

Computer vision and machine learning at Adelaide

Chunhua Shen

Australian Centre for Robotic Vision; and
School of Computer Science, The University of Adelaide



THE UNIVERSITY
of ADELAIDE

Australian Centre for Visual Technologies

- Largest computer vision centre at Australia, with ~70 staff and PhD students, including:
 - 4 full professors
 - 7 tenure-track/tenured staff
- Main hub of two major Gov. projects:
 - ARC Centre of Excellence for Robotic Vision (\$20M, 7 yrs)
 - Data to Decisions CRC Centre (\$25M, 5 yrs)

My team at Adelaide: 20+ PhD students and Postdoc researchers (4 more joining in 2015)

Ⓜ A formal, brief biography

Chunhua Shen is a Professor at School of Computer Science, [University of Adelaide](#). He is a Project Leader and Chief Investigator at the Australian Research Council Centre of Excellence for Robotic Vision ([ACRV](#)), for which he leads the project on machine learning for robotic vision. Before he moved to Adelaide as a Senior Lecturer, he was with the computer vision program at [NICTA](#) (National ICT Australia), Canberra Research Laboratory for about six years. His research interests are in the intersection of computer vision and statistical machine learning. Recent work has been on real-time object detection, large-scale image retrieval and classification, and scalable nonlinear optimization.

He studied at Nanjing University, at [Australian National University](#), and received his PhD degree from the [University of Adelaide](#). From 2012 to 2016, he holds an Australian Research Council Future Fellowship. He is serving as Associate Editor of IEEE Transactions on Neural Networks and Learning Systems.

「中文简介」

沈春华博士现任澳大利亚阿德莱德大学计算机科学学院教授(终身教职)。2011之前在澳大利亚国家信息通讯技术研究院堪培拉实验室的计算机视觉组工作近6年。目前主要从事统计机器学习以及计算机视觉领域的研究工作。主持多项科研课题,在重要国际学术会议上发表论文100余篇。

沈春华博士曾在南京大学(本科及硕士)、澳大利亚国立大学(硕士)学习,并在阿德莱德大学获得计算机视觉方向的博士学位。2012年被澳大利亚研究理事会(Australian Research Council)授予Future Fellowship。

www.cs.adelaide.edu.au/~chhshen



THE UNIVERSITY
of ADELAIDE

GROUP MEMBERS

Supervision of PhD students and Postdoctoral researchers:

Graduate students

Bohan Zhuang	Adelaide
Qichang Hu	Adelaide
Yuanzhouhan Cao	Adelaide
Ruizhi Qiao	Adelaide
Fayao Liu	Adelaide
Josh Boys	Adelaide ; co-supervised with Anton van den Her
Yao Li	Adelaide
Teng Li	Adelaide
Junjie Zhang	Adelaide
Hui Li	Adelaide
Yuchao Jiang	Adelaide ; jointly supervised with Ian Reid
Chamara Saroj Weerasekera	Adelaide ; jointly supervised with Ian Reid
Chongyu Liu	Adelaide ; jointly supervised with Qinfeng Shi
Bo Li	Visiting from Northwestern Polytechnical Univer
Peng Wang	Visiting from University of Queensland , 2014---20
Zetao Chen	Visiting from Queensland University of Technolo
Zongyuan Ge	Visiting from Queensland University of Technolo
Lei Zhang	Visiting from Northwestern Polytechnical Univer

Postdoctoral researchers

Dr. Peng Wang	Project: MRF inference, SDP optimisation, object detection (D2L)
Dr. Lingqiao Liu	Project: Large-scale image classification (D2DCRC), medical ima
Dr. Lin Wu	Project: Video surveillance (ACRV); jointly supervised with Ant
Dr. Qi Wu	Project: Image captioning (D2DCRC); co-supervised with Anton
Dr. Zifeng Wu	Project: Large-scale image classification (D2DCRC); co-supervis
Dr. Guosheng Lin	Project: Large-scale image classification, deep learning (ACRV and den Hengel)

Post-doctoral researcher positions to work on large-scale image classification using deep learning

 Posted on [April 1, 2015](#) by [Ian](#)

The Australian Centre for Visual Technologies (ACVT) has two vacancies for post-doctoral researchers to work on deep learning with applications to large-scale image classification, such as on the ImageNet dataset. Successful applicants will have the opportunity to work with the world-class team of researchers on some of the most fundamental, and challenging problems in Computer Vision.

This is an opportunity for a researcher with a background in machine learning methods, in particular, deep learning such as Convolutional Neural Networks, to apply their experience to large-scale problems in image understanding.

Researchers will work with members of the ACVT, including Prof. Anton van den Hengel, and Prof Chunhua Shen, and will join a large, and growing, group of world-class researchers working in this rapidly developing area. This position forms part of the ACVT involvement in the Data 2 Decisions Cooperative Research Centre.

Duration: 2 years initially, but with the option to extend

Start date: as soon as possible (posted April 2015).

<http://tinyurl.com/pjhx8dc>

PhD scholarships available too!

Glenelg beach: 9km from UofA



Henley beach: 9.7km from UofA



Brighton beach: 15km from UofA





top 10 most liveable cities 2014

1. Melbourne, Australia
2. Vienna, Austria
3. Vancouver, Canada
4. Toronto, Canada
5. Adelaide, Australia



Acknowledgements: most of the hard work was done by my (ex-) students and postdocs. Credit goes to them. Among many others, in particular I'd mention:

- Guosheng Lin (2011~present, now postdoc)
- Fayo Liu (2011~present, PhD student)
- Yao Li (2013~present, PhD student)
- Lingqiao Liu (2010~present, now postdoc)
- Sakrapee Paul Paisitkriangkrai (2006~2015; departed)
- Peng Wang (2008~present, now postdoc)

Agenda

1. What we did: boosting, sdp, etc.
2. What we are doing:
 - deep learning
 - structured output learning
 - deep structured output learning
3. Future work

Boosting

- Boosting builds a very accurate classifier by combining rough and only moderately accurate classifiers.
- Boosting procedures
 - Given a set of labeled training examples
 - On each round
 - 1 The booster devises a distribution (importance) over the example set
 - 2 The booster requests a weak hypothesis/classifier/learner with low error
 - Upon convergence, the booster combine the weak hypothesis into a single prediction rule.

Why boosting works

- Let \mathcal{H} be a class of base classifier
 $\mathcal{H} = \{h_j(\cdot) : \mathcal{X} \rightarrow \mathbb{R}\}, j = 1 \cdots N$, a boosting algorithm seeks for a convex combination:

$$F(\mathbf{w}) = \sum_{j=1}^N w_j h_j(\mathbf{x})$$

- Statistical view [Friedman et al. 2000], maximum margin [Schapire et al. 1998], still there are open questions [Mease & Wyner 2008]
- The Lagrange dual problems of AdaBoost, LogitBoost and soft-margin LPBoost with generalized hinge loss are all entropy maximization problems [Shen & Li 2010 TPAMI]

A duality view of boosting

Explicitly find a **meaningful** Lagrange dual for some boosting algorithms

Dual of AdaBoost

The Lagrange dual of AdaBoost is a Shannon entropy maximization problem:

$$\max_{r, \mathbf{u}} \frac{r}{T} - \overbrace{\sum_{i=1}^M u_i \log u_i}^{\text{reg. in dual}}, \quad \text{s.t.} \sum_{i=1}^M y_i u_i H_i \leq -r \mathbf{1}^\top, \mathbf{u} \geq 0, \mathbf{1}^\top \mathbf{u} = 1.$$

Here $H_i = [H_{i1} \dots H_{iN}]$ denotes i -th row of H , which constitutes the output of all weak classifiers on \mathbf{x}_i .

A duality view of boosting

Primal of AdaBoost (Note the auxiliary variables $z_i, i = 1, \dots$)

$$\begin{aligned} \min_{\mathbf{w}} \quad & \log \left(\sum_{i=1}^M \exp z_i \right), \\ \text{s.t.} \quad & z_i = -y_i H_i \mathbf{w} \quad (\forall i = 1, \dots, M), \\ & \mathbf{w} \geq 0, \mathbf{1}^\top \mathbf{w} = \frac{1}{T}. \end{aligned}$$

Dual of boosting algorithms are entropy regularized LPBoost.

algorithm	loss in primal	entropy reg. LPBoost in dual
adaboost	exponential loss	Shannon entropy
logitboost	logistic loss	binary relative entropy
soft-margin $\ell_p(p > 1)$ LPBoost	generalized hinge loss	Tsallis entropy

Average margin vs. margin variance

Why AdaBoost just works?

Theorem:

AdaBoost approximately maximizes the **average margin** and at the same time minimizes the **variance of the margin distribution** under the assumption that the margin follows a Gaussian distribution.

Proof: See [Shen & Li 2010 TPAMI]. Main tools used:

- 1 Central limit theorem;
- 2 Monte Carlo integral.

Average margin vs. margin variance

What this theorem tells us:

- ① We should focus on optimizing the overall **margin distribution**. Almost all previous work on boosting has focused on a large **minimum margin**.
- ② Answered an open question in [Reyzin & Schapire 2006], [Mease & Wyner 2008]
- ③ We can design new boosting algorithm to directly maximize the average margin and minimize the margin variance [Shen & Li, 2010 TNN]

Margin distribution boosting

$$\max_{\mathbf{w}} \bar{\rho} - \frac{1}{2}\sigma^2, \text{ s.t. } \mathbf{w} \geq 0, \mathbf{1}^\top \mathbf{w} = T.$$

It is equivalent to

$$\begin{aligned} \min_{\mathbf{w}, \boldsymbol{\rho}} \quad & \frac{1}{2} \boldsymbol{\rho}^\top A \boldsymbol{\rho} - \mathbf{1}^\top \boldsymbol{\rho}, \\ \text{s.t. } \quad & \mathbf{w} \geq 0, \mathbf{1}^\top \mathbf{w} = T, \\ & \rho_i = y_i H_i \mathbf{w}, \forall i = 1, \dots, M. \end{aligned}$$

Its dual is

$$\min_{r, \mathbf{u}} \quad r + 1/(2T) (\mathbf{u} - \mathbf{1})^\top A^{-1} (\mathbf{u} - \mathbf{1}), \text{ s.t. }, \sum_{i=1}^M y_i u_i H_i \leq r \mathbf{1}^\top.$$

Fully corrective boosting for regularised risk minimisation

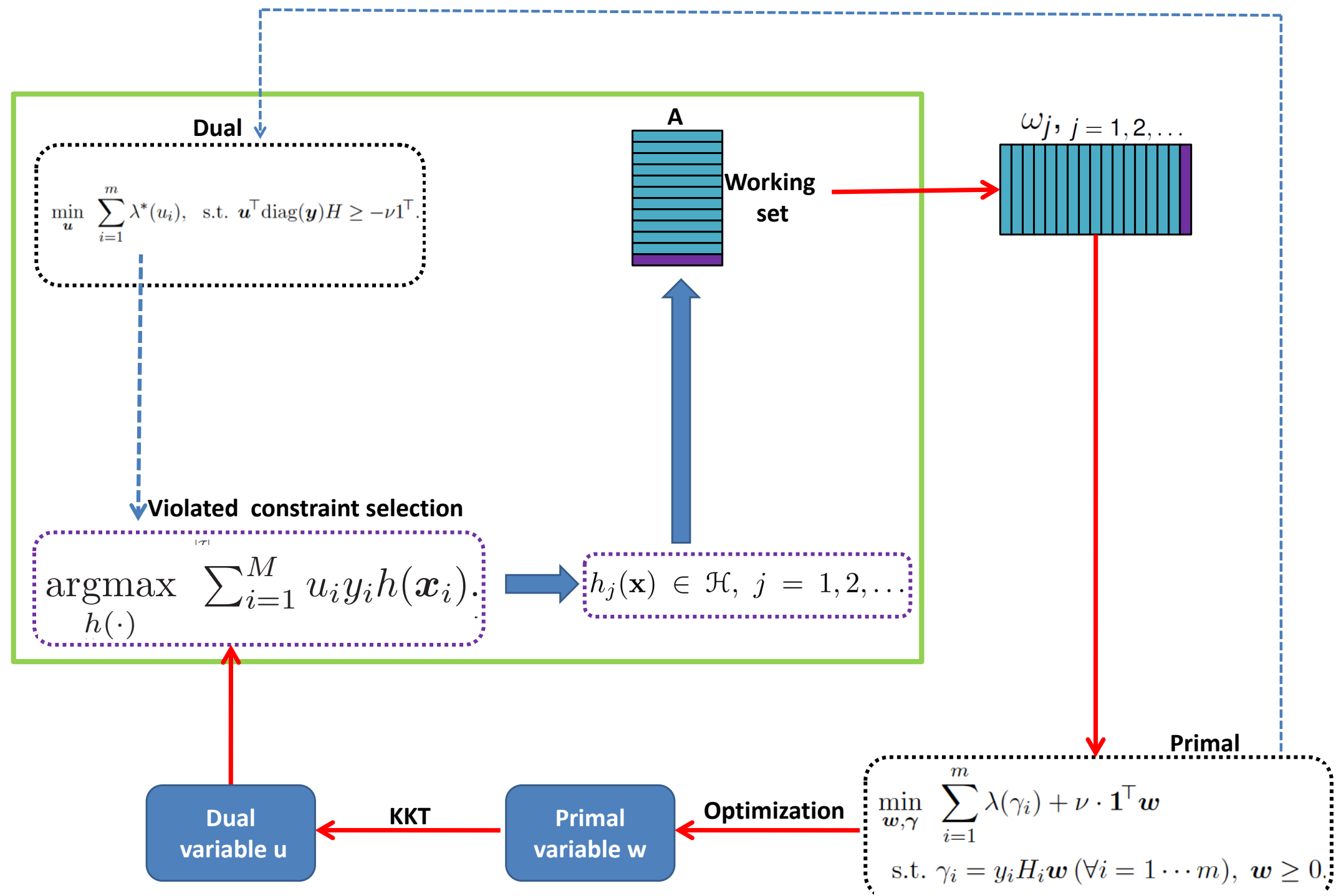
- ① A general framework that can be used to designed new boosting algorithms.
- ② The proposed boosting framework, termed CGBoost, can accommodate various loss functions and different regularizers in a totally-corrective optimization way.

Boosting via column generation

- ① Samples' margins γ and weak classifiers' clipped edges \mathbf{d}^+ are dual to each other.
- ② ℓ_p regularization in primal corresponds to ℓ_q regularization in dual with $1/p + 1/q = 1$.

