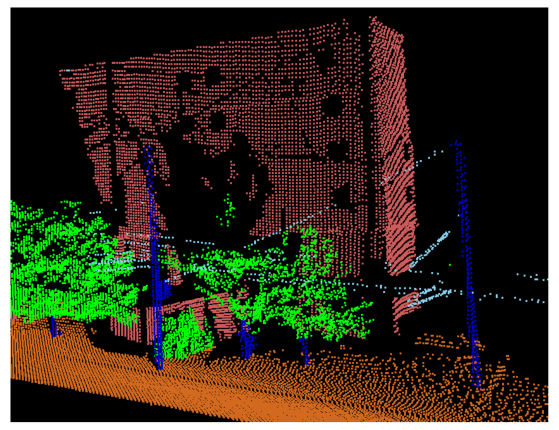
Learning Message-Passing Inference Machines for Structured Prediction

The Robotics Institute, Carnegie Mellon University
Stephane Ross, Daniel Munoz, Martial Hebert & J. Andrew Bagnell

Overview: Structured Prediction for Scene Labeling

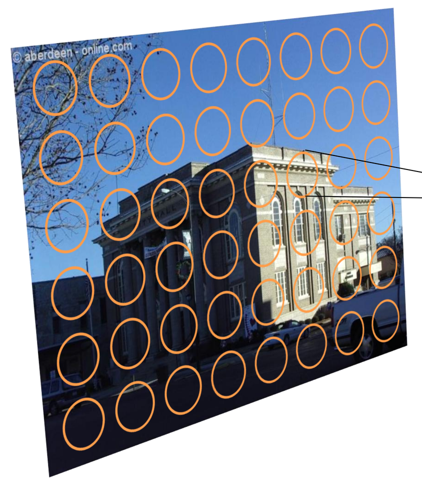
Predict a class label at every site (pixel/superpixel/segment) in a scene:

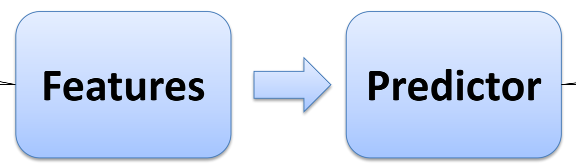


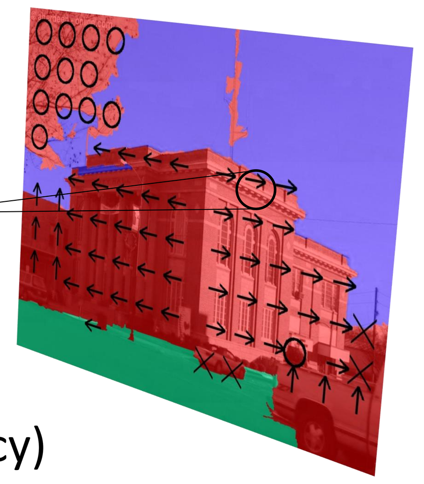
3D Point Cloud Classification [2]

Surface Layout Estimation [6]

