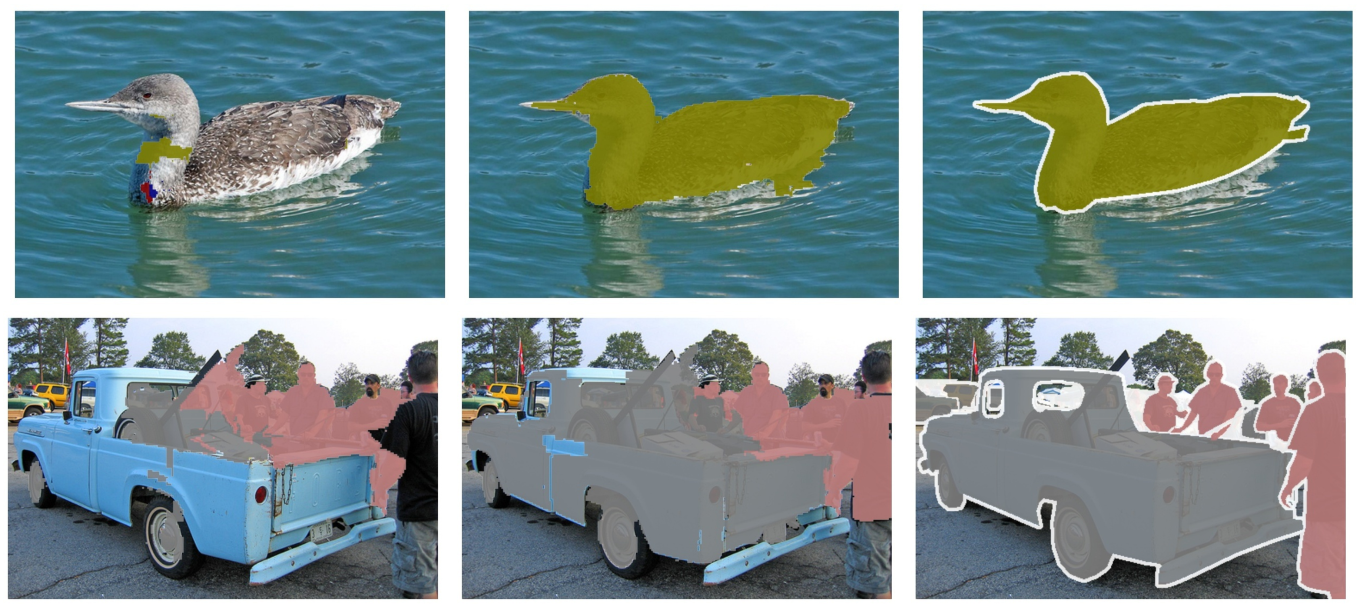
[1] TU Darmstadt, [2] MPI-Intelligent Systems, [3] Microsoft Research

PROBLEM

Semantic Image Segmentation: Assign a semantic class in $\{1, \ldots, C\}$ to each pixel of an image, e.g. *person*, *dog* or *car*, etc.