	Primal	Dual
ℓ_1	$\min \sum_{i=1}^m \lambda(\gamma_i) + \nu \ \mathbf{w}\ _1$	$\min \sum_{i=1}^m \lambda^*(-u_i) + r \ \mathbf{d}^+\ _\infty$
ℓ_2	$\min \sum_{i=1}^m \lambda(\gamma_i) + \nu \ \mathbf{w}\ _2^2$	$\min \sum_{i=1}^m \lambda^*(-u_i) + r \ \mathbf{d}^+\ _2^2$
ℓ_∞	$\min \sum_{i=1}^m \lambda(\gamma_i) + \nu \ \mathbf{w}\ _\infty$	$\min \sum_{i=1}^m \lambda^*(-u_i) + r \ \mathbf{d}^+\ _1$
	$\lambda(\gamma)$: loss in primal	$\ \mathbf{d}^+\ _q$: loss in dual
	$\ \mathbf{w}\ _p$: regularization in primal	$\lambda^*(\mathbf{u})$: regularization in dual

Boosting via column generation



Boosting via column generation

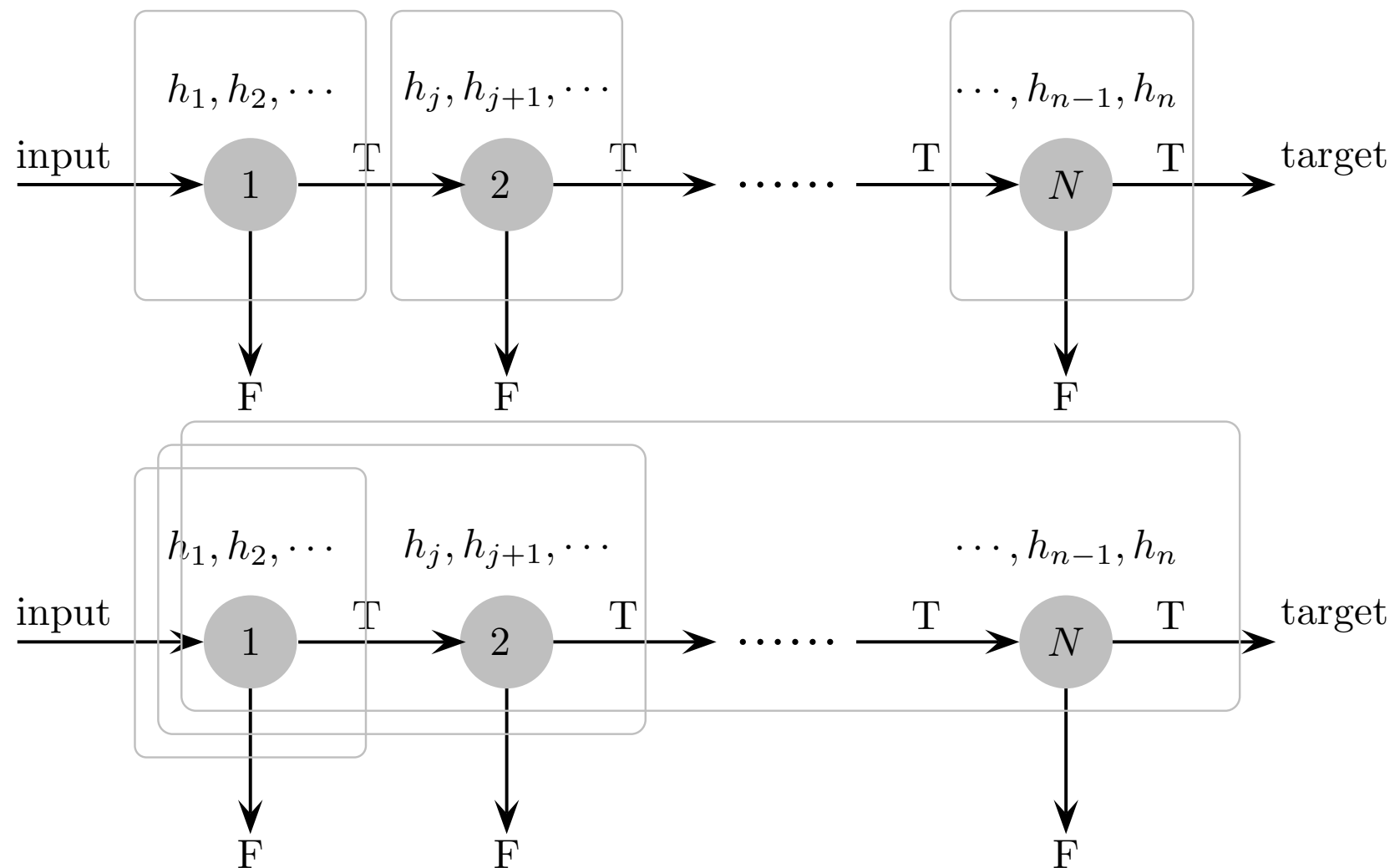
- We now have a general framework for designing *fully-corrective* boosting methods, to minimise *arbitrary*:
 - convex loss + convex regularisation
- It converges faster with *on par* test accuracy compared with conventional stage-wise boosting (such as AdaBoost, logistic boosting)

Refs: TPAMI2010, TNN2010, NN2013

Applications of this general boosting framework

1:

Cascade classifiers (1) standard cascade (2) multi-exit cascade.
Only those classified as true detection by all nodes will be true targets.



Boosting for node classifier learning

Biased Minimax Probability Machines:

$$\max_{\boldsymbol{w}, b, \gamma} \gamma \quad \text{s.t.} \quad \begin{cases} \left[\inf_{\boldsymbol{x}_1 \sim (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)} \Pr\{\boldsymbol{w}^\top \boldsymbol{x}_1 \geq b\} \right] \geq \gamma, \\ \left[\inf_{\boldsymbol{x}_2 \sim (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)} \Pr\{\boldsymbol{w}^\top \boldsymbol{x}_2 \leq b\} \right] \geq \gamma_0. \end{cases}$$

Let's consider a special case: $\gamma_0 = 0.5$: The 2nd class will have a classification accuracy around 50%.

Refs: ECCV2010, IJCV2013

2: Direct approach to Multi-class boosting; sharing features in multi-class boosting

We generalize this idea to the entire training set and introduce slack variables ξ to enable soft-margin. The primal problem that we want to optimize can then be written as

$$\begin{aligned} \min_{W, \xi} \quad & \sum_{i=1}^m \xi_i + \nu \|W\|_1 \\ \text{s.t.} \quad & \delta_{r, y_i} + H_i: \mathbf{w}_{y_i} \geq 1 + H_i: \mathbf{w}_r - \xi_i, \forall i, r, \\ & W \geq 0. \end{aligned}$$

$$\cdot \nu \|W\|_{1,2}$$

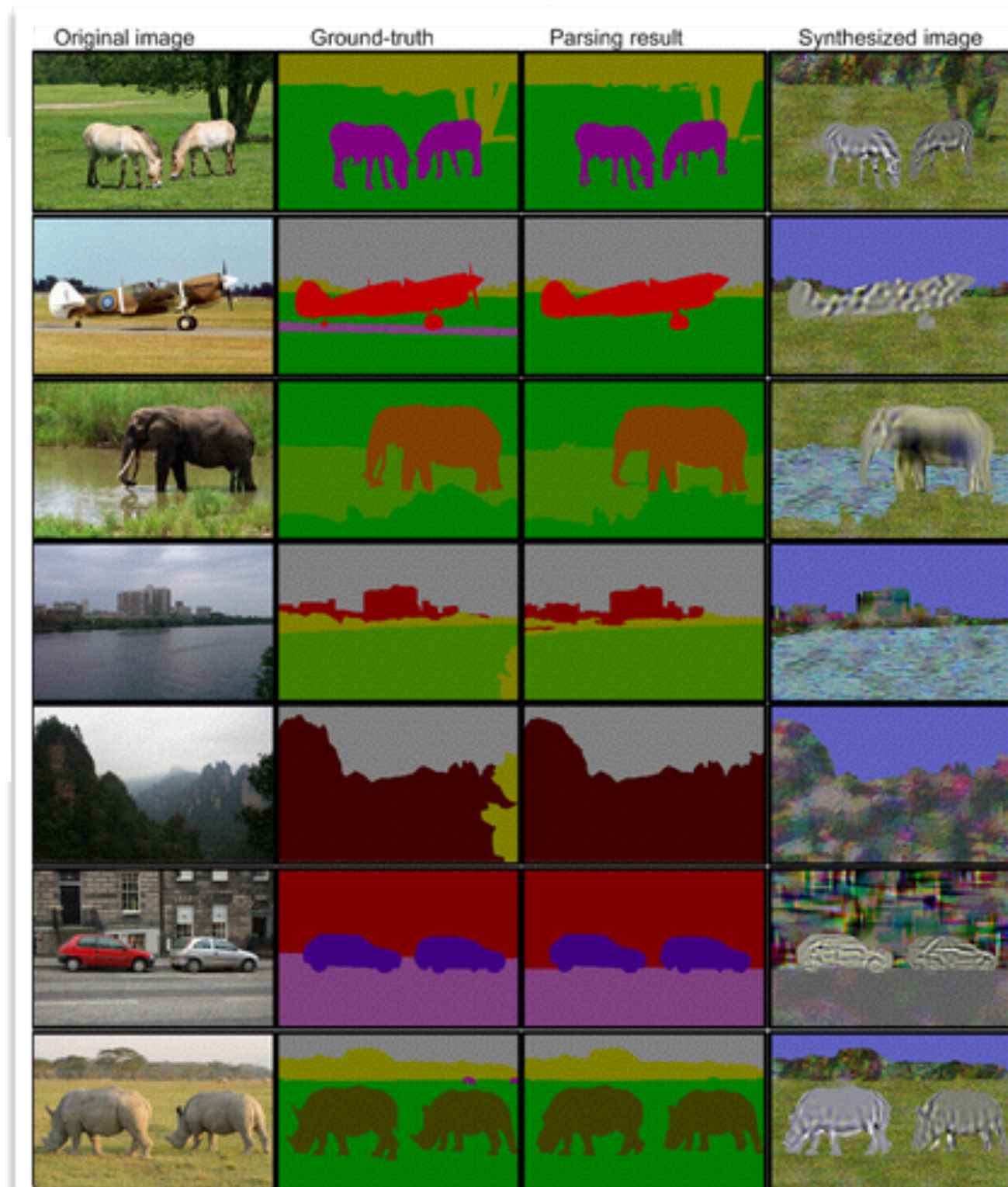
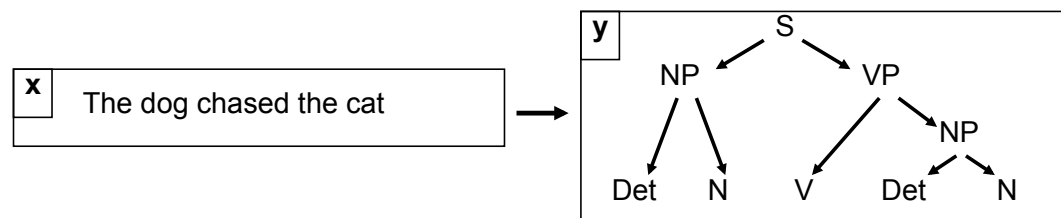
Here $\nu > 0$ is the regularization parameter.

3: Structured output boosting

Natural Language Parsing

Given a sequence of words x , predict the parse tree y .

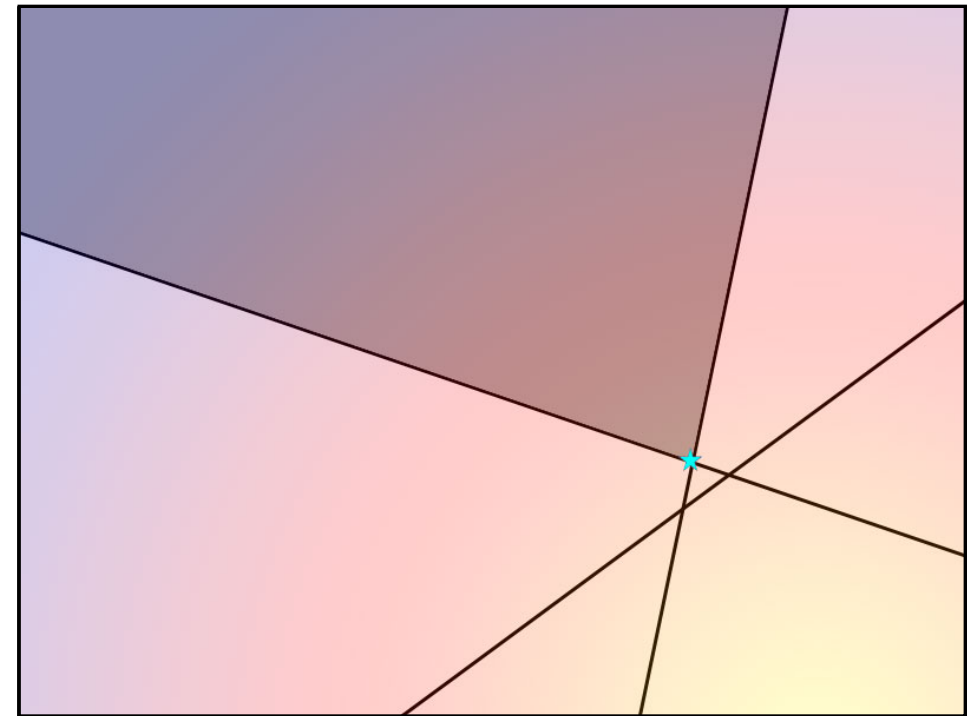
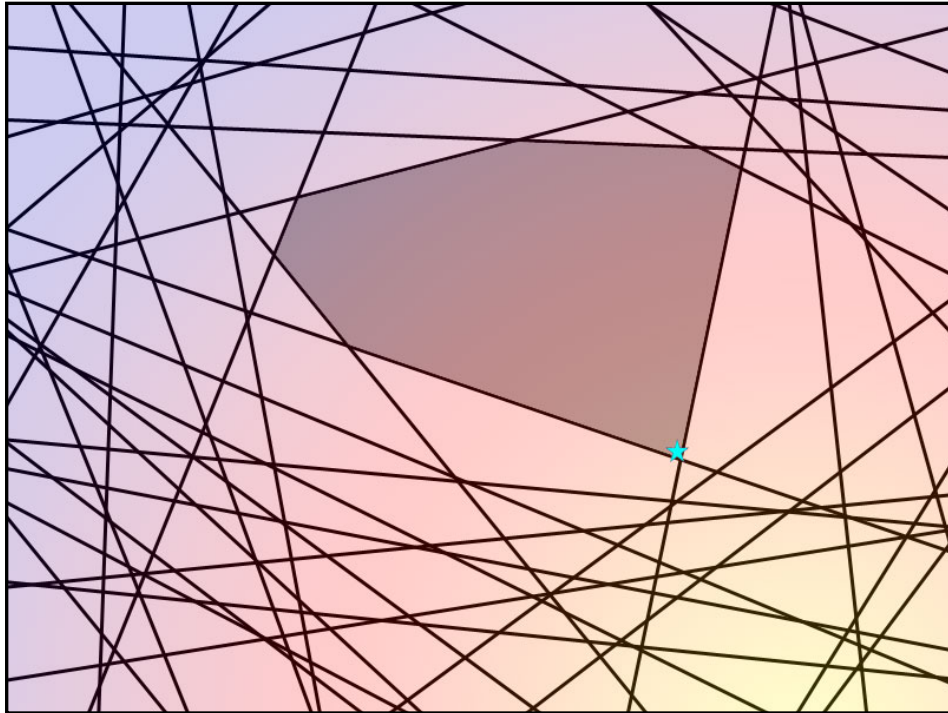
Dependencies from structural constraints, since y has to be a tree.



Structured SVM

Original SVM Problem

- Exponential constraints
- Most are dominated by a small set of “important” constraints



Structural SVM Approach

- Repeatedly finds the next most violated constraint...
- ...until set of constraints is a good approximation.

This is so-called the “cutting plane” method

Structured Boosting

- The discriminant function we want to learn:

$$F : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$$

structured weak learner

$$F(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w}) = \boldsymbol{w}^\top \Psi(\boldsymbol{x}, \boldsymbol{y}) = \sum_j w_j \psi_j(\boldsymbol{x}, \boldsymbol{y})$$

with $\boldsymbol{w} \geq 0$. As in other structured learning models, the process for predicting a structured output (or inference) is to find an output \boldsymbol{y} that maximizes the joint compatibility function:

$$\boldsymbol{y}^* = \operatorname{argmax}_{\boldsymbol{y}} F(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{w}) = \operatorname{argmax}_{\boldsymbol{y}} \boldsymbol{w}^\top \Psi(\boldsymbol{x}, \boldsymbol{y}).$$

Structured Boosting

Primal:

$$\begin{aligned} \min_{\mathbf{w} \geq 0, \boldsymbol{\xi} \geq 0} \quad & \mathbf{1}^\top \mathbf{w} + \frac{C}{m} \mathbf{1}^\top \boldsymbol{\xi} \\ \text{s.t.} \quad & \mathbf{w}^\top \left[\Psi(\mathbf{x}_i, \mathbf{y}_i) - \Psi(\mathbf{x}_i, \mathbf{y}) \right] \geq \Delta(\mathbf{y}_i, \mathbf{y}) - \xi_i \\ & \forall i = 1, \dots, m; \text{ and } \forall \mathbf{y} \in \mathcal{Y}. \end{aligned}$$

- Exponentially many variables and constraints
- More challenging than structured SVM and boosting

Structured Boosting

- Let's put aside the difficulty of many constraints in the primal, and using the CG framework to design boosting

Dual:

$$\begin{aligned} \max_{\mu \geq 0} \quad & \sum_{i, \mathbf{y}} \mu_{(i, \mathbf{y})} \Delta(\mathbf{y}_i, \mathbf{y}) \\ \text{s.t.} \quad & \sum_{i, \mathbf{y}} \mu_{(i, \mathbf{y})} \delta \Psi_i(\mathbf{y}) \leq \mathbf{1}, \\ & 0 \leq \sum_{\mathbf{y}} \mu_{(i, \mathbf{y})} \leq \frac{C}{m}, \forall i = 1, \dots, m. \end{aligned}$$

Structured Boosting

Algorithm 1 Column generation for StructBoost

- 1: **Input:** training examples $(\mathbf{x}_1; \mathbf{y}_1), (\mathbf{x}_2; \mathbf{y}_2), \dots$; parameter C ; termination threshold ϵ_{cg} , and the maximum iteration number.
 - 2: **Initialize:** for each i , ($i = 1, \dots, m$), randomly pick any $\mathbf{y}_i^{(0)} \in \mathcal{Y}$, initialize $\mu_{(i, \mathbf{y})} = \frac{C}{m}$ for $\mathbf{y} = \mathbf{y}_i^{(0)}$, and $\mu_{(i, \mathbf{y})} = 0$ for all $\mathbf{y} \in \mathcal{Y} \setminus \mathbf{y}_i^{(0)}$.
 - 3: **Repeat**
 - 4: — Find and add a weak structured learner $\psi^*(\cdot, \cdot)$ by solving the subproblem (7) or (11).
 - 5: — Call Algorithm 2 to obtain \mathbf{w} and μ . Cutting plane
 - 7: **Until** either (8) is met or the maximum number of iterations is reached.
 - 8: **Output:** the discriminant function $F(\mathbf{x}, \mathbf{y}; \mathbf{w}) = \mathbf{w}^\top \Psi(\mathbf{x}, \mathbf{y})$.
-

