3-D Scene Analysis via Sequenced Predictions over Points and Regions

Xuehan Xiong

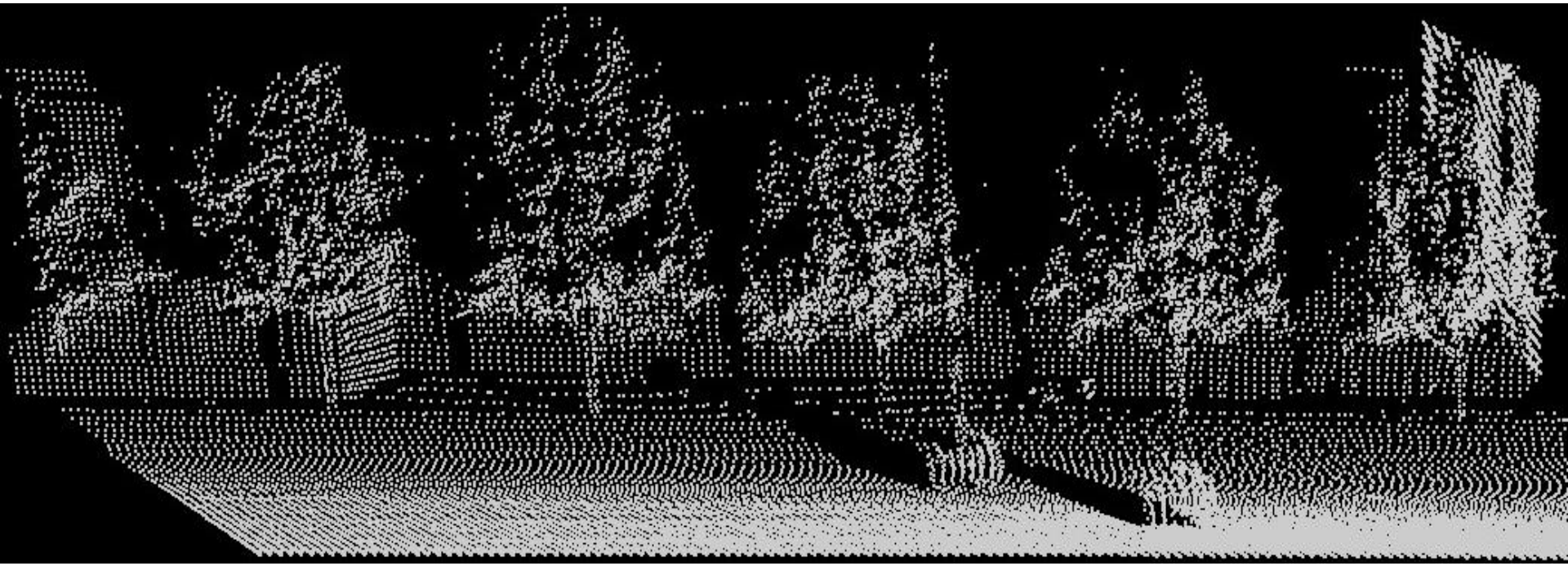
Daniel Munoz

Drew Bagnell

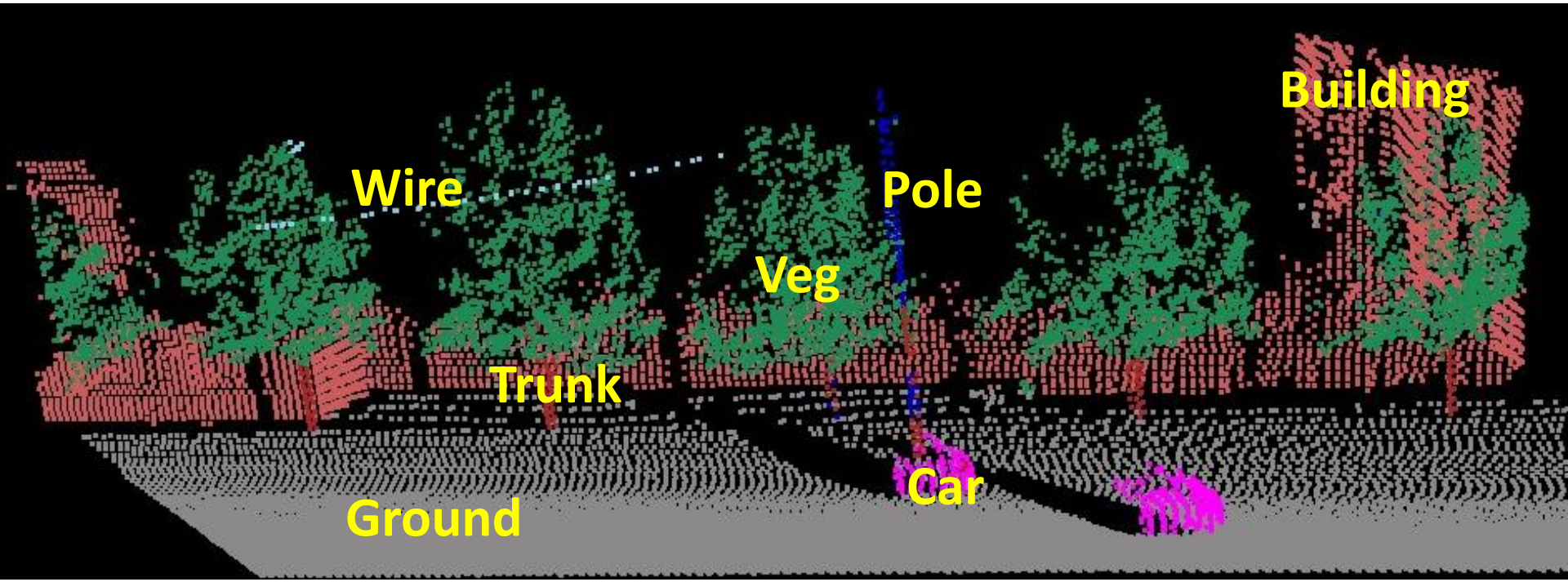
Martial Hebert

Carnegie Mellon
THE ROBOTICS INSTITUTE

Problem: 3D Scene Understanding



Solution: Contextual Classification



Classical Approach: Graphical Models