Naive Approach: Independent Predictions



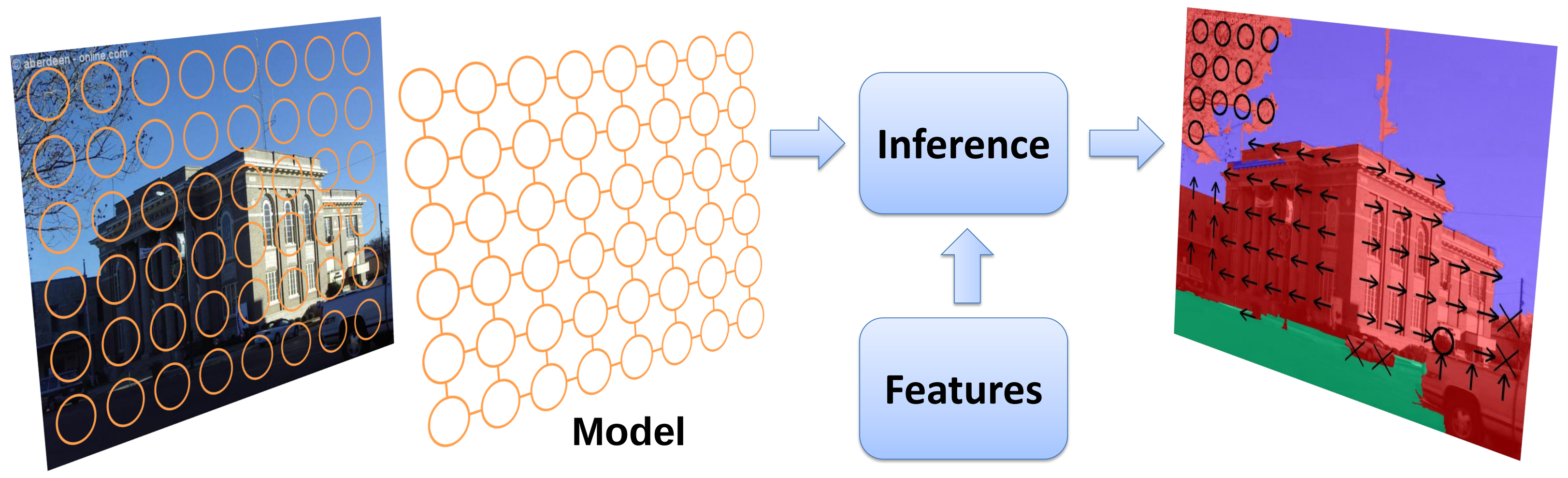




✔ Simple, fast and tractable

✘ No context (usually poor accuracy)

Typical Approach: Graphical Models



✔ Exploits context (models spatial relationships between neighbors)

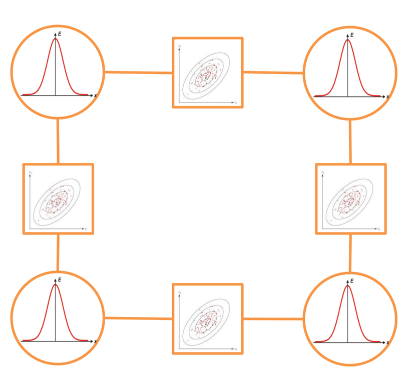
✘ Intractable without approximations (e.g. Tree model/Loopy BP/Graph-cut with limited class of potentials)

✘ Optimization with approximations can lead to poor performance [5]

<u>Pairwise CRF Model</u>

- Models the conditional joint distribution over labelings Y of the scene given input features X:

$$\mathbf{P}(\mathbf{Y}\,|\,\mathbf{X}) \propto \prod_{i \in V} \phi(\mathbf{Y}_i, \mathbf{X}_i) \prod_{(i,j) \in E} \varphi(\mathbf{Y}_i, \mathbf{Y}_j, \mathbf{X}_{ij})$$



- Learn node ϕ and edge potentials ψ from training data.

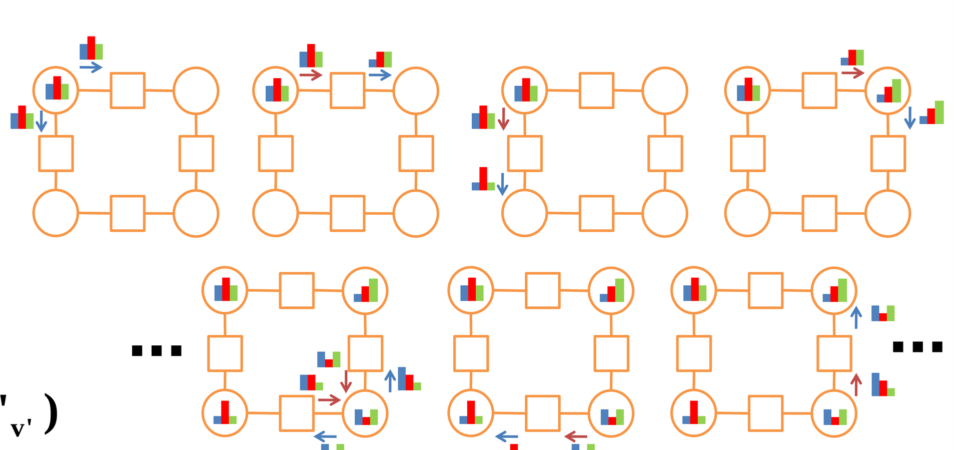
<u>Loopy Belief Propagation (BP)</u>

- Approximate inference procedure to find most likely labeling.
- Iteratively visits all nodes & edges in the graph to update neighbors' messages until convergence:

$$\mathbf{P}(\mathbf{Y}_v = \mathbf{y}_v) \propto \phi(\mathbf{y}_v, \mathbf{x}_v) \prod_{f' \in N_v} \mathbf{m}_{f'v}(\mathbf{y}_v)$$

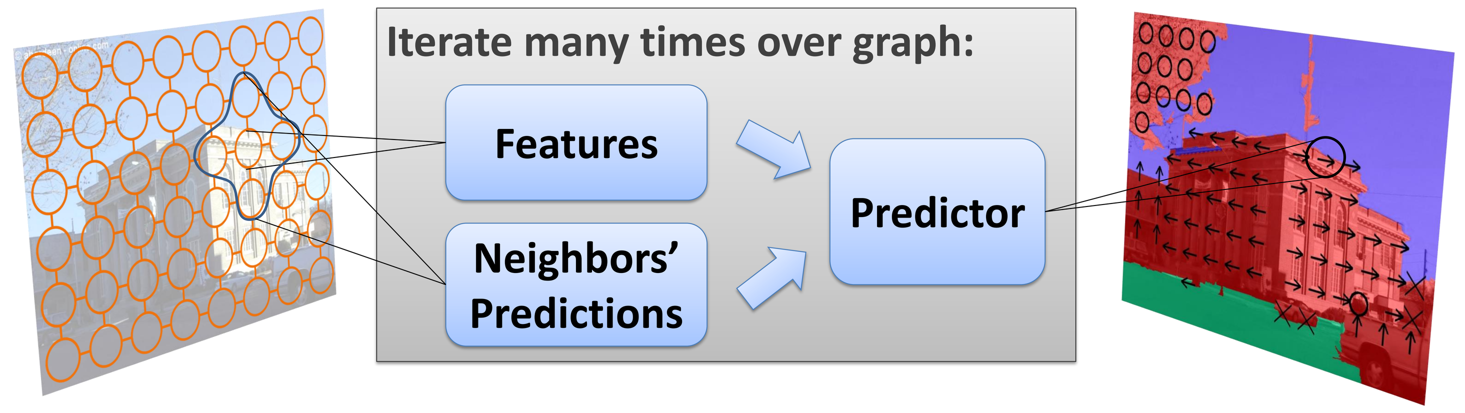
$$\mathbf{m}_{vf}(\mathbf{y}_v) \propto \phi(\mathbf{y}_v, \mathbf{x}_v) \prod_{f' \in N_v \setminus f} \mathbf{m}_{f'v}(\mathbf{y}_v)$$

$$\mathbf{m}_{fv}(\mathbf{y}_v) \propto \sum_{y'_f | y'_v = y_v} \varphi(\mathbf{y}'_f, \mathbf{x}_f) \prod_{v' \in N_f \setminus v} \mathbf{m}_{v'f}(\mathbf{y}'_{v'})$$



Proposed Approach: Inference Machines

Inference procedure treated as a black box function to optimize, instead of learning a graphical model ([3,4] are special cases of our approach):



✔ Exploits context

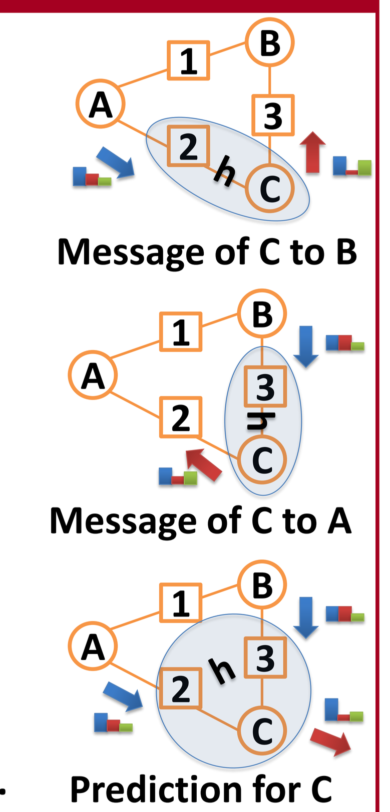
✔ Exploits discriminative power of classifiers