Ref: TPAMI2014, <http://arxiv.org/abs/1302.3283>

#4: Optimising ROC curves for pedestrian detection (ICCV'13)



Approach	INRIA	ETH	TUD-Brussels	Caltech-USA
Sketch tokens [16] (Prev. best on INRIA [†])	13.3%	N/A	N/A	N/A
DBN-Mut [19] (Prev. best on ETH [†])	N/A	41.1%	N/A	48.2%
MultiFtr+Motion+2Ped [18] (Prev. best on TUD-Brussels)	N/A	N/A	50.5%	N/A
SDtSVM [20] (Prev. best on Caltech-USA)	N/A	N/A	N/A	36.0%
Roerei [1] (2-nd best on INRIA [†] & ETH [†])	13.5%	43.5%	64.0%	48.4%
Ours (sp-Cov+M+O+LUV+LBP)	11.2%	36.5%	43.2%	29.4%
Ours (sp-Cov+M+O+LUV+LBP + pAUC ^{struct})	10.9%	36.2%	43.2%	29.2%

#5: Learning hash functions using column generation (and cutting planes) (ICML'13; ECCV'2014)

Hashing functions *vs.* weak learners in boosting
Using a triplet based loss function for NN search (ICML'13)

With Structured output boosting, we can also optimise a ranking based loss (multivariate measure), ECCV'14

Some not-that-relevant work on hashing: CVPR'13,14,15; ICCV'13, TPAMI'14

Boosting-like scalable semidefinite programming

To learn a p.s.d. X ,

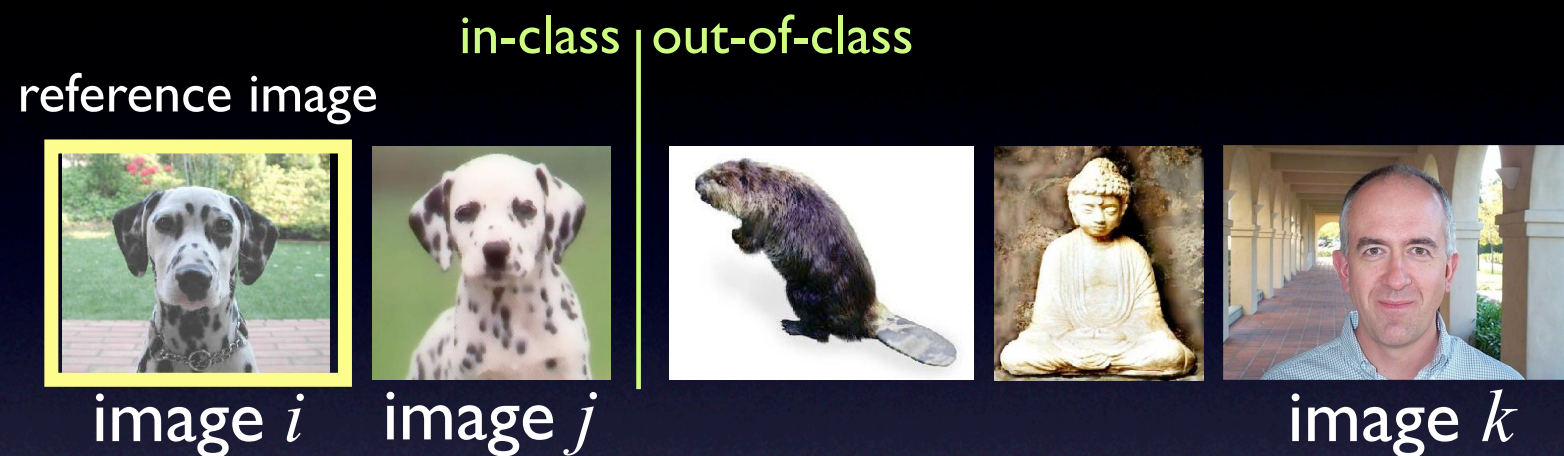
$$0.2 \text{ --- } + 0.3 \text{ / } + 0.5 \text{ | } = \begin{pmatrix} 0.35 & 0.15 \\ 0.15 & 0.65 \end{pmatrix} = \text{ellipse} = 0.29 \text{ \textcolor{green}{/}} + 0.71 \text{ \textcolor{red}{/}}$$

$$X = \sum_i w_i Z_i, \quad \text{with } w_i > 0, \text{ rank}(Z_i) = 1, \text{ trace}(Z_i) = 1$$

Here weak learners are rank-1 trace-1 matrices, instead of classifiers/
regressors.

Boosting-like scalable semidefinite programming

ranking: learn from **triplets** of training images



$$D\left(\begin{array}{c} \text{man} \\ \text{image } k \end{array}, \begin{array}{c} \text{dog} \\ \text{image } i \end{array}\right) > D\left(\begin{array}{c} \text{dog} \\ \text{image } j \end{array}, \begin{array}{c} \text{dog} \\ \text{image } i \end{array}\right)$$
$$D_{ki} > D_{ji}$$

Refs: NIPS'08,09, JMLR'13

Scalable semidefinite programming

$$\begin{aligned} \min_{X, \xi} \quad & \text{Tr}(X) + \frac{C_2}{m} \sum_{r=1}^m \xi_r \\ \text{s.t.} \quad & \langle A_r, X \rangle \geq 1 - \xi_r, r = 1, \dots, m, \\ & \xi \succcurlyeq 0, X \succcurlyeq 0. \end{aligned}$$

from L1 to L2 (trace to Frobenius)

$$\begin{aligned} \min_{X, \xi} \quad & \frac{1}{2} \|X\|_F^2 + \frac{C_3}{m} \sum_{r=1}^m \xi_r \\ \text{s.t.} \quad & \langle A_r, X \rangle \geq 1 - \xi_r, r = 1, \dots, m, \\ & \xi \succcurlyeq 0, X \succcurlyeq 0. \end{aligned}$$

Scalable semidefinite programming

- The original dual problem can be simplified into

$$\max_{\mathbf{u}} \sum_{r=1}^m u_r - \frac{1}{2} \|(\hat{A})_-\|_F^2, \text{ s.t. } \frac{C_3}{m} \succcurlyeq \mathbf{u} \succcurlyeq 0.$$

Ref:

Shen et al. CVPR2011

with $\hat{A} = -\sum_{r=1}^m u_r A_r$.

- Now no matrix variable; no p.s.d constraint!
- The objective function is first-order differentiable but not second-order differentiable \longrightarrow Quasi-Newton like L-BFGS-B applicable
- L-BFGS-B converges in 20 to 30 iterations in all experiments
- Computational complexity is $O(t \cdot D^3)$, $t \in [20, 30]$.
- Much more scalable: $O(D^{6.5}) \longrightarrow O(t \cdot D^3)$

Scalable semidefinite programming

Given a convex optimization problem, it's beneficial to study its dual problem

$$\min_{\mathbf{x}} \mathbf{x}^\top \mathbf{A} \mathbf{x}, \text{ s.t. } \mathbf{x} \in \{-1, 1\}^n$$

$\mathbf{A} \in \mathcal{S}_n$ symmetric matrix but not necessary p.s.d. : Binary quadratic problem

variable must be binary: NP hard

$$\min_{\mathbf{x}} \mathbf{x}^\top \mathbf{A} \mathbf{x}, \text{ s.t. } \mathbf{x} \in \{-1, 1\}^n$$

SDP relaxation: Introducing: $\mathbf{X} = \mathbf{x} \mathbf{x}^\top$

$$\min_{\mathbf{X} \succeq \mathbf{0}} \langle \mathbf{X}, \mathbf{A} \rangle, \text{ s.t. } \text{diag}(\mathbf{X}) = \mathbf{e}, \text{rank}(\mathbf{X}) = 1.$$

Scalable semidefinite programming

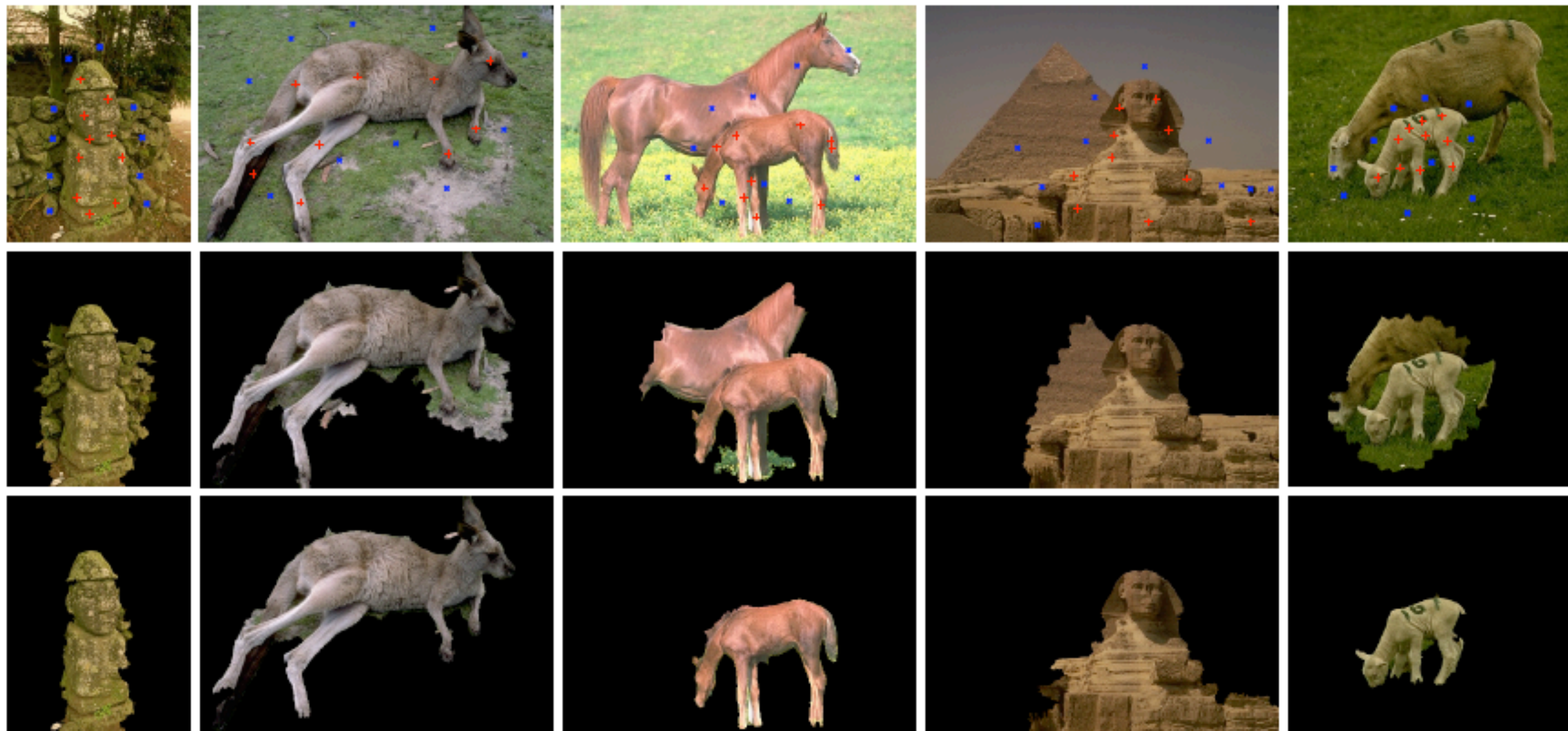
Faster sdp formulation in the dual

$$\min_{\mathbf{X} \succeq \mathbf{0}} \langle \mathbf{X}, \mathbf{A} \rangle, \quad \text{s.t.} \quad \|\mathbf{X}\|_F^2 - \eta^2 \leq 0$$

$$\min_{\mathbf{X} \succeq \mathbf{0}} \langle \mathbf{X}, \mathbf{A} \rangle + \sigma(\|\mathbf{X}\|_F^2 - \eta^2)$$

$$\begin{aligned} \max_{\mathbf{u}} \quad & -\frac{1}{4\sigma} \|\mathbf{C}(\mathbf{u}) - \mathbf{b}\|_F^2 - \sigma\eta^2, \\ \text{s.t.} \quad & u_j \geq 0, \quad \forall j = p+1, \dots, m, \\ & \mathbf{C}(\mathbf{u}) = \sum_{i=1}^m u_i \mathbf{B}_i + \mathbf{A}. \end{aligned}$$

Scalable semidefinite programming



Ref: CVPR2013

Scalable semidefinite programming: Fully connected CRF inference

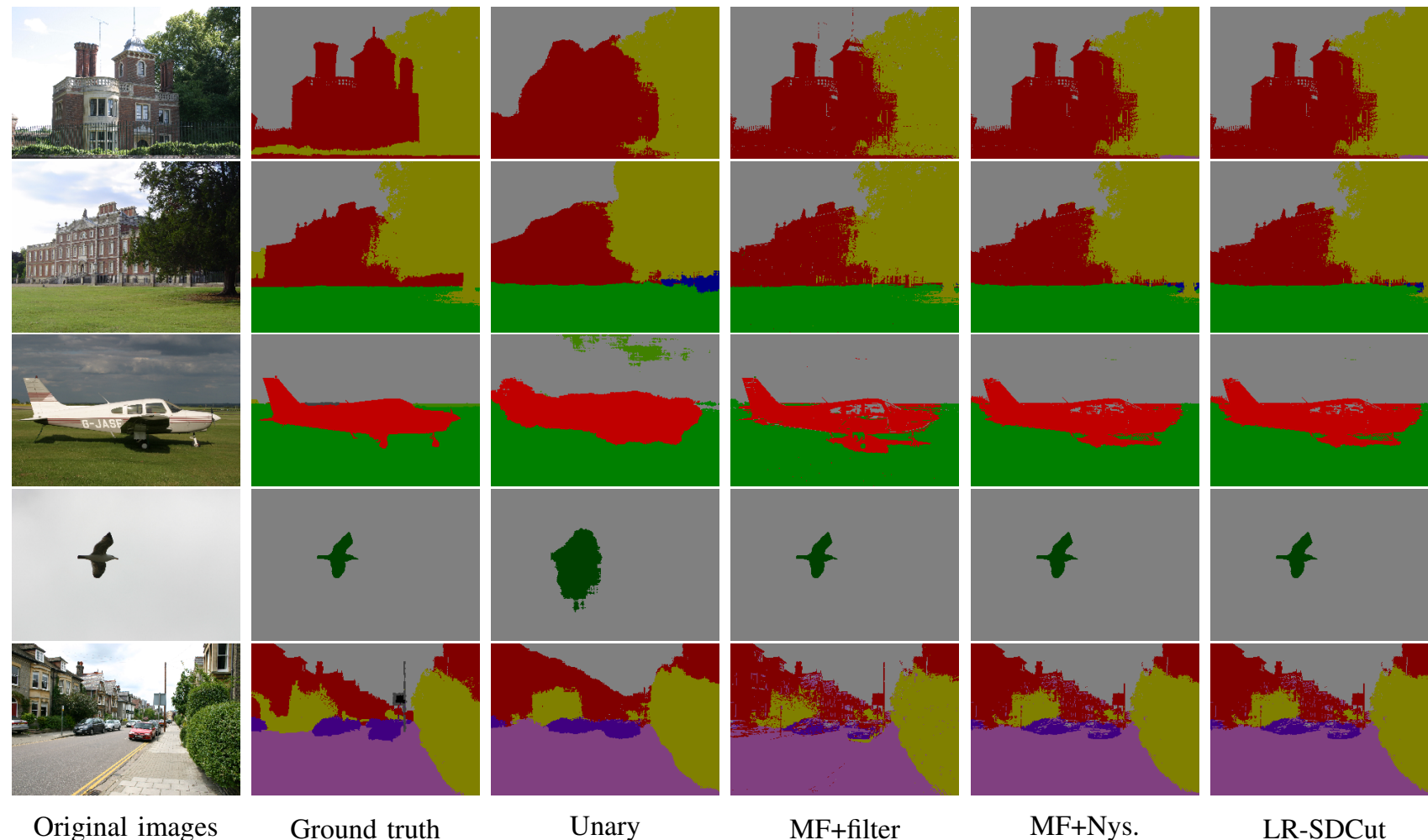
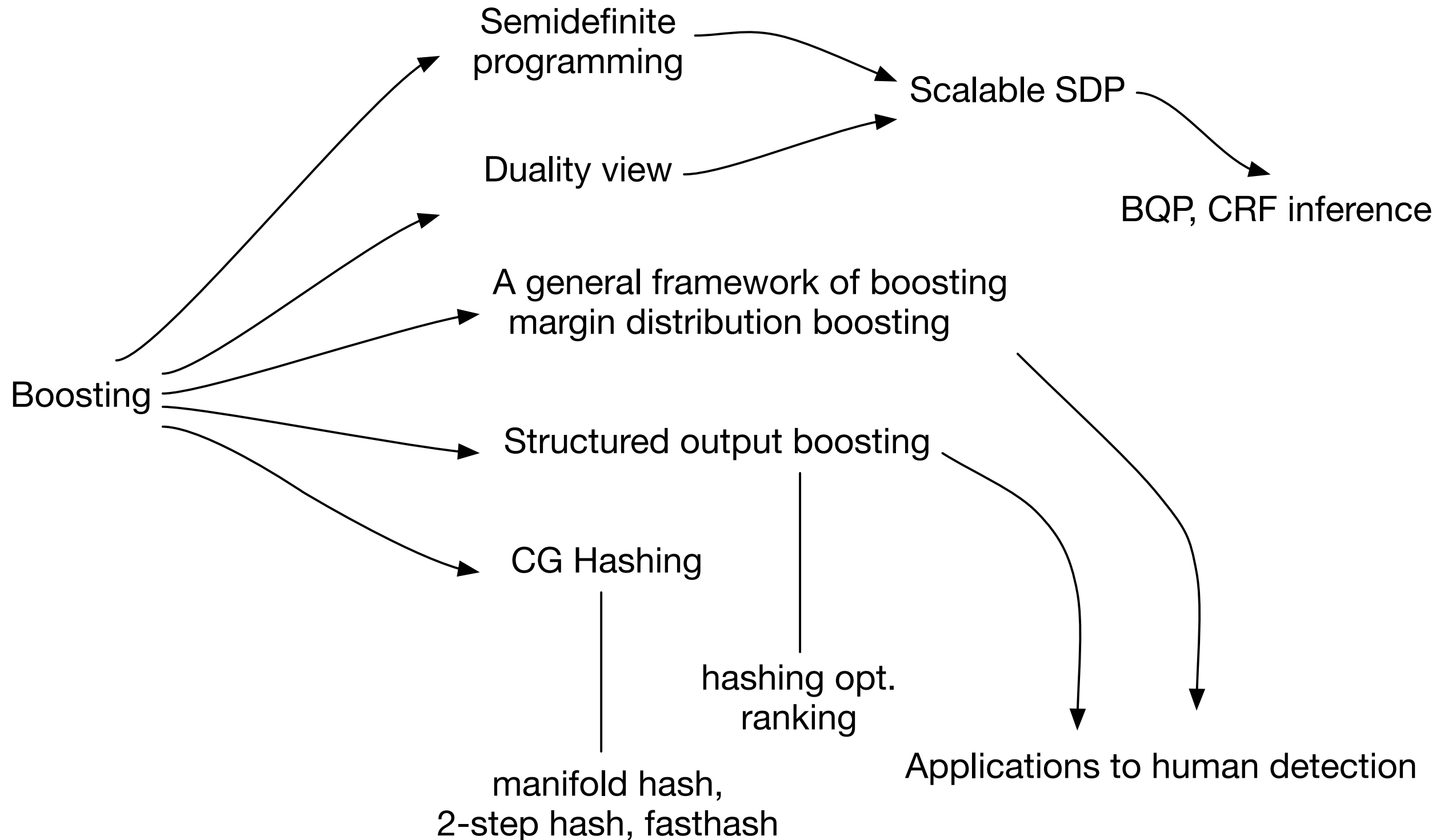