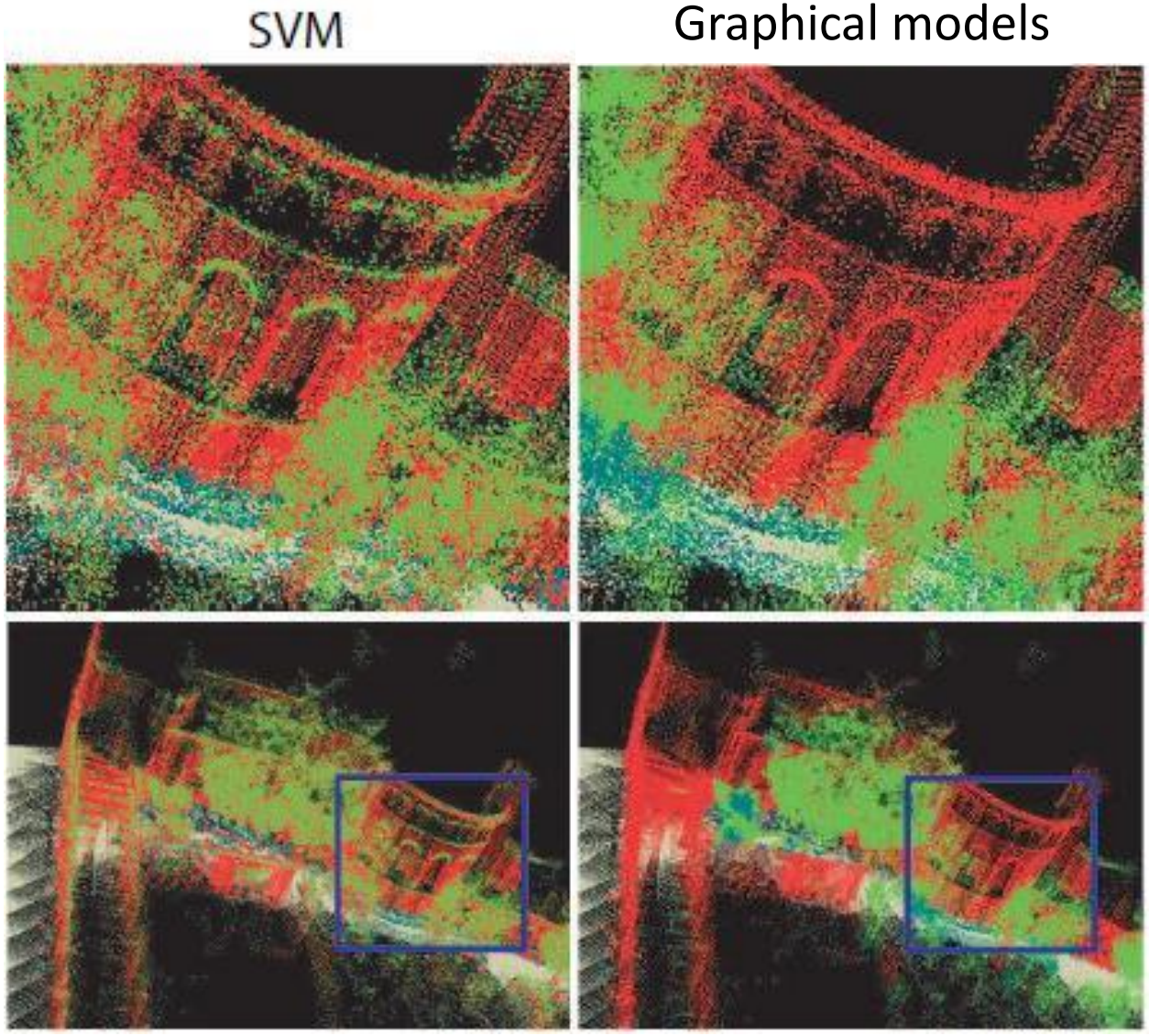


Fig. from Anguelov, et al. CVPR 2005

- Intractable inference

  Belief propagation
  Mean field
  MCMC

- Difficult to train

  Kulesza NIPS 2007
  Wainwright JMLR 2006
  Finley & Joachims ICML 2008

- Limited success

  Anguelov, et al. CVPR 2005
  Triebel, et. al. IJCAI 2007
  Munoz, et al. CVPR 2009