✔ Strong theoretical guarantees

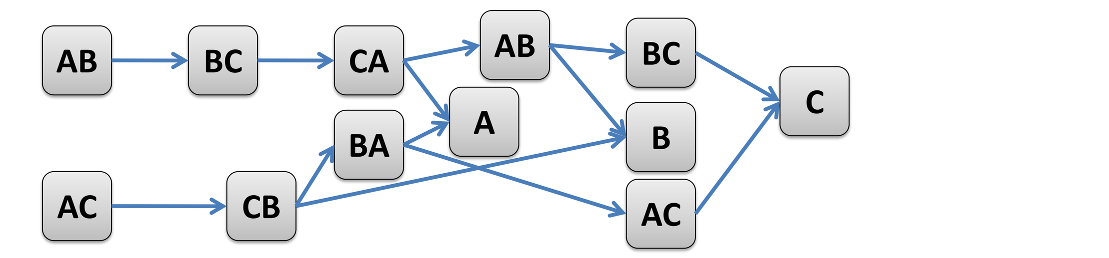
✔ Accuracy/speed trade-off via more/less complex predictors

Given any message-passing algorithm (e.g. Mean-Field, BP):

- Keep the same algorithm structure.
- Message updates are replaced by a learned predictor h's outputs; e.g. logistic regressor.
- h outputs messages (a distribution over labels) to send to neighbors, given input features and neighbors' messages.
- h minimizes the loss of the inference's output prediction.



Issue: Predictor's outputs change its future inputs (non-iid)