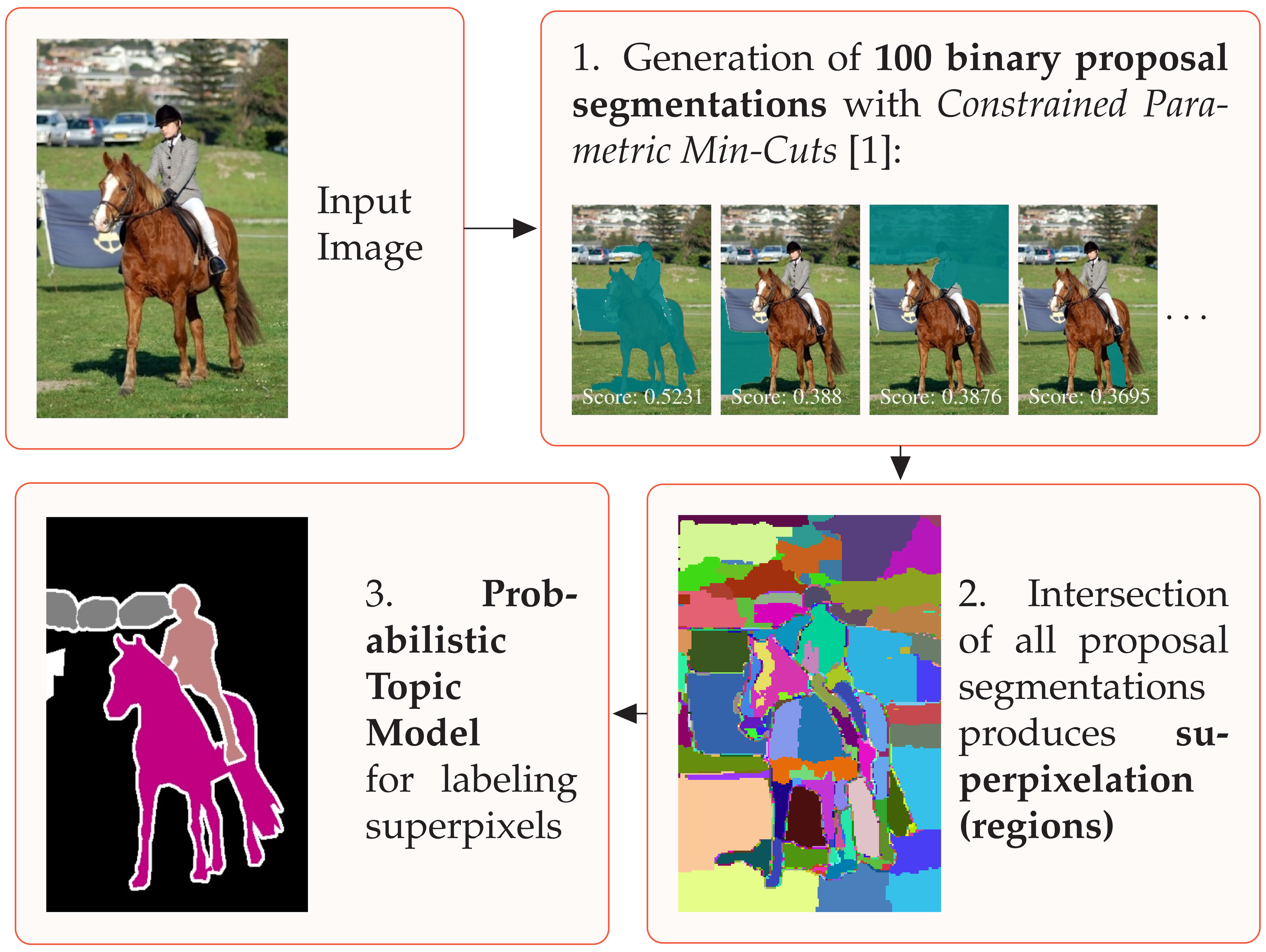
EXAMPLE RESULTS



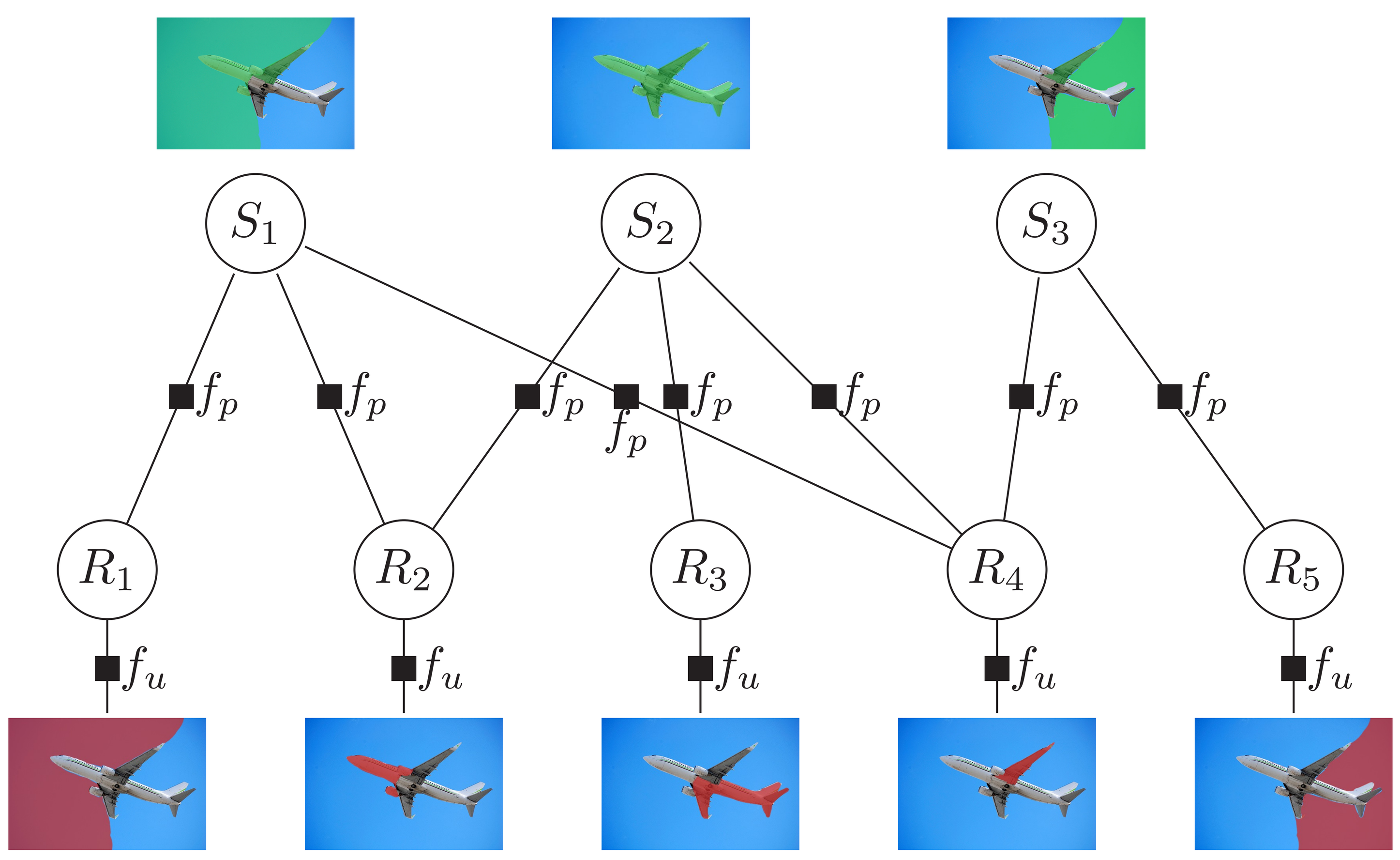
unary-only prediction   Pottics   ground-truth