Classical Approach: Graphical Models

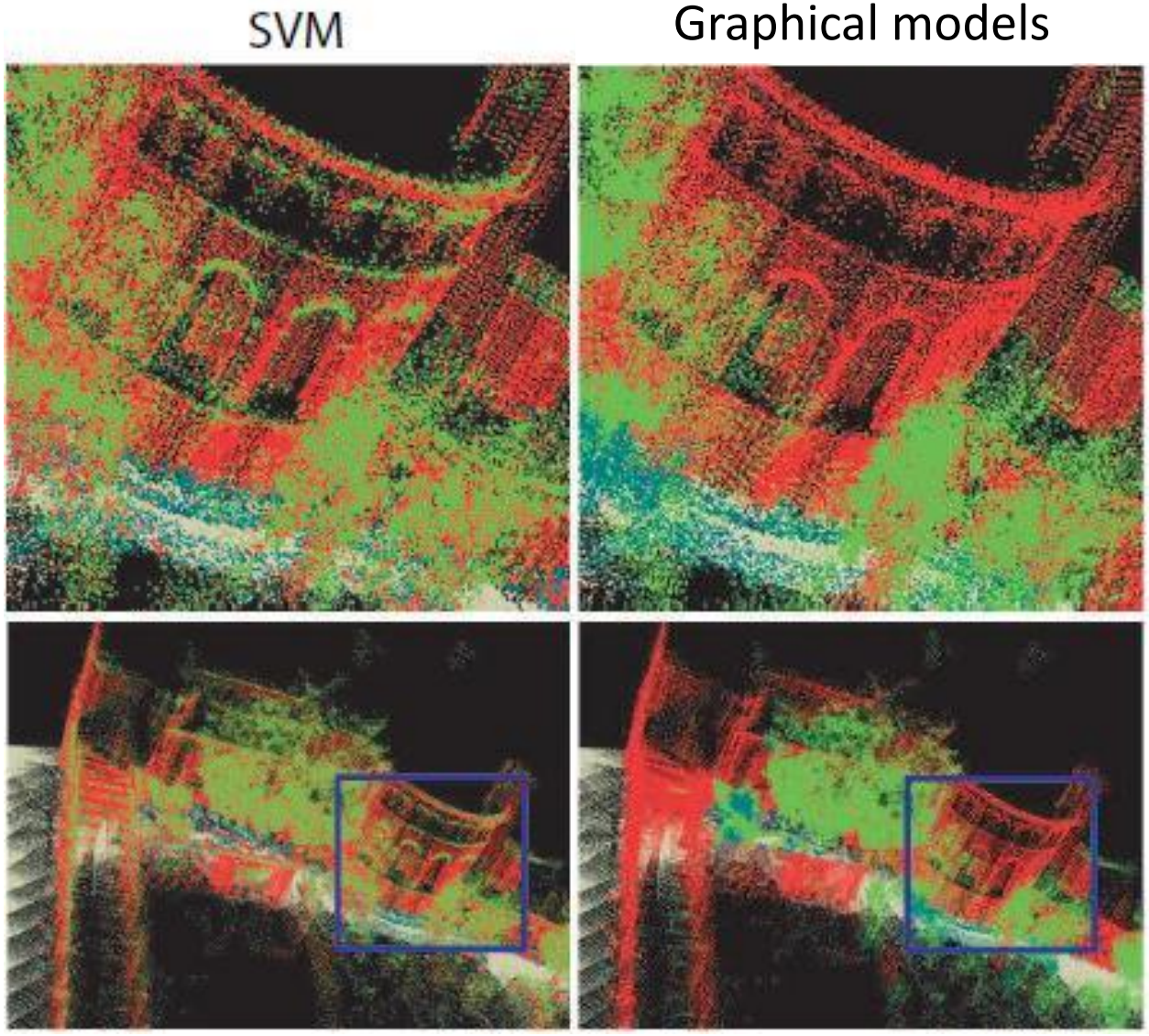


Fig. from Anguelov, et al. CVPR 2005

- Intractable inference

  Belief propagation
  Mean field
  MCMC

- Difficult to train

  Kulesza NIPS 2007
  Wainwright JMLR 2006
  Finley & Joachims ICML 2008

- Limited success

  Anguelov, et al. CVPR 2005
  Triebel, et. al. IJCAI 2007
  Munoz, et al. CVPR 2009