Fig. 1: Qualitative results of image segmentation. Original images and the corresponding ground truth are shown in the first two columns. The third column demonstrates the segmentation results based only on unary terms. The results of mean field methods with different matrix-vector product approaches are illustrated in the fourth and fifth columns. Our methods achieves similar visual performance with mean field methods.

- Efficient SDP inference for fully-connected CRFs based on low-rank decomposition, CVPR2015
- Efficient semidefinite branch-and-cut for MAP-MRF inference, arXiv:1404.5009
- Large-scale binary quadratic optimization using semidefinite relaxation and applications, arXiv: 1411.7564

Summary

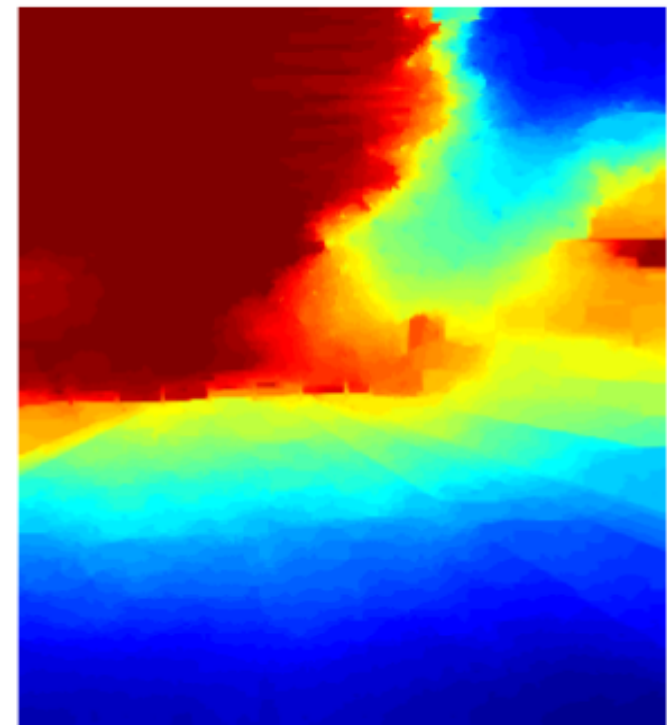
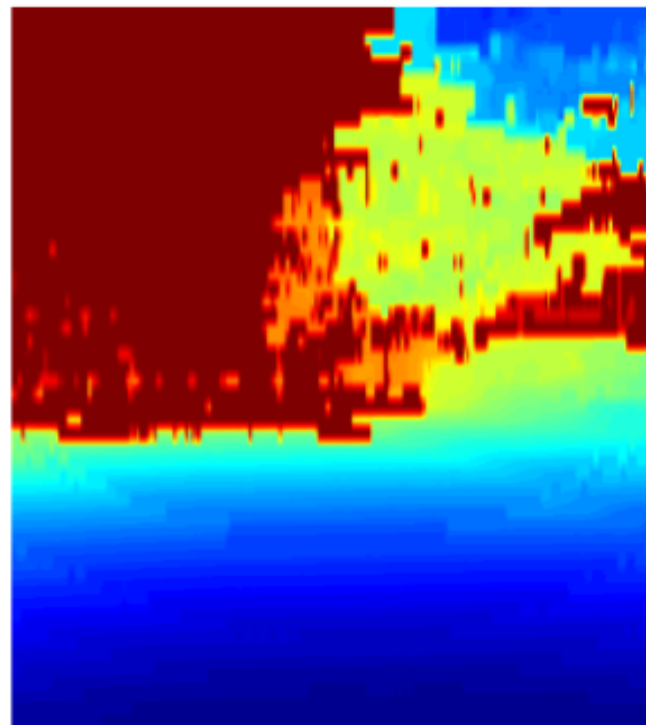
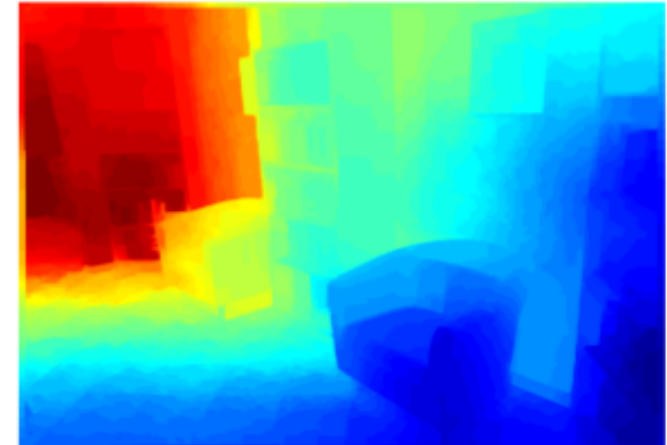
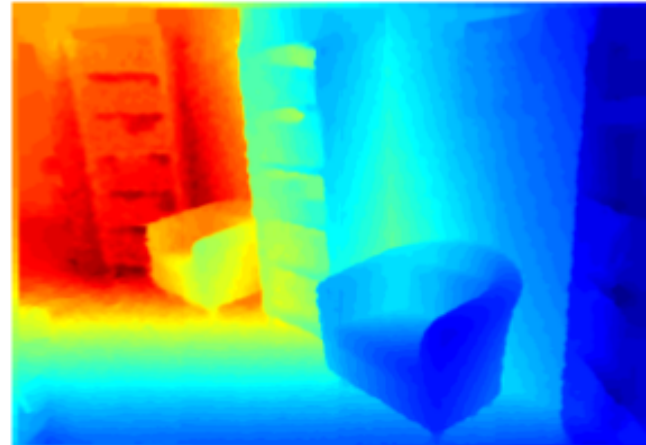
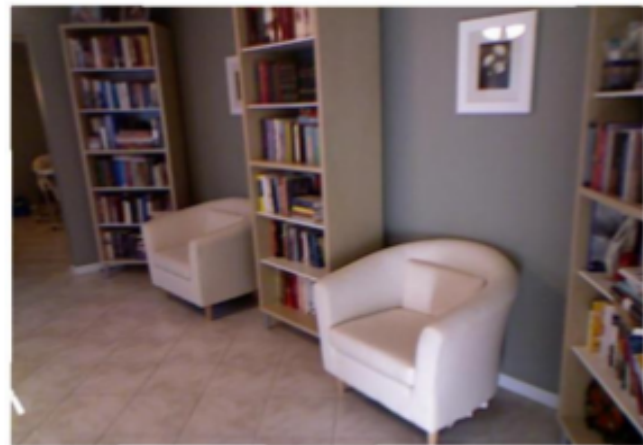


Deep learning

Deep learning

- Better encoding of CNN features
 - new Fisher vector encoding (Liu NIPS'14)
 - cross layer pooling (Liu CVPR'15)
 - mid-level feature mining (Li CVPR'15)
- Deep structured output learning:
 - deep continuous CRF, depth estimation (Liu CVPR'15)
 - piecewise training of CRF, pixel labelling (Lin arXiv'15)
 - deep message-passing machines (Lin, 2015)
- Other work
 - image captioning (Wu, 2015)
 - face recognition (Zhuang et al.)
 - text in the wild (Li et al.)
 - object detection