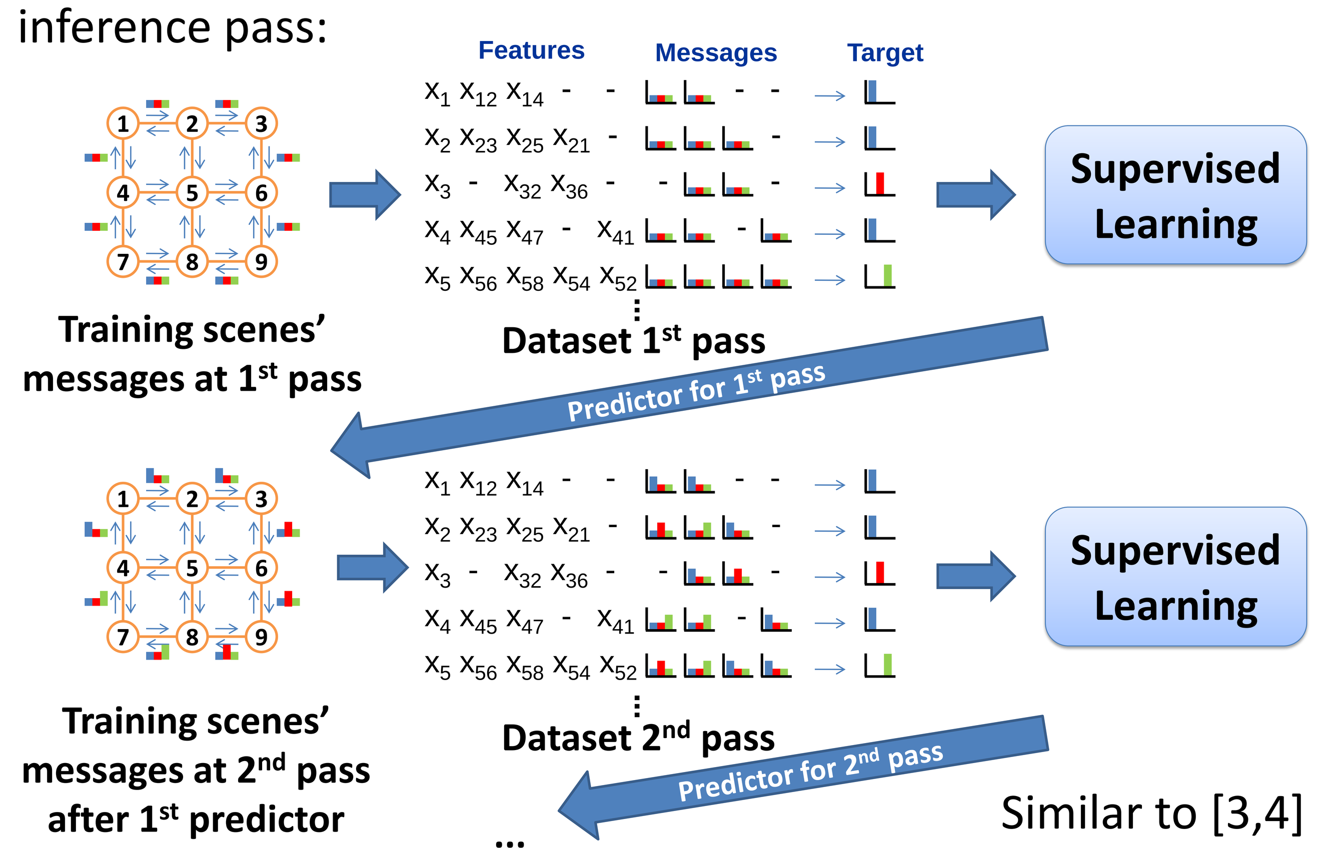


Interdependencies between outputs for 3 async. passes from graph on the right.

General Learning Procedures for Inference Machines

Learning Synchronous Message-Passing Inference

Forward Training: Learn a separate predictor for each inference pass:

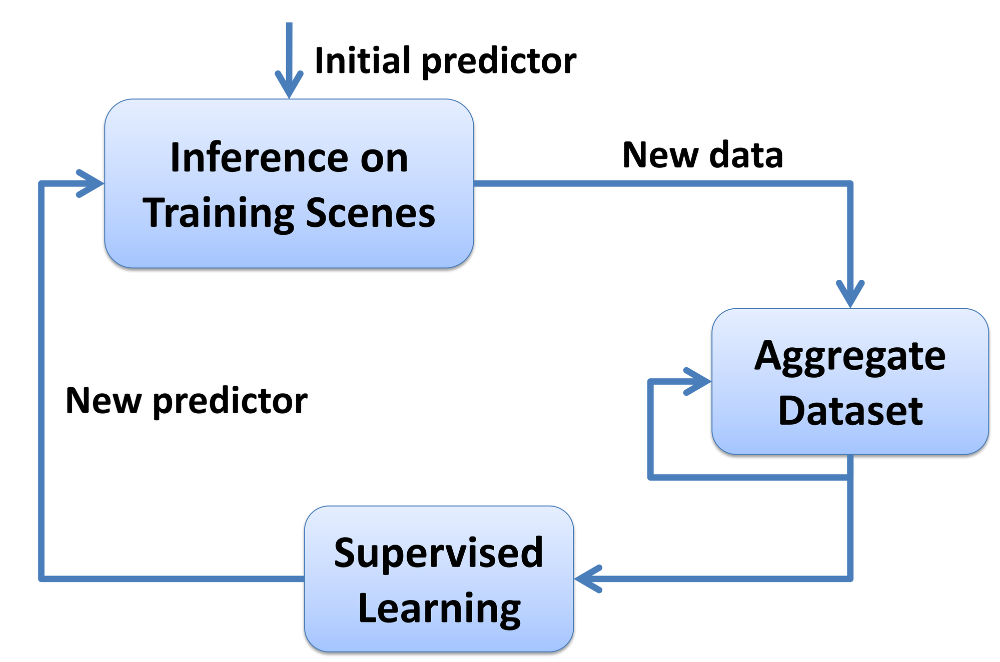


Similar to [3,4]

Learning Asynchronous Message-Passing Inference

Learn one predictor for all inference passes:

Dataset Aggregation (DAgger) [1]



Theoretical Guarantees

Forward Training:
If the predictors have ε avg. loss on the supervised learning task, then the sequence will have ε avg. loss during inference.