Classical Approach: Graphical Models

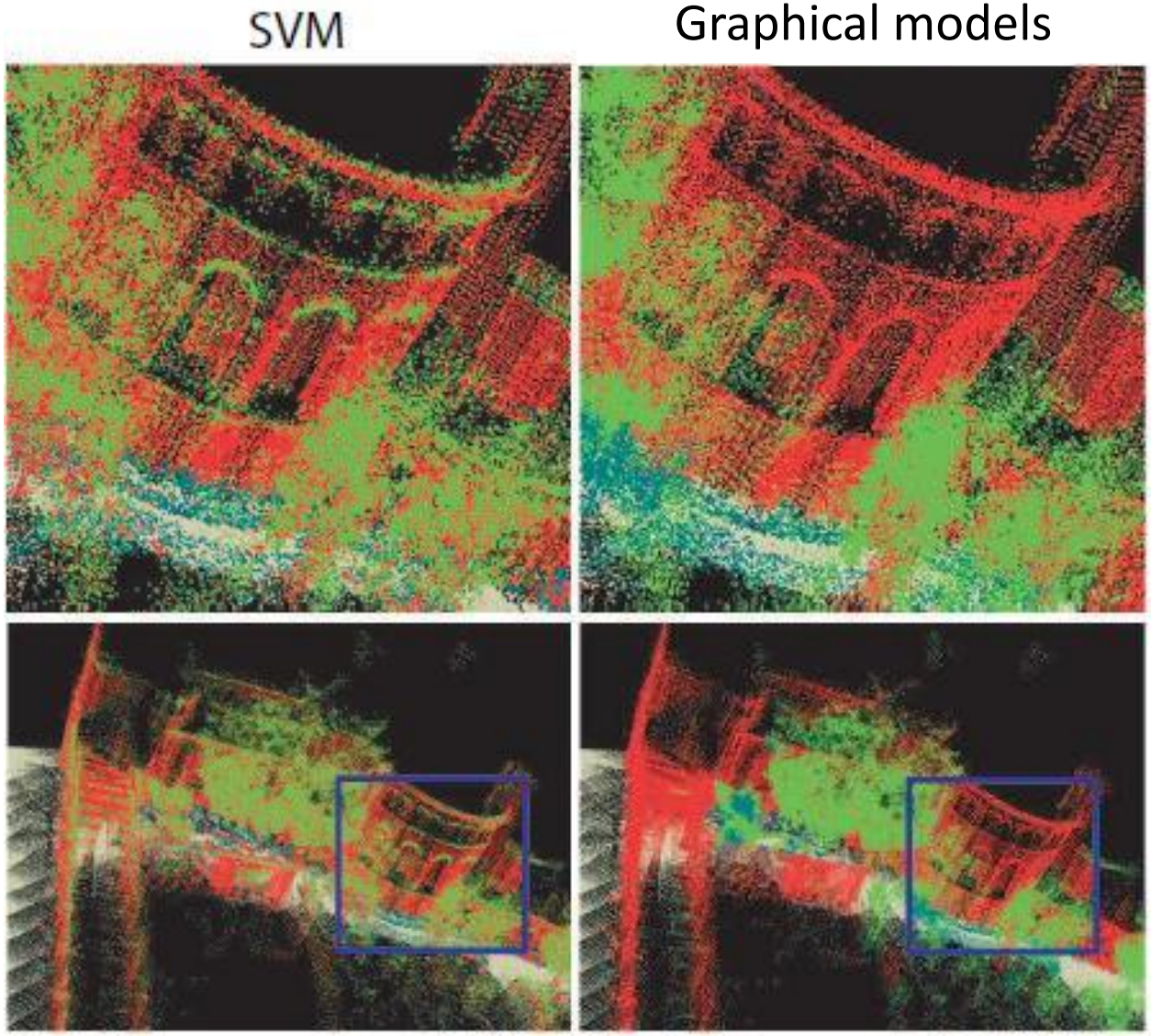


Fig. from Anguelov, et al. CVPR 2005

- Intractable inference

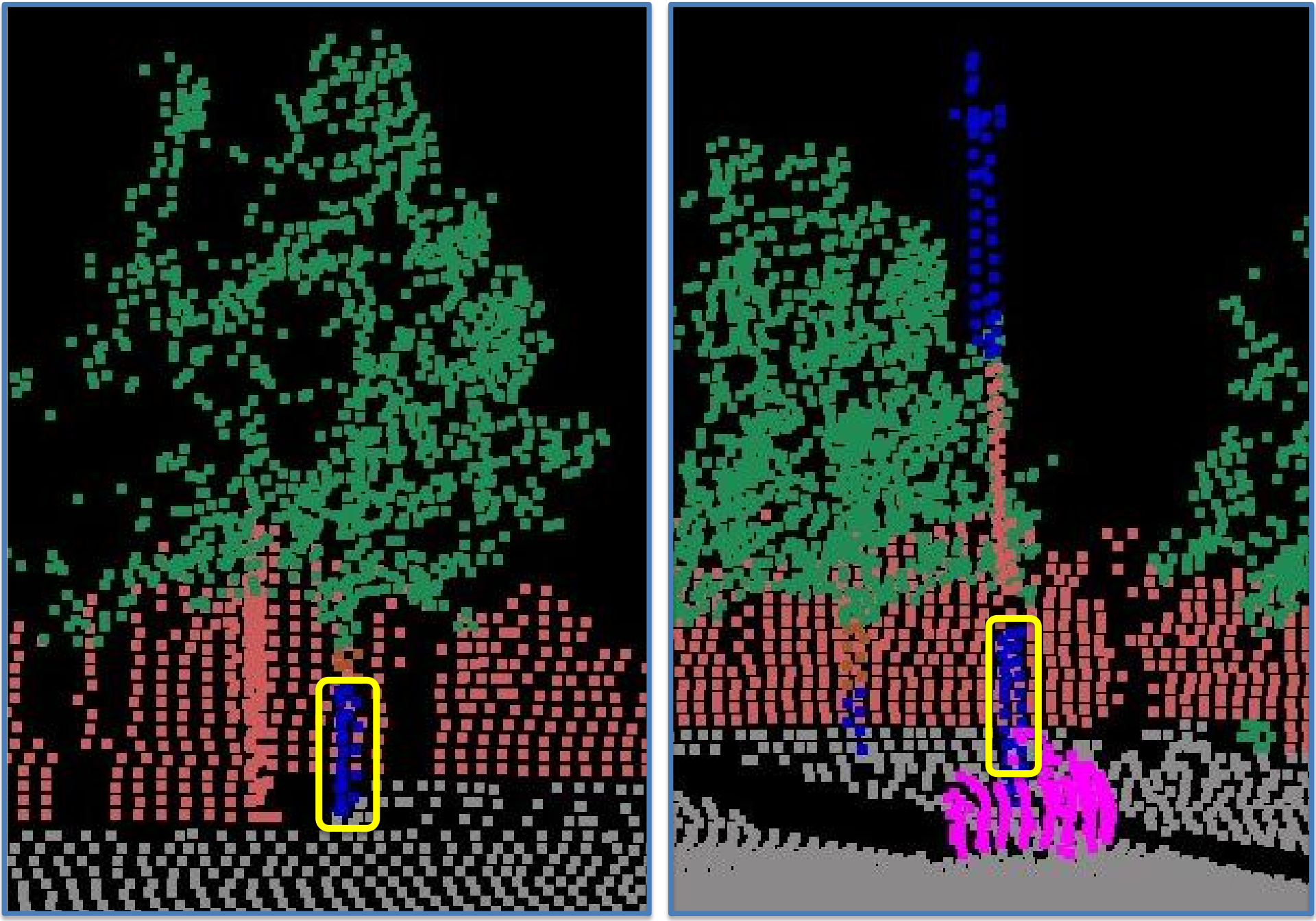