Depth Estimation From Single Molecular Images

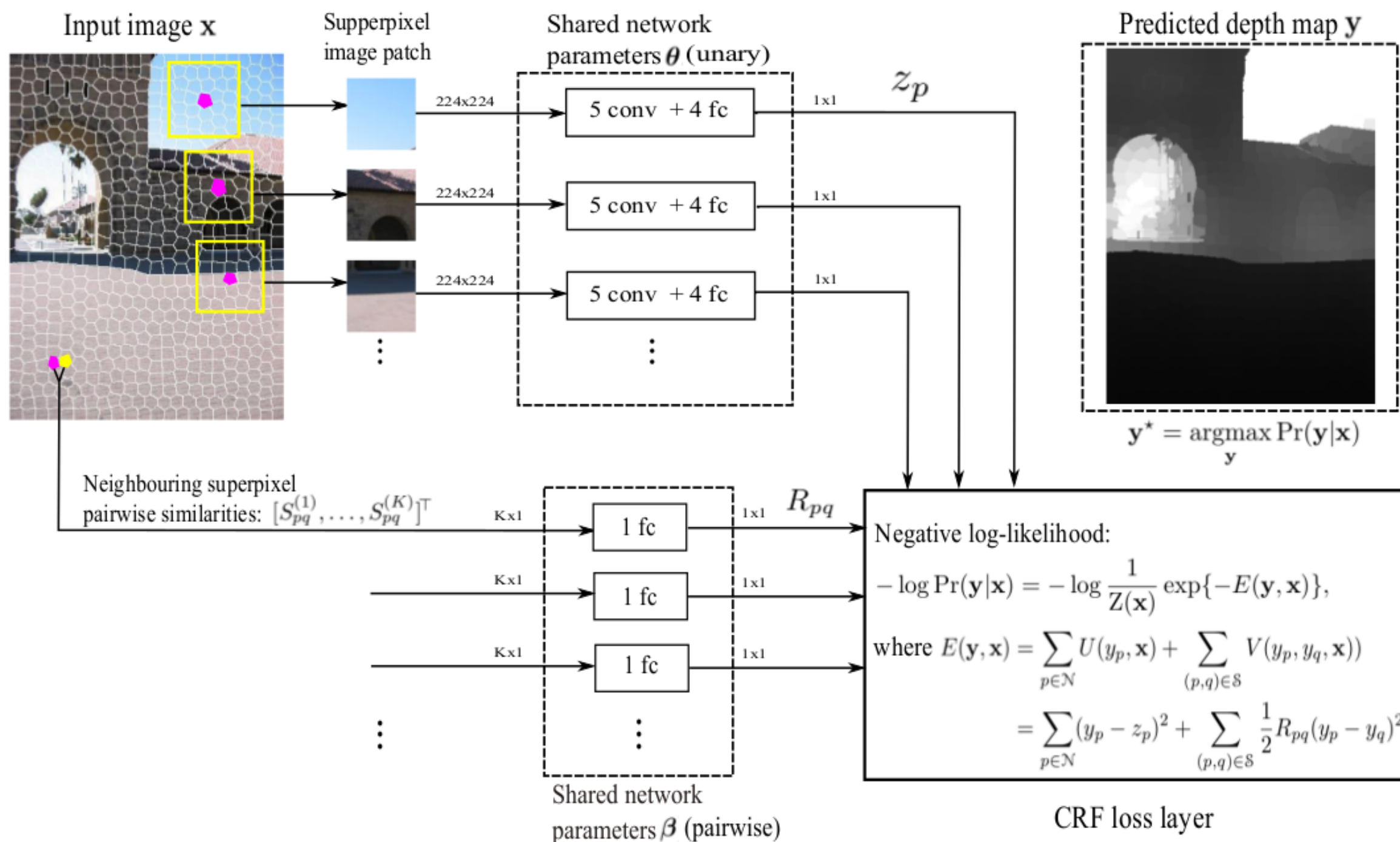


Test image

Ground-truth

Our prediction

Deep Convolutional Neural Fields

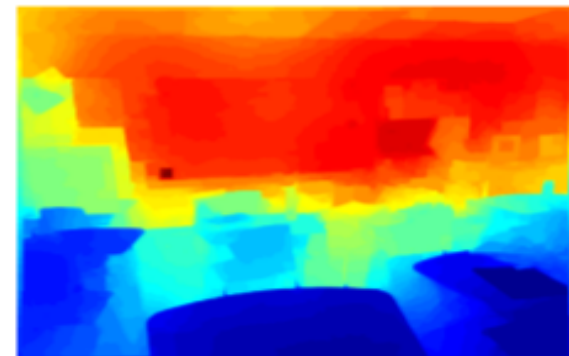
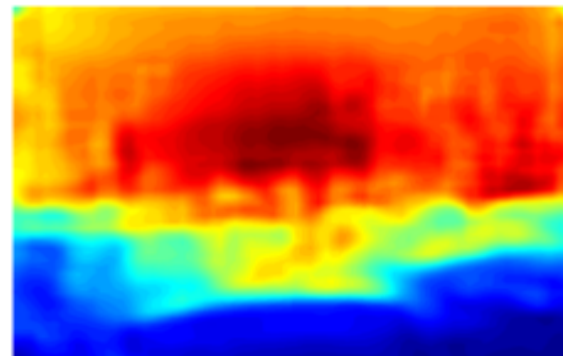
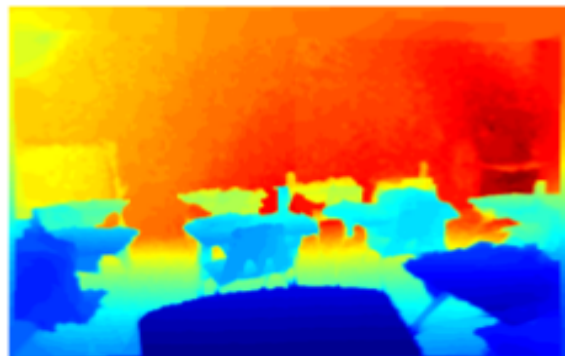
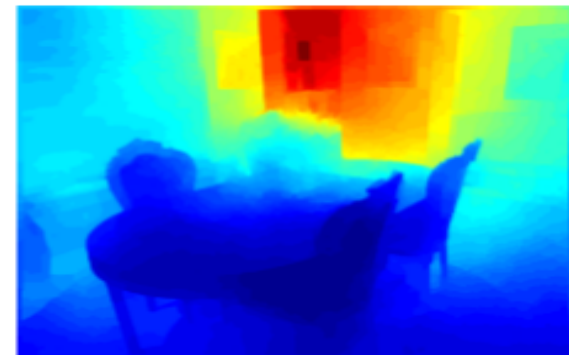
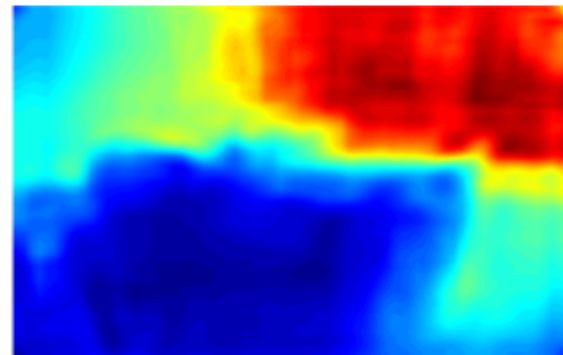
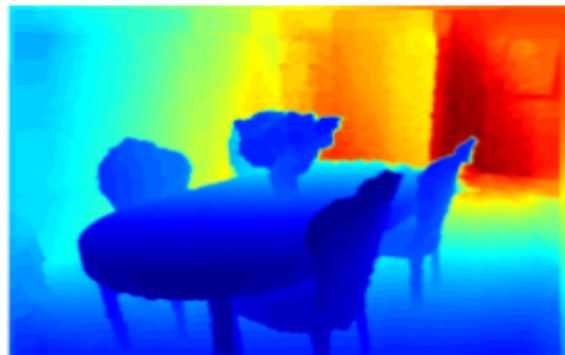
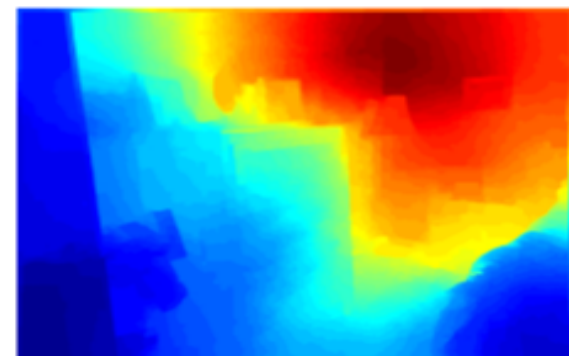
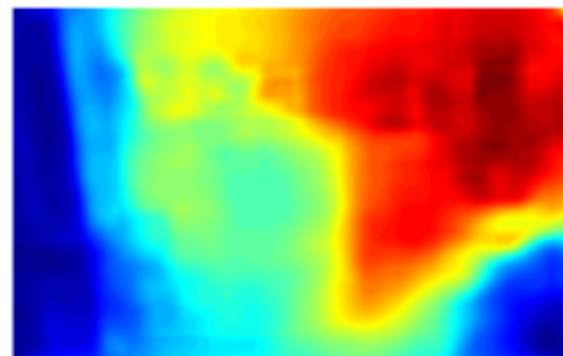
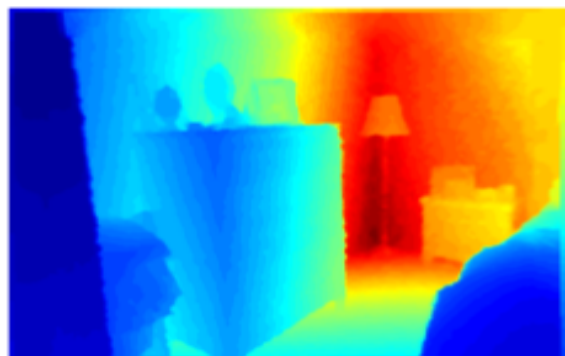
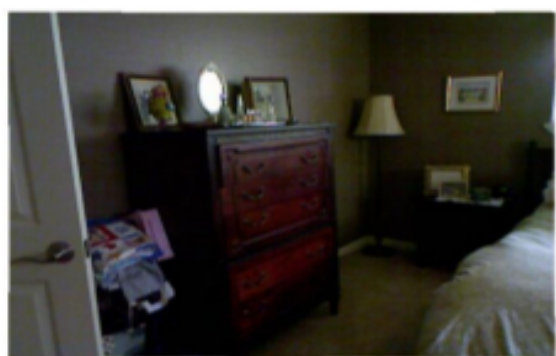
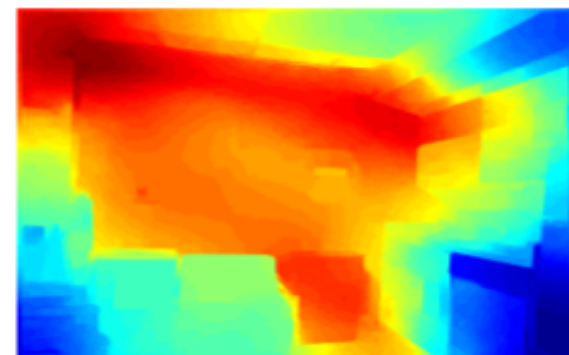
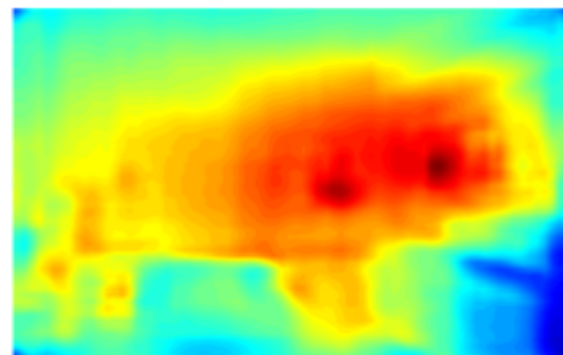
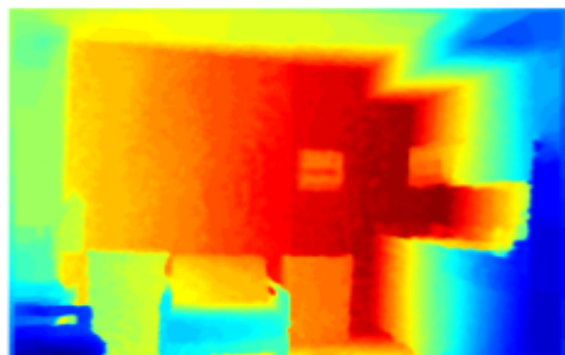


State-of-the-art comparison

Method	Error (lower is better)			Accuracy (higher is better)		
	rel	log10	rms	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Make3d [22]	0.349	-	1.214	0.447	0.745	0.897
DepthTransfer [7]	0.35	0.131	1.2	-	-	-
Discrete-continuous CRF [23]	0.335	0.127	1.06	-	-	-
Ladicky et al. [1]	-	-	-	0.542	0.829	0.941
Eigen et al. [3]	0.215	-	0.907	0.611	0.887	0.971
DCNF-FCSP (pre-train)	0.234	0.095	0.842	0.604	0.885	0.973
DCNF-FCSP (fine-tune)	0.213	0.087	0.759	0.650	0.906	0.976

Ref: CVPR'05

Prediction code and trained models: <https://bitbucket.org/fayao/dcnf-fcsp>

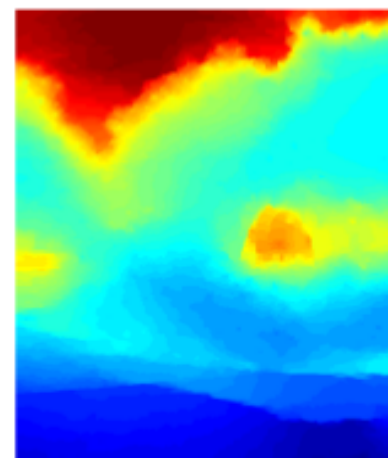
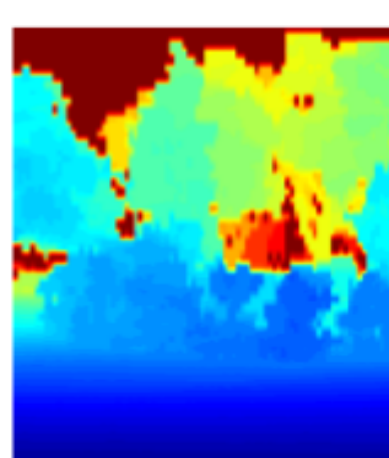
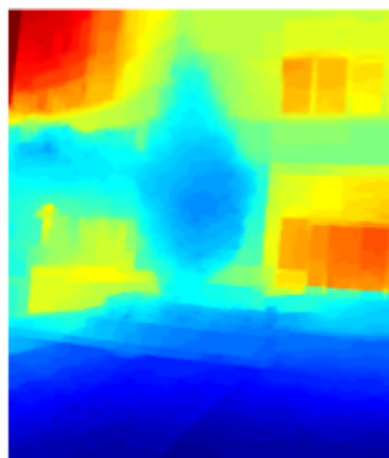
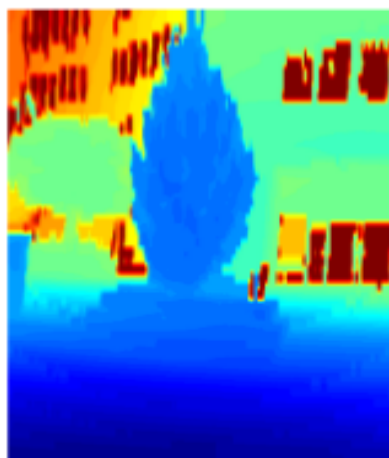
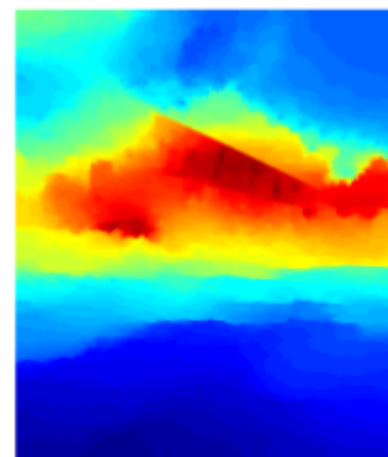
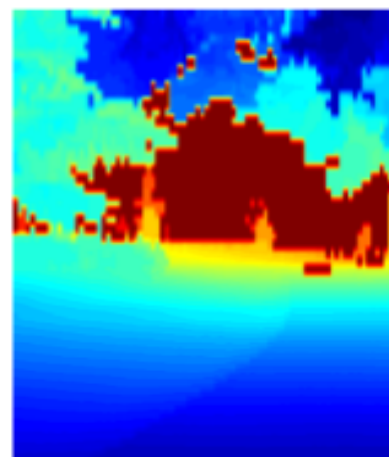
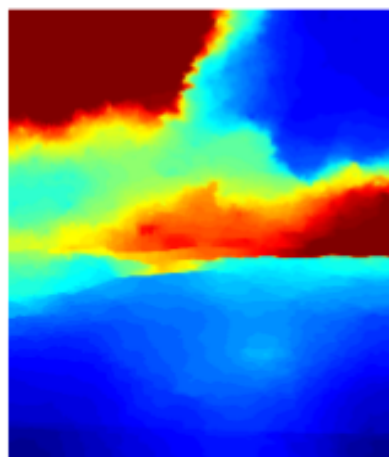
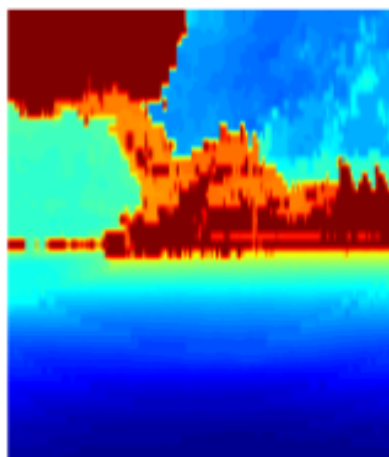
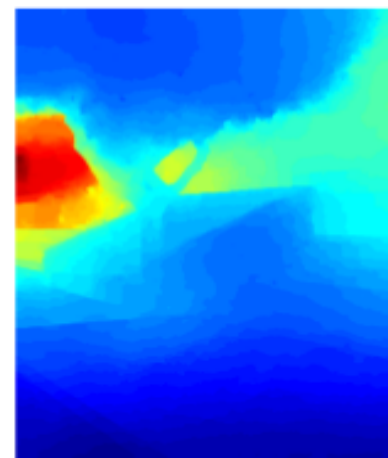
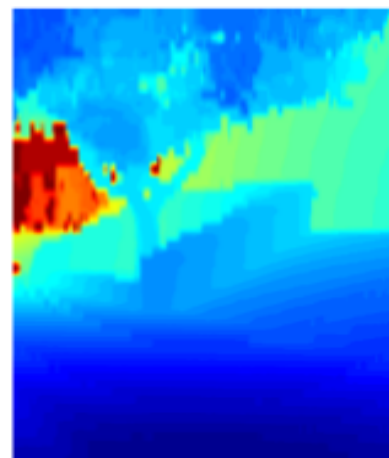
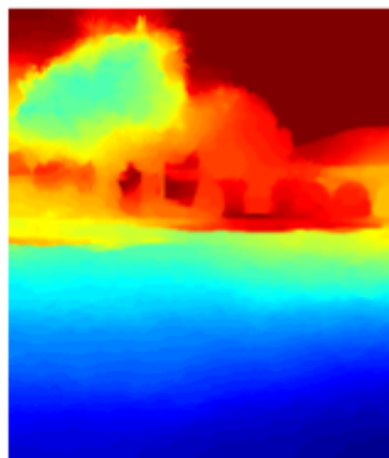
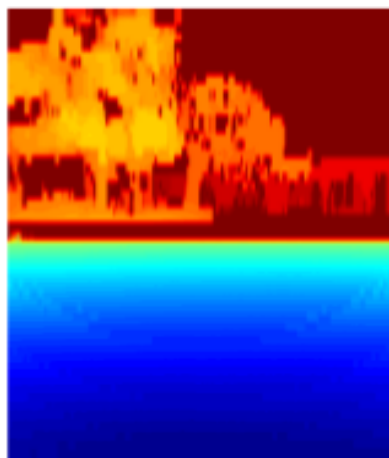


Test image

Ground-truth

Eigen etal. (NIPS2014)

Ours



Test image

Ground-truth

Our predictions

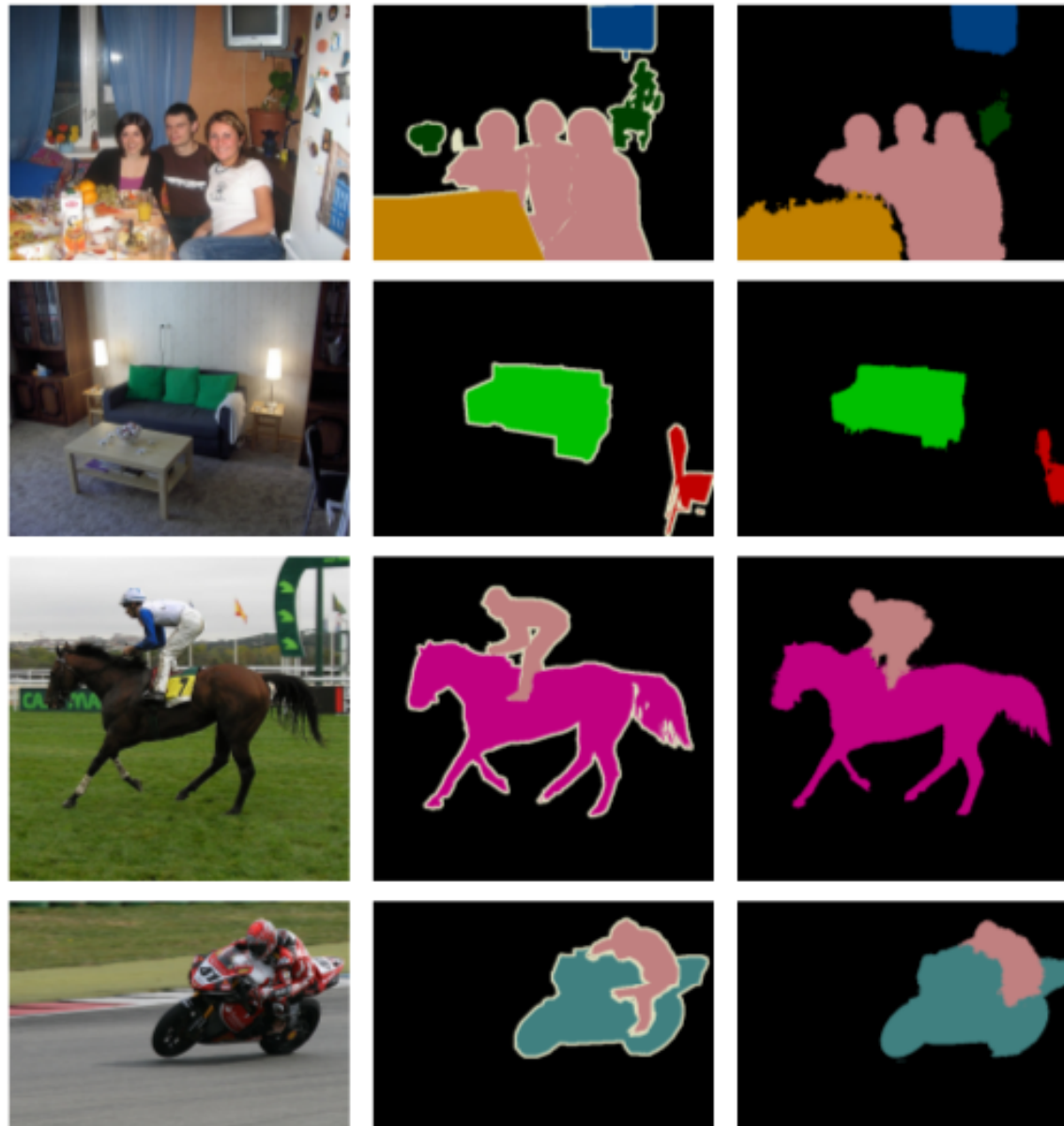
Test image

Ground-truth

Our predictions

- Deep convolutional neural fields for monocular image depth estimations
- Combining deep CNN and continuous CRF
- General learning framework

Semantic Segmentation



(a) Testing

(b) Truth

(c) Predict

Semantic Segmentation

Exploring context

Modelling various spatial relations

e.g., a car appears over a road, a glass appears over a table

Combining the strength of CRFs and CNNs

CNNs: powerful representations

CRFs: complex relation modeling.