ALGORITHM OVERVIEW

Main Idea: Capture higher-order spatial relations of segment labels through latent topics, defined on multiple binary figure-ground segmentations (from CPMC [1]) – efficient inference through bipartite graph structure.



PROBABILISTIC TOPIC MODEL



Variables:

- $S_i \in \{1, \ldots, T\}$ (latent) assigns a topic to each binary proposal segmentation
- $R_j \in \{1, \ldots, C\}$ specifies the class label for all pixels in the j^{th} superpixel

Factors:

- f_u: pre-trained unary potential based on TextonBoost; incorporates image information such as color, location, HOG, bounding box detections [3]
- f_p: pairwise potential connects all pairs of binary proposal segmentations and superpixels that overlap: $f_p(S_i, R_j) = w_{S_i, R_j} \rightarrow$ models the consistency of a semantic class and a latent topic

Learning: parameters w_{tc} are learned with a Structured Output SVM [2].

NUMERICAL RESULTS

VOC (2010) %	UO*	RP*	SP*	Potts*	1T*	Pottics
background	80.3	80.5	77.5	78.8	78.7	**80.8**
plane	27.6	27.4	12.8	22.4	10.1	**41.0**
bicycle	0.6	0.6	0.0	0.6	0.1	**3.9**
bird	11.9	11.9	2.3	10.7	2.8	**22.1**
boat	16.0	16.1	4.8	13.7	4.8	**25.3**
bottle	15.2	14.9	2.3	12.7	3.8	**24.2**
bus	33.0	33.1	29.0	32.2	22.4	**41.3**
car	43.3	44.2	37.3	43.2	31.4	**52.8**
cat	28.8	**30.4**	25.4	26.5	25.0	25.3
chair	5.2	5.5	3.4	5.4	2.5	**6.4**
cow	10.7	10.5	2.2	7.8	2.5	**20.2**
table	5.4	4.7	1.0	4.5	2.7	**12.5**
dog	12.2	**12.3**	7.3	11.9	4.5	11.5
horse	16.1	16.3	3.7	13.4	6.1	**18.6**
motorbike	28.4	29.4	20.9	26.9	16.5	**34.7**
person	34.6	35.7	33.2	35.2	28.4	**37.1**
plant	11.0	10.3	8.4	9.7	5.7	**16.2**
sheep	20.0	20.9	9.9	17.4	10.8	**21.0**
sofa	12.8	**12.9**	5.7	11.7	7.4	12.3
train	30.1	30.3	19.1	28.8	18.2	**39.9**
tv	23.2	22.0	8.0	20.9	4.2	**27.6**
total	22.2	22.4	14.8	20.7	13.8	**27.4**

SUMMARY

- Novel CRF model beyond Potts for semantic image segmentation
- Latent topics may have a meaningful interpretation, such as *"bus and car appear together"*
- Substantial improvement over unary-only prediction and simple Gibbs-models
- Efficient inference (here: iterated conditional modes – ICM) as topic assignments and superpixel labels form a bipartite graph

REFERENCES & SOURCE CODE

[1] Carreira, J., Sminchisescu, C.: Constrained Parametric Min-Cuts for Automatic Object Segmentation. In: CVPR (2010)

[2] Yu, C.N., Joachims, T.: Learning Structural SVMs with Latent Variables. In: ICML (2009)

[3] Krähenbühl, P., Koltun, V.: Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In: NIPS (2011)

The source code is available at `http://github.com/chrodan/pottics`

* (UO) Pixel prediction [3]; (RP) f_u-only prediction; (SP) Unary-potential for pixels in a proposal segmentation; (Potts) Potts-Model on superpixels; (1T) Pottics restricted to 1 topic