  Belief propagation
  Mean field
  MCMC

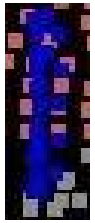
- Difficult to train

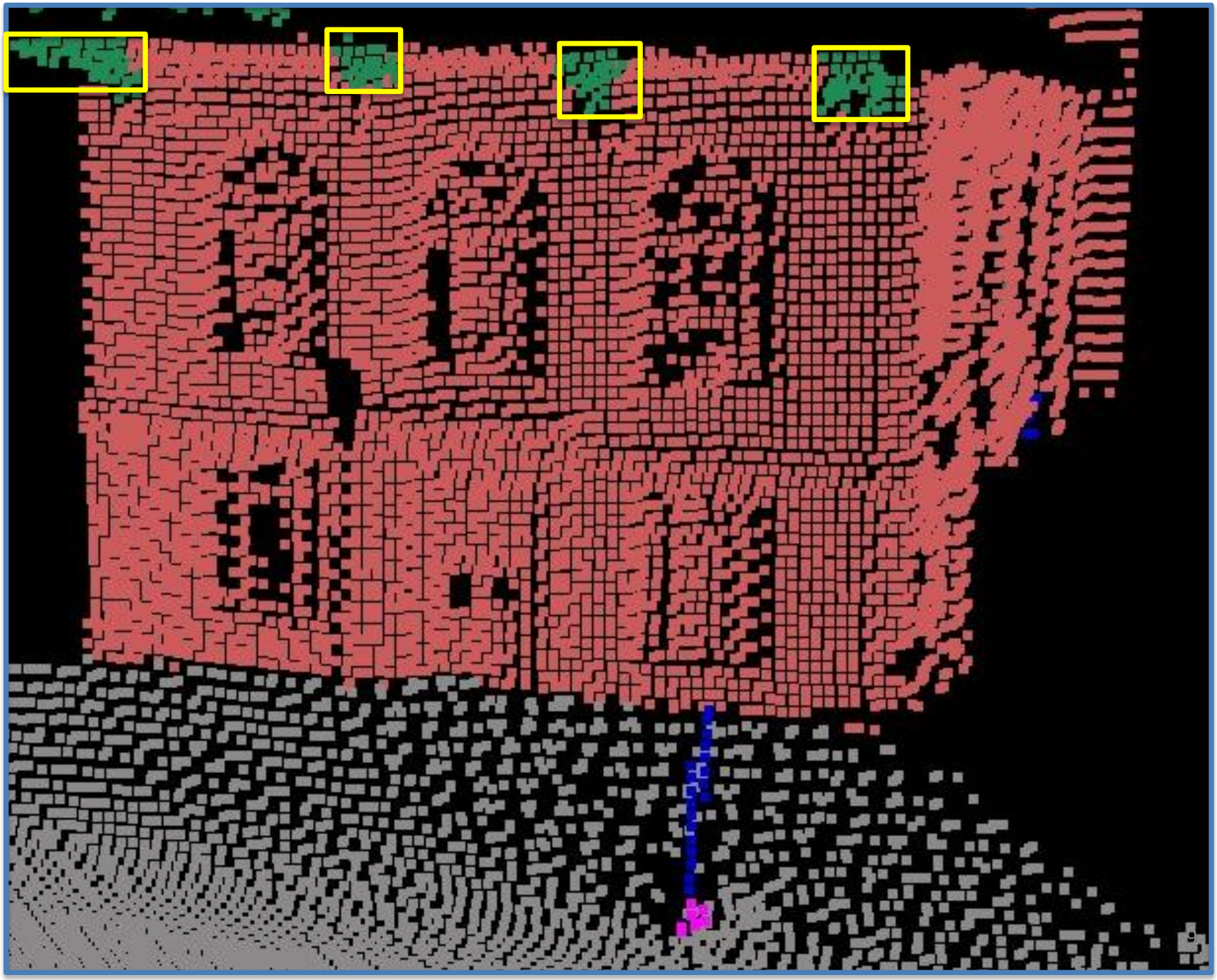
  Kulesza
  Wainwright
  Finley & Joachims ICML 2008