Efficient piecewise training

Avoid repeated inference

CNN based general pairwise potential

both unary and pairwise potential: multi-scale FCNNs.

Learning multi-scale FCNNs

capture rich background context

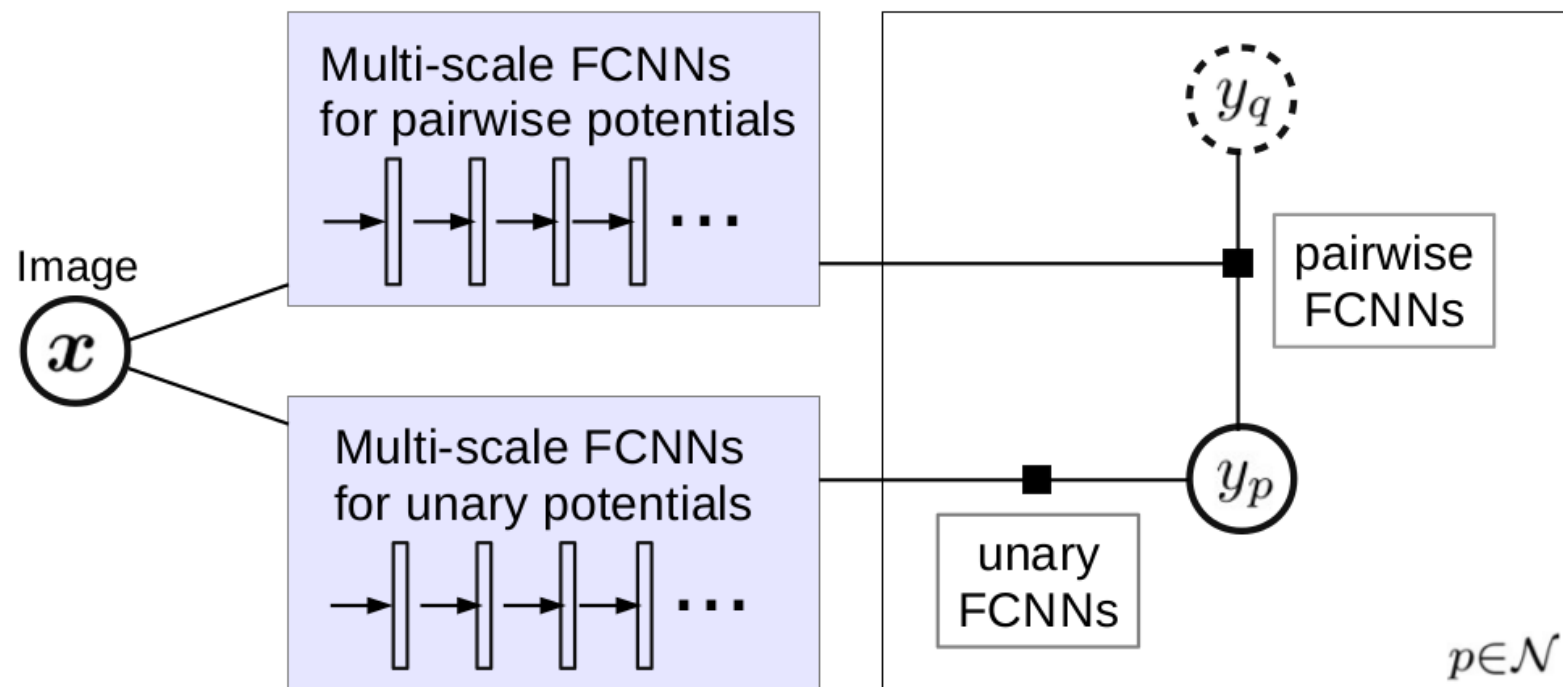


Figure 1: An illustration of our general CRF graph. Both our unary and pairwise potentials are formulated as multi-scale FCNNs which are learned in an end-to-end fashion.

Spatial relations

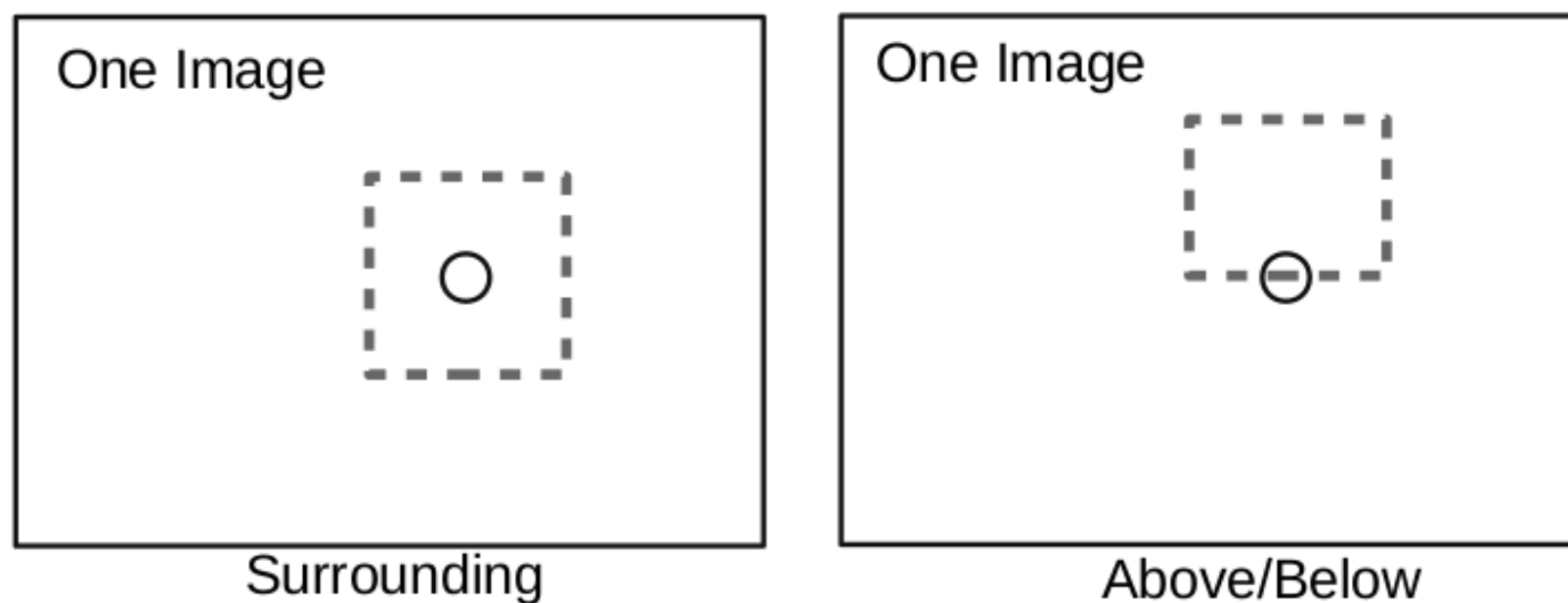
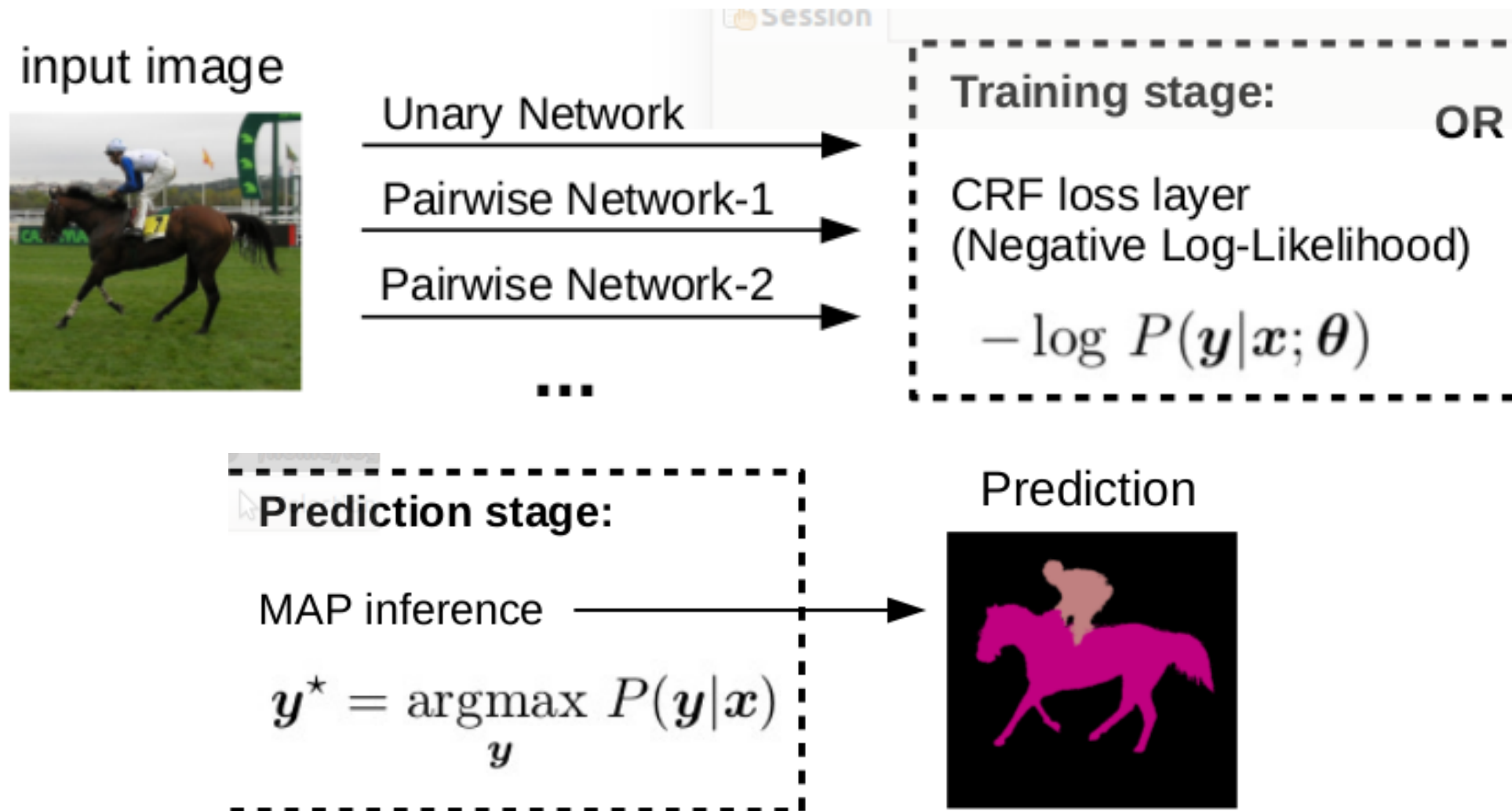


Figure 3: An illustration of two types of spatial relations, which corresponds to two types of pairwise potential functions. A node is connected to all other nodes which lie within the dash box.

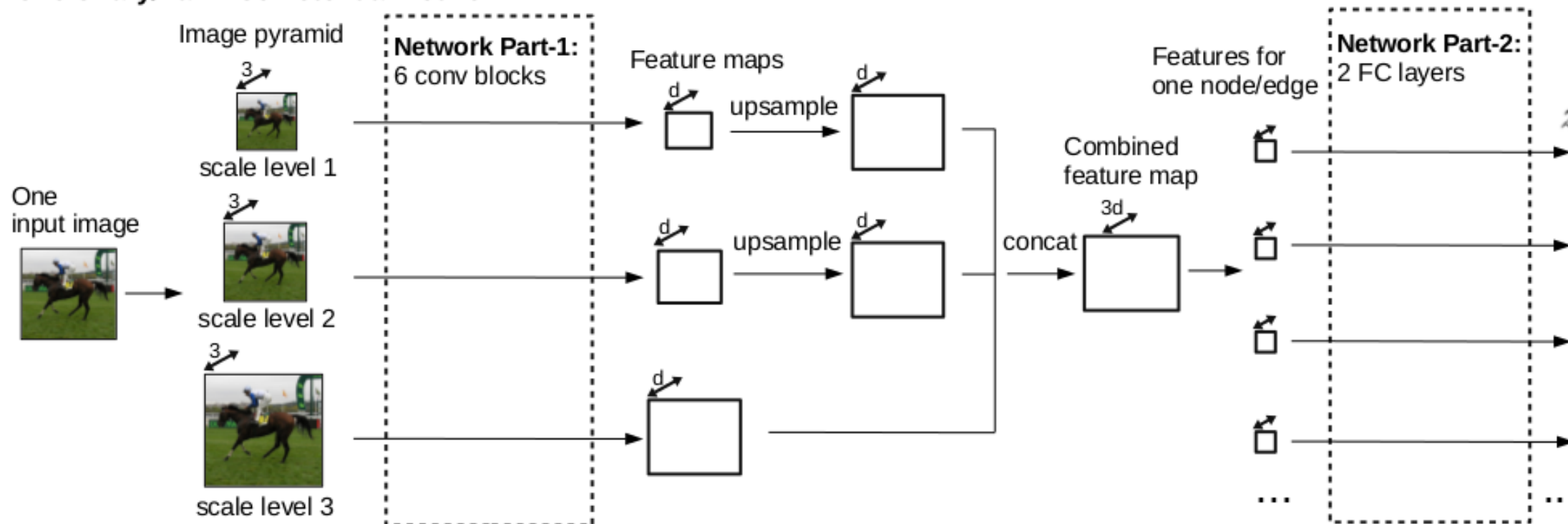
Overview



An illustration of the training and prediction process for one input image.