DAgger:
In N iterations, if we achieve ε avg. loss on the aggregate dataset, we are guaranteed to find a predictor h that has avg. loss of ε + O(1/N) during inference.

For both: Performance at inference is guaranteed to be at least as good as naive independent predictions.

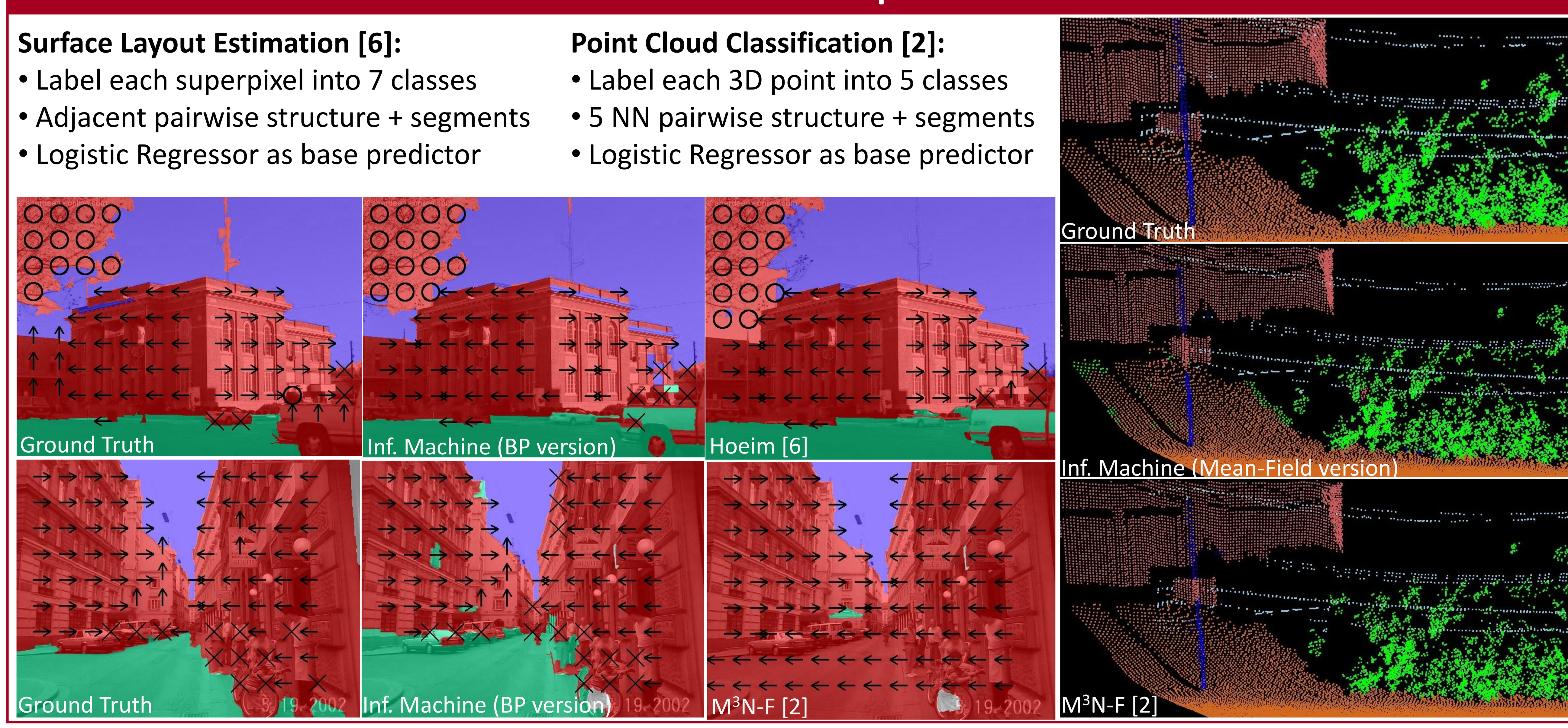
Experimental Results

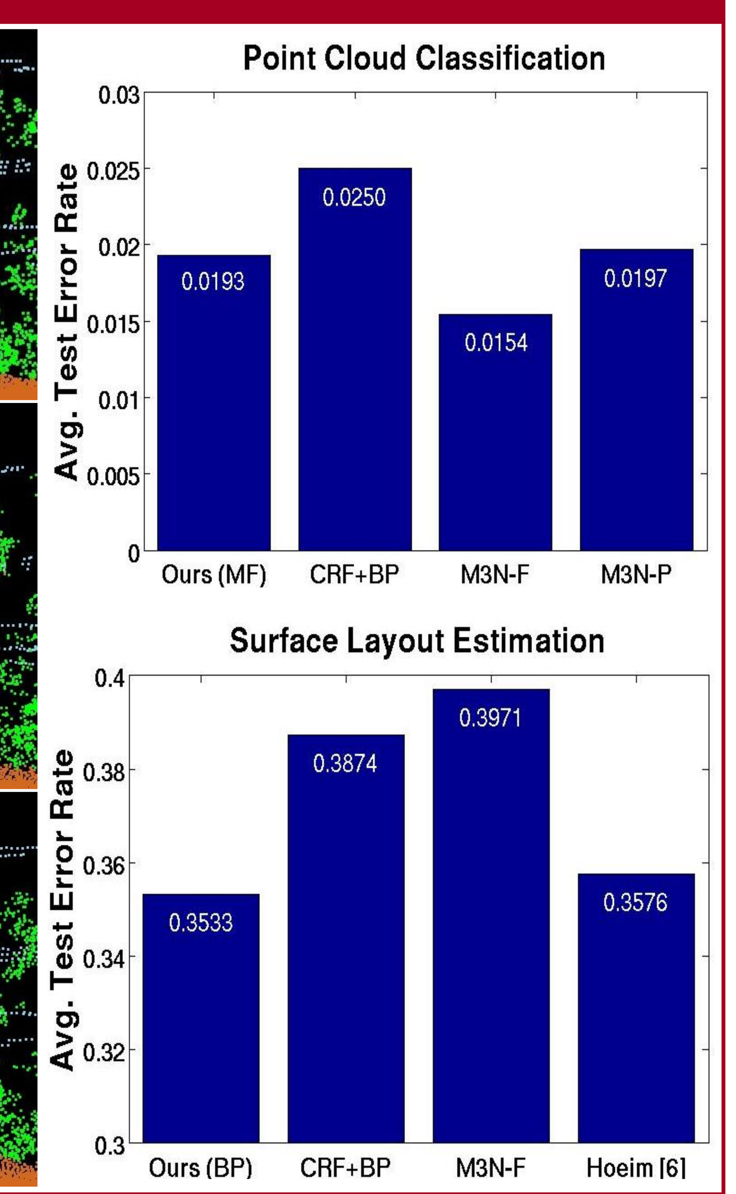
Surface Layout Estimation [6]:

- Label each superpixel into 7 classes
- Adjacent pairwise structure + segments
- Logistic Regressor as base predictor

Point Cloud Classification [2]:

- Label each 3D point into 5 classes
- 5 NN pairwise structure + segments
- Logistic Regressor as base predictor





[1] S. Ross, G. Gordon & J. A. Bagnell. *A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning*. AISTATS 2011.
[2] D. Munoz, J. A. Bagnell, N. Vandapel & M. Hebert. *Contextual Classification with Functional Max-Margin Markov Networks*. CVPR 2009.
[3] Z. Tu & X. Bai. *Auto-context and Its application to High-level Vision Tasks and 3D Brain Image Segmentation. PAMI 2010.*
[4] D. Munoz, J. A. Bagnell & M. Hebert. *Stacked Hierarchical Labeling*. ECCV 2010.
[5] A. Kulesza & F. Pereira. *Structured learning with approximate inference*. NIPS 2008.
[6] D. Hoeim, A. A. Efros & M. Hebert. *Recovering Surface Layout from an Image*. IJCV 2007