- Limited success

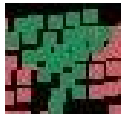
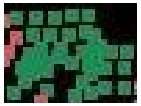
  Anguelov, et al. CVPR 2005
  Triebel, et. al. IJCAI 2007
  Munoz, et al. CVPR 2009









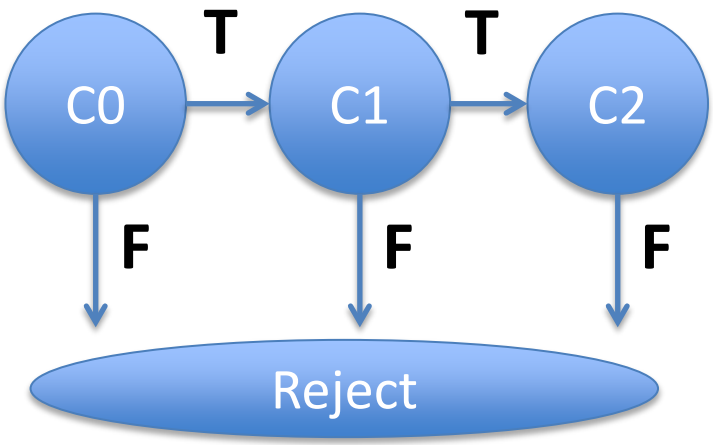



Our Approach: Inference Machines

- Train an inference **procedure**, not a model.
  - To encode spatial layout and long range relations
  - Daume III 2006, Tu 2008, Bagnell 2010, Munoz 2010

Our Approach: Inference Machines

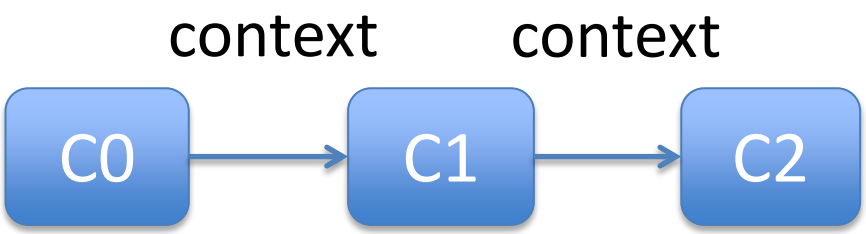
- Train an inference **procedure**, not a model.
  - To encode spatial layout and long range relations
  - Daume III 2006, Tu 2008, Bagnell 2010, Munoz 2010
- Inference via sequential prediction



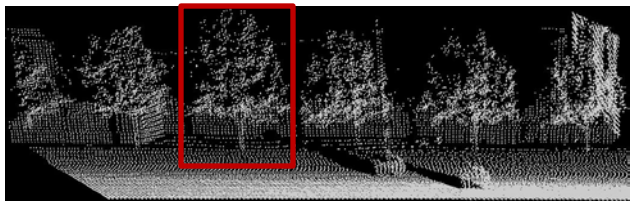
E.g. Viola-Jones 2001

Our Approach: Inference Machines

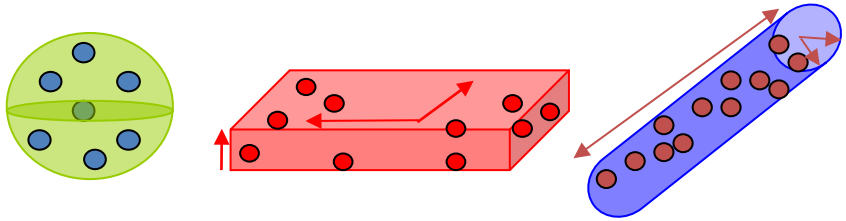
- Train an inference **procedure**, not a model.
  - To encode spatial layout and long range relations
  - Daume III 2006, Tu 2008, Bagnell 2010, Munoz 2010
- Inference via sequential prediction