Details

One Unary/Pairwise Potential Network:

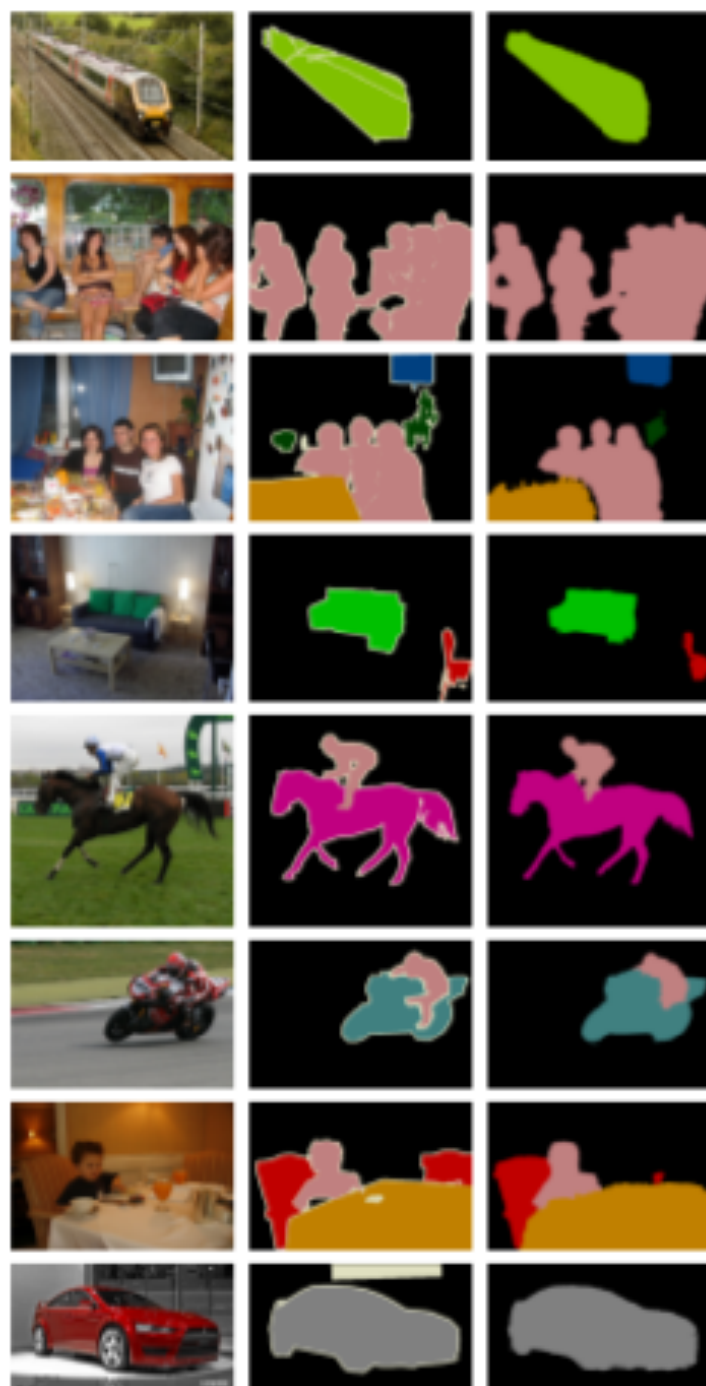


Network Part-1:

Conv block 1:	Conv block 2:	Conv block 3:	Conv block 4:	Conv block 5:	Conv block 6:
3 x 3 conv 64	3 x 3 conv 128	3 x 3 conv 256	3 x 3 conv 512	3 x 3 conv 512	7 x 7 conv 4096
3 x 3 conv 64	3 x 3 conv 128	3 x 3 conv 256	3 x 3 conv 512	3 x 3 conv 512	3 x 3 conv 512
2 x 2 pooling	2 x 2 pooling	3 x 3 conv 256	3 x 3 conv 512	3 x 3 conv 512	3 x 3 conv 512
		2 x 2 pooling	2 x 2 pooling	2 x 2 pooling	

Network Part-2:

2 fully-connected layers:
Fc 512
Fc 21(unary) or Fc 441(pairwise)



(a) Testing

(b) Truth

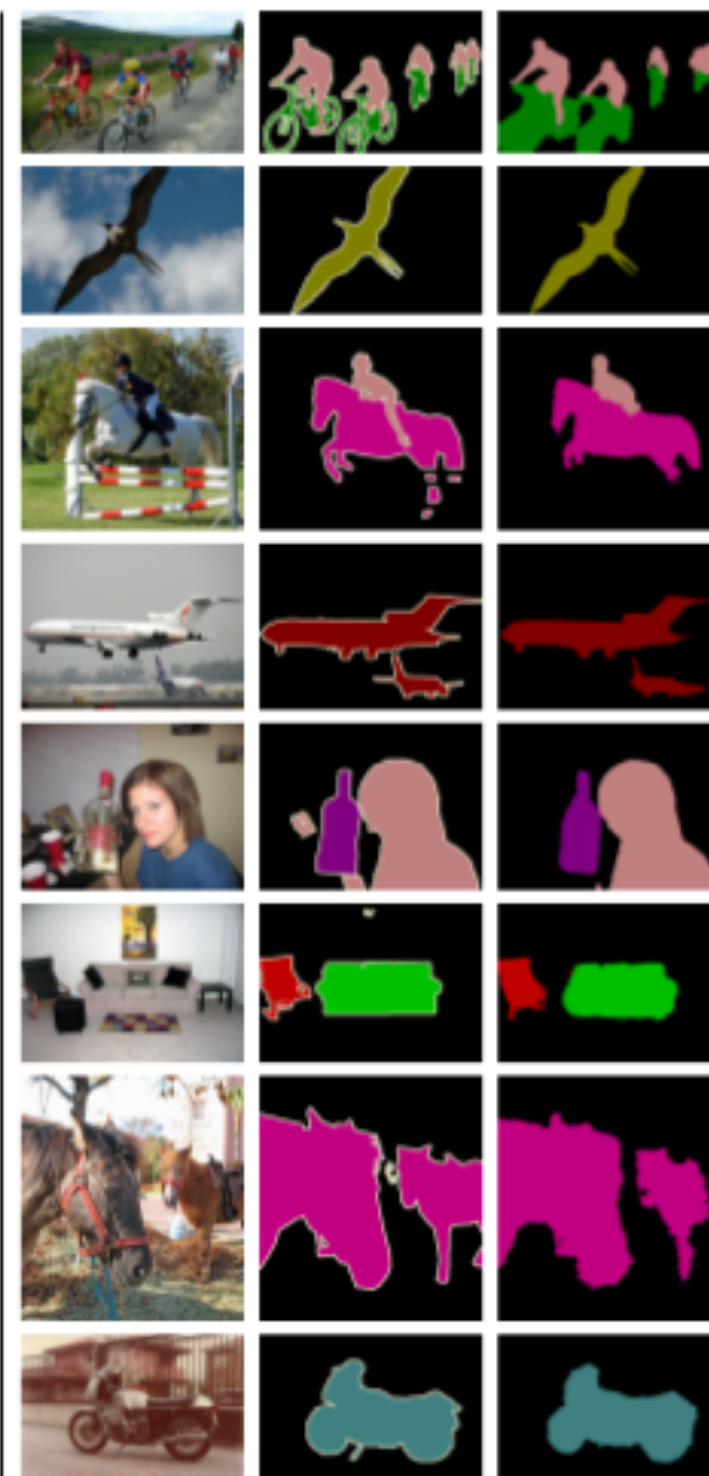
(c) Predict



(d) Testing

(e) Truth

(f) Predict



(g) Testing

(h) Truth

(i) Predict

Pascal VOC leaderboard, as of 31 May 2015

Entries equivalent to a selected submission are determined by bootstrapping the performance measure, and assessing if the differences between the selected submission and the others are not statistically significant (see sec 3.5 in [VOC 2014 paper](#)).

Average Precision (AP %)

	mean	aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor	submission date
	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	
▶ MSRA_BoxSup [?]	75.2	89.8	38.0	89.2	68.9	68.0	89.6	83.0	87.7	34.4	83.6	67.1	81.5	83.7	85.2	83.5	58.6	84.9	55.8	81.2	70.7	18-May-2015
▶ Oxford_TVG_CRF_RNN_COCO [?]	74.7	90.4	55.3	88.7	68.4	69.8	88.3	82.4	85.1	32.6	78.5	64.4	79.6	81.9	86.4	81.8	58.6	82.4	53.5	77.4	70.1	22-Apr-2015
▶ DeepLab-MSc-CRF-LargeFOV-COCO-CrossJoint [?]	73.9	89.2	46.7	88.5	63.5	68.4	87.0	81.2	86.3	32.6	80.7	62.4	81.0	81.3	84.3	82.1	56.2	84.6	58.3	76.2	67.2	26-Apr-2015
▶ Adelaide_Context_CNN_CRF_VOC [?]	72.9	89.7	37.6	77.4	62.1	72.9	88.1	84.8	81.9	34.4	80.0	55.9	79.3	82.3	84.0	82.9	59.7	82.8	54.1	77.5	70.3	25-May-2015
▶ DeepLab-CRF-COCO-LargeFOV [?]	72.7	89.1	38.3	88.1	63.3	69.7	87.1	83.1	85.0	29.3	76.5	56.5	79.8	77.9	85.8	82.4	57.4	84.3	54.9	80.5	64.1	18-Mar-2015
▶ POSTECH_EDeconvNet_CRF_VOC [?]	72.5	89.9	39.3	79.7	63.9	68.2	87.4	81.2	86.1	28.5	77.0	62.0	79.0	80.3	83.6	80.2	58.8	83.4	54.3	80.7	65.0	22-Apr-2015
▶ Oxford_TVG_CRF_RNN_VOC [?]	72.0	87.5	39.0	79.7	64.2	68.3	87.6	80.8	84.4	30.4	78.2	60.4	80.5	77.8	83.1	80.6	59.5	82.8	47.8	78.3	67.1	22-Apr-2015
▶ DeepLab-MSc-CRF-LargeFOV [?]	71.6	84.4	54.5	81.5	63.6	65.9	85.1	79.1	83.4	30.7	74.1	59.8	79.0	76.1	83.2	80.8	59.7	82.2	50.4	73.1	63.7	02-Apr-2015
▶ MSRA_BoxSup [?]	71.0	86.4	35.5	79.7	65.2	65.2	84.3	78.5	83.7	30.5	76.2	62.6	79.3	76.1	82.1	81.3	57.0	78.2	55.0	72.5	68.1	10-Feb-2015
▶ DeepLab-CRF-COCO-Strong [?]	70.4	85.3	36.2	84.8	61.2	67.5	84.6	81.4	81.0	30.8	73.8	53.8	77.5	76.5	82.3	81.6	56.3	78.9	52.3	76.6	63.3	11-Feb-2015
▶ DeepLab-CRF-LargeFOV [?]	70.3	83.5	36.6	82.5	62.3	66.5	85.4	78.5	83.7	30.4	72.9	60.4	78.5	75.5	82.1	79.7	58.2	82.0	48.8	73.7	63.3	28-Mar-2015
▶ DeepLab-CRF-MSc [?]	67.1	80.4	36.8	77.4	55.2	66.4	81.5	77.5	78.9	27.1	68.2	52.7	74.3	69.6	79.4	79.0	56.9	78.8	45.2	72.7	59.3	30-Dec-2014
▶ DeepLab-CRF [?]	66.4	78.4	33.1	78.2	55.6	65.3	81.3	75.5	78.6	25.3	69.2	52.7	75.2	69.0	79.1	77.6	54.7	78.3	45.1	73.3	56.2	23-Dec-2014
▶ CRF_RNN [?]	65.2	80.9	34.0	72.9	52.6	62.5	79.8	76.3	79.9	23.6	67.7	51.8	74.8	69.9	76.9	76.9	49.0	74.7	42.7	72.1	59.6	10-Feb-2015

Efficient piecewise training of deep structured models for semantic segmentation
<http://arxiv.org/abs/1504.01013>

Image captioning using LSTMs

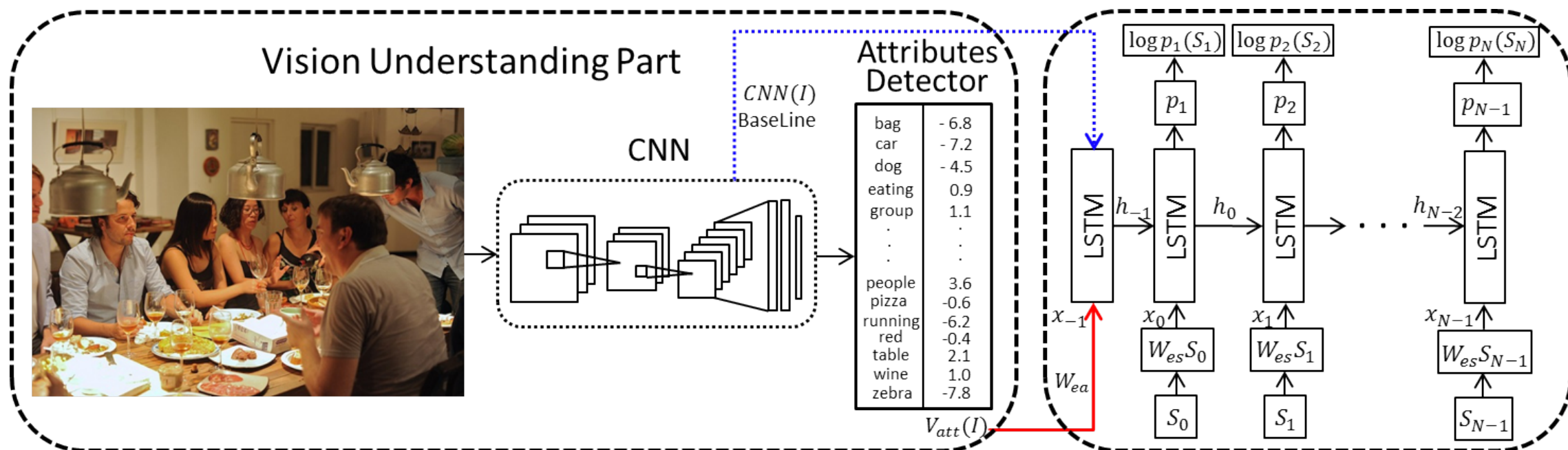


Figure 1: Our two-stage image captioning framework. The first stage is the vision understanding part, which learns a mapping between an image and semantic attributes through CNN. The second stage is the language generation part, which learns a mapping from input attributes vector (red arrow) to a sequence of words through LSTM. In the end-to-end baseline mode, CNN features are input to the LSTM directly (blue dash arrow), without the attributes detector.

Image captioning using LSTMs

State-of-art-Flickr30k	BLEU-1	BLEU-2	BLEU-3	BLEU-4	\mathcal{PPL}
Karpathy & Li (NeuralTalk)[18]	0.57	0.37	0.24	0.16	-
Chen & Zintick (Mind's Eye) [5]	-	-	-	0.13	19.10
Google(NIC)[38]	0.66	-	-	-	-
Donahue et al. (LRCN)[9]	0.59	0.39	0.25	0.16	-
Mao et al. (m-Rnn-AlexNet)[29]	0.54	0.36	0.23	0.15	35.11
Mao et al. (m-Rnn-VggNet)[29]	0.60	0.41	0.28	0.19	20.72
Xu et al. (Hard-Attention)[40]	0.67	0.44	0.30	0.20	-
BaseLine-Flickr30k					
VggNet+LSTM	0.57	0.38	0.25	0.17	18.83
VggNet-PCA+LSTM	0.59	0.40	0.26	0.17	18.92
GoogLeNet+LSTM	0.58	0.39	0.26	0.17	18.77
Ours-Flickr30k					
gt-attributes-Sampling [†]	0.73	0.53	0.38	0.27	15.36
gt-attributes-BeamSearch [†]	0.78	0.57	0.42	0.30	14.88
predict-attributes-Sampling	0.63	0.43	0.28	0.19	17.57
predict-attributes-BeamSearch	0.67	0.46	0.31	0.20	17.01

Table 2: BLEU-1,2,3,4 and \mathcal{PPL} metrics compared to other state-of-the-art methods and our base-line on Flickr30k dataset. [†] indicates ground truth attributes labels are used. Our \mathcal{PPL} s are based on Flickr30k word dictionaries of size 7414.

misc

Face recognition

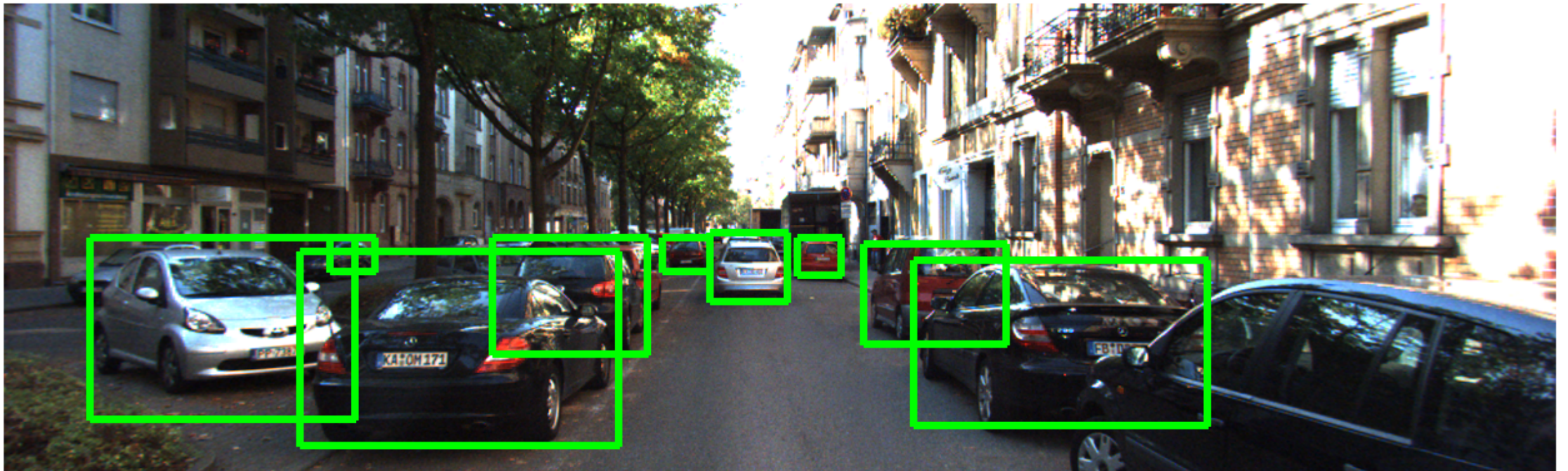
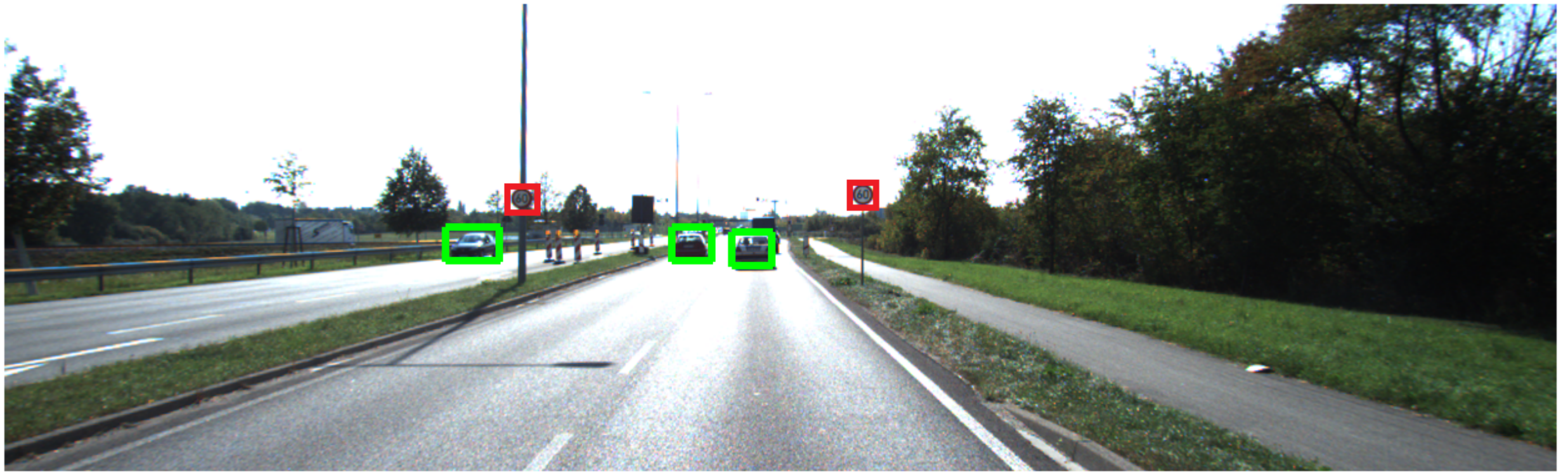
LFW: ~98%. Trained with 0.5M labelled faces of 10k classes