Ours

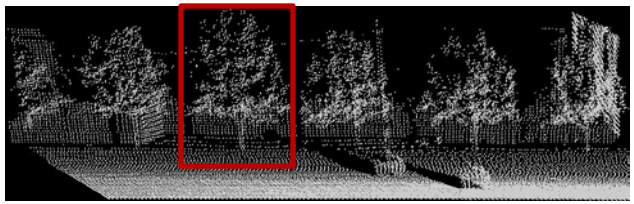


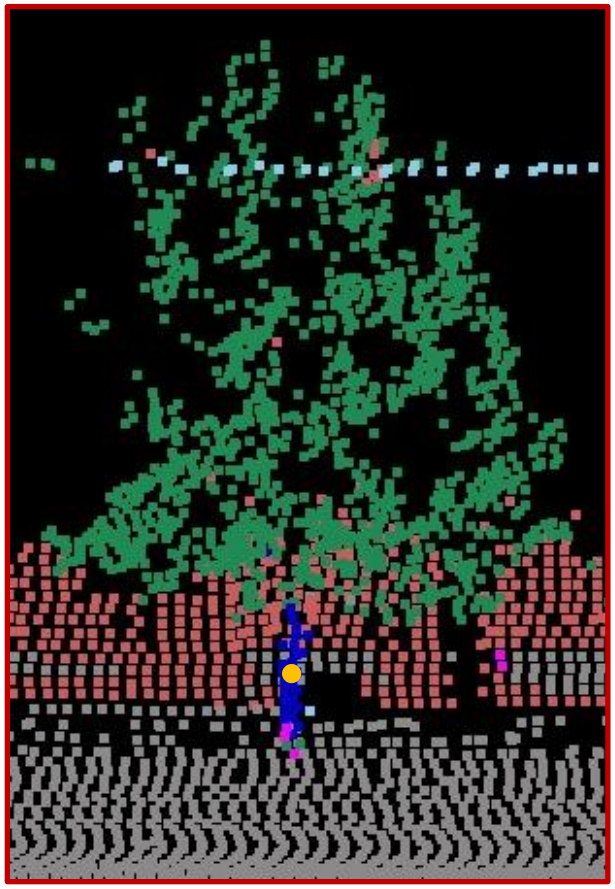
Example features



$\vec{x}_i^{(0)}$: point features

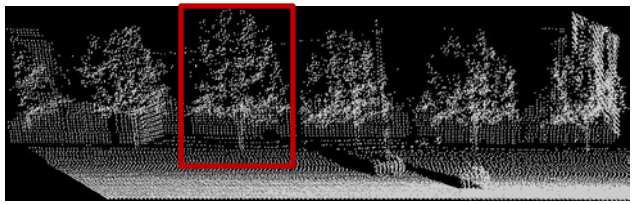
$$\hat{Y}^{(0)} = \mathrm{LogReg}^{(0)}(X^{(0)})$$