Classification

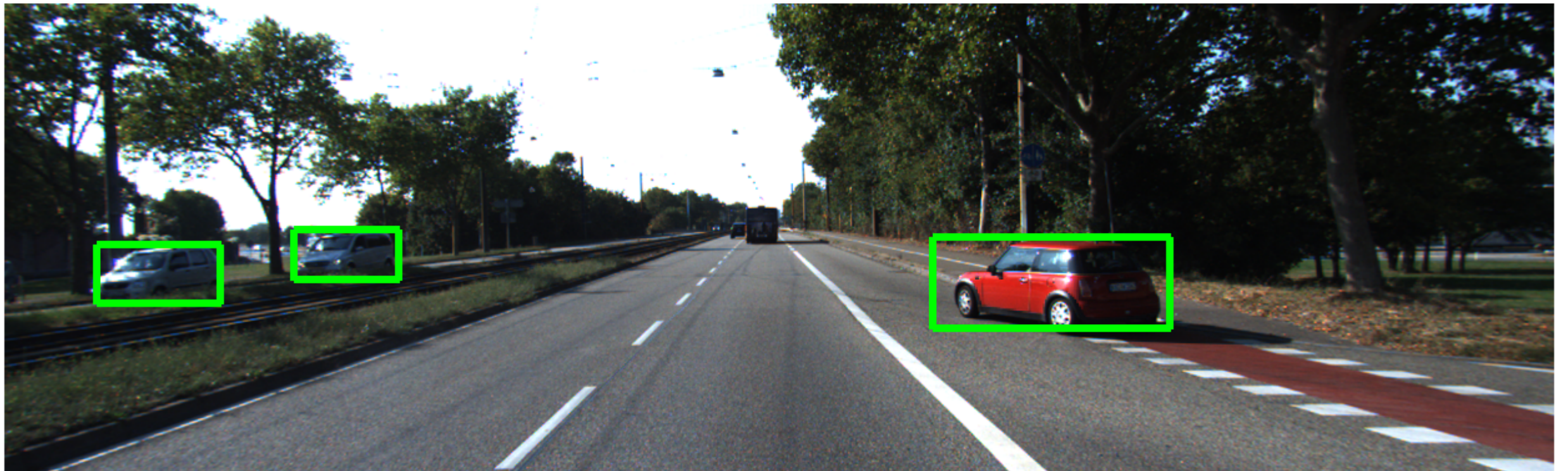
Training GoogleNet took ~3 days with 8 K40's (data parallelisation)
Software developed on top of a fork of Caffe. We only implemented data parallelisation.

Working on various object detection problems now

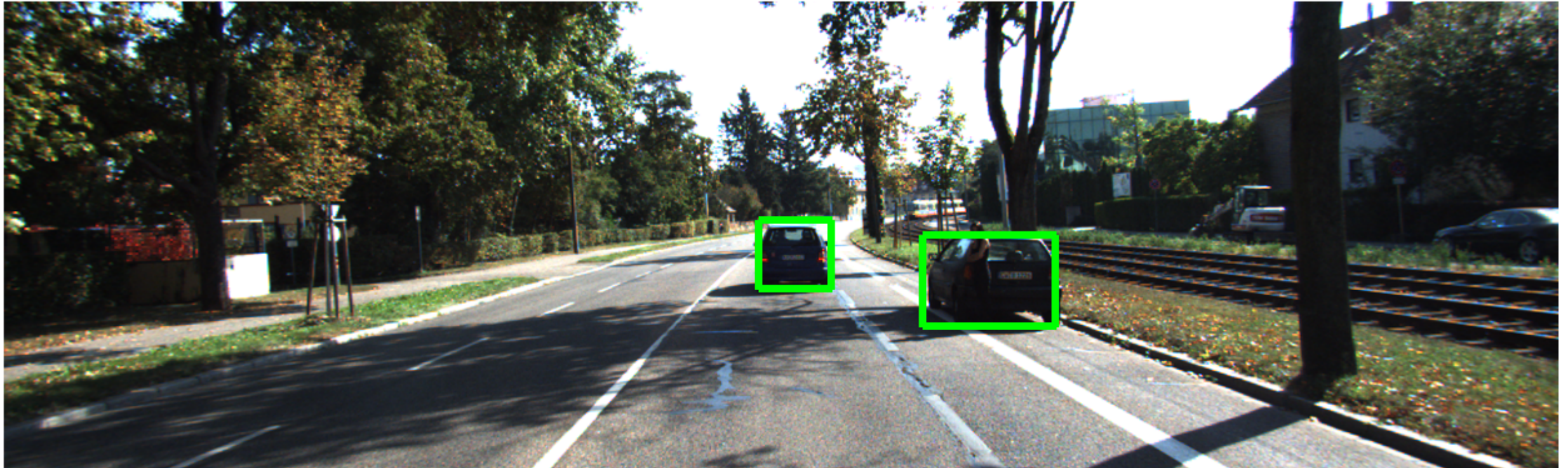
Car and sign detection



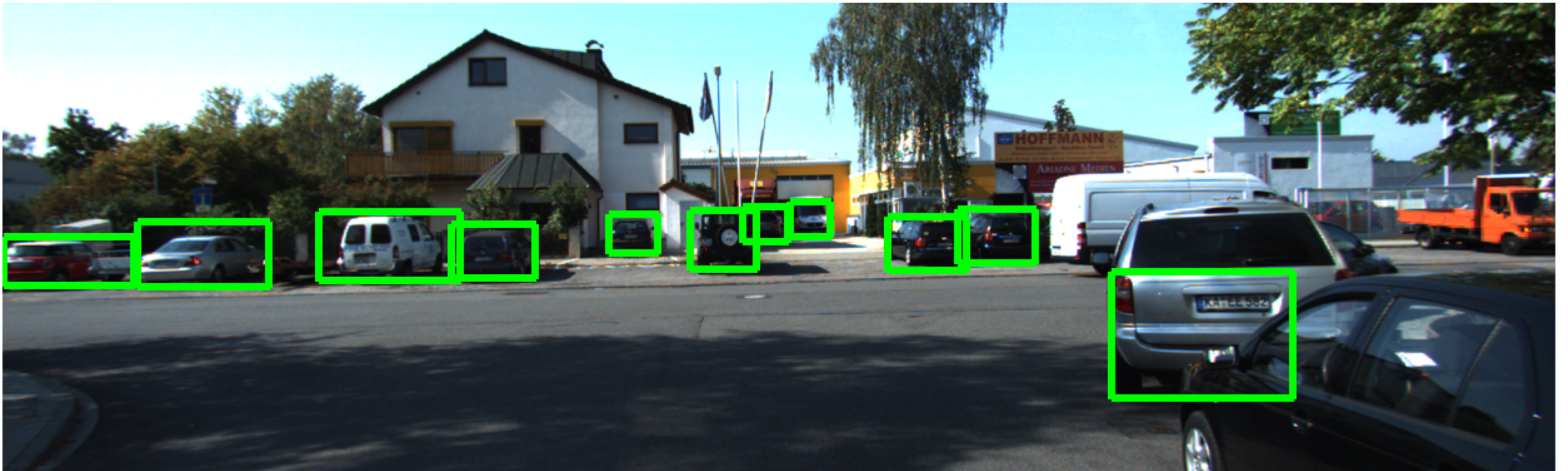
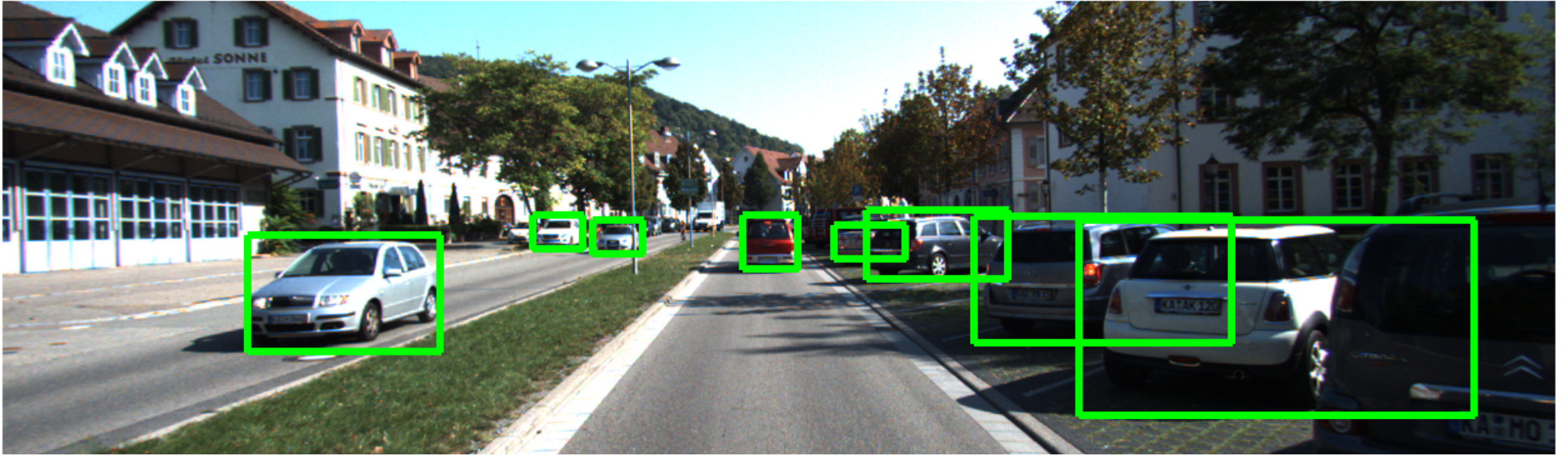
Car and sign detection



Car and sign detection



Car and sign detection



Text in the wild

- *ICDAR2003 Text in wild detection results*

Method	Precision	Recall	F-measure
<i>Our Method</i>	<i>0.84</i>	<i>0.70</i>	<i>0.76</i>
<i>Max et al. (ECCV2014)</i>	<i>0.89</i>	<i>0.66</i>	<i>0.75</i>
<i>Huang et al. (ECCV2014)</i>	<i>0.84</i>	<i>0.67</i>	<i>0.75</i>
<i>Neumann et al. (ICDAR2011)</i>	<i>0.65</i>	<i>0.64</i>	<i>0.63</i>

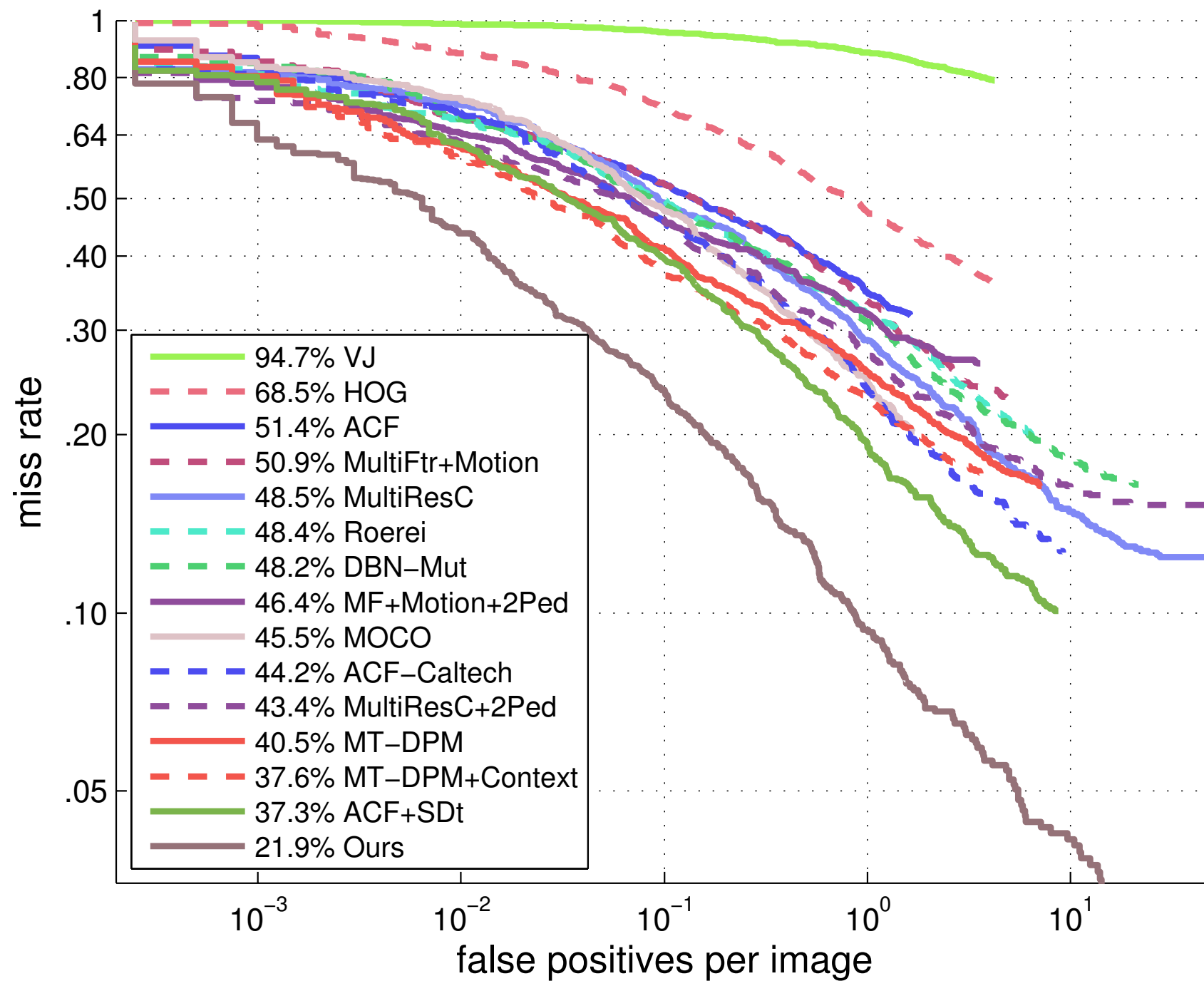


THE UNIVERSITY
of ADELAIDE

Car license plate detection



Pedestrian detection (boosting w. trees. not DL)



<http://arxiv.org/pdf/1409.5209.pdf>

<https://github.com/chhshen/pedestrian-detection>

What we are planning to do in the near future:

- Efficient deep structured output learning (highly nonsubmodular models?)
- Dense prediction
e.g., per-pixel depth estimation and label prediction
- Multi-modal learning (cross domain transfer?)
e.g., how can we take advantage of the knowledge embedded in Wikipedia for vision problems?

That's all. Questions?



cs.adelaide.edu.au/~chhshen