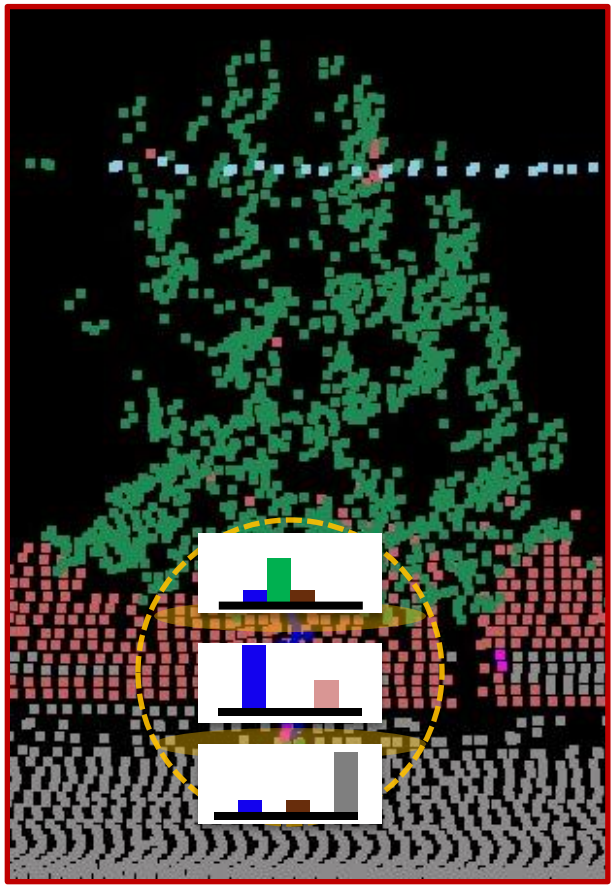


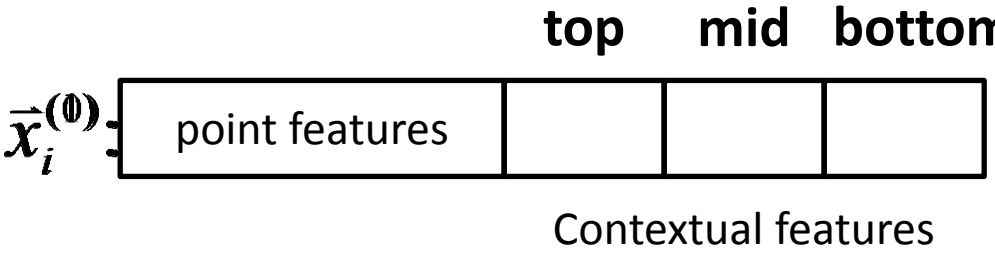


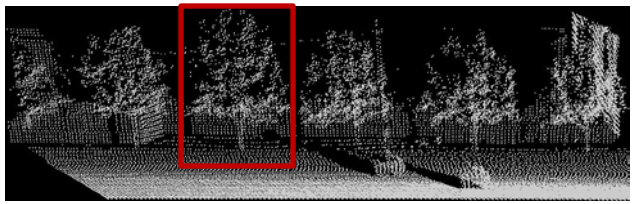
$\vec{x}_i^{(0)}$: point features

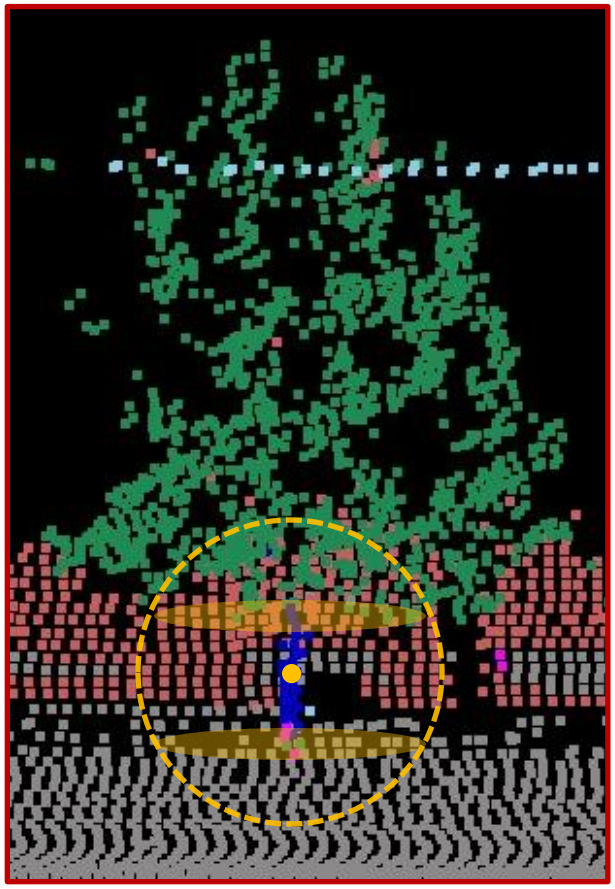
$$\arg\max(\hat{Y}^{(0)})$$

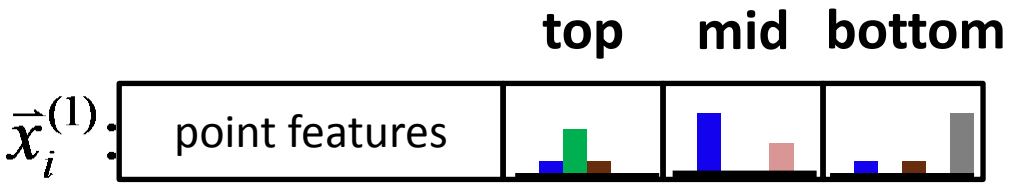




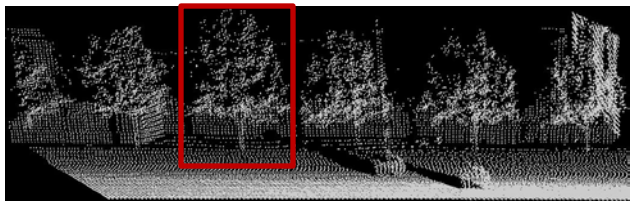

top
mid
bottom
$\vec{x}_i^{(0)}$
point features
Contextual features

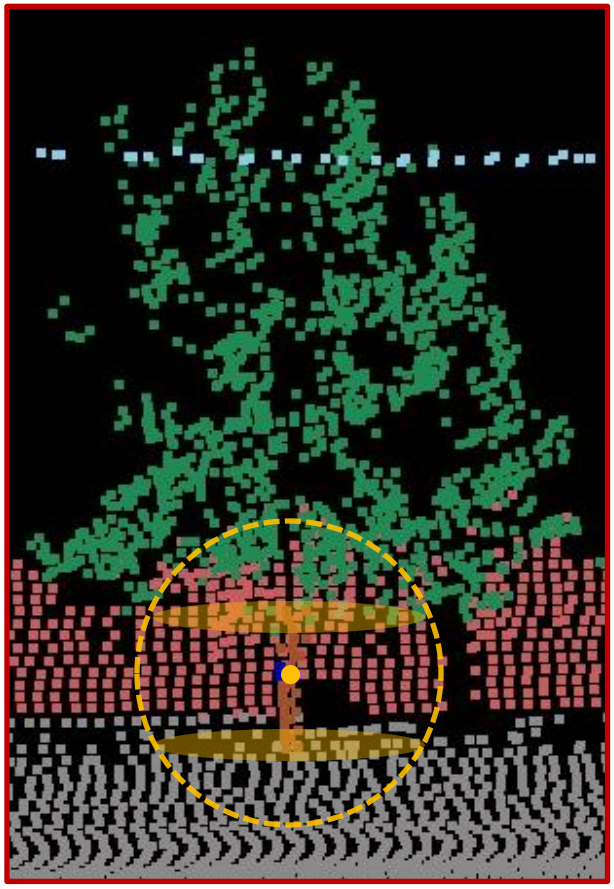






$$\hat{Y}^{(1)} = \text{LogReg}^{(1)}(X^{(1)})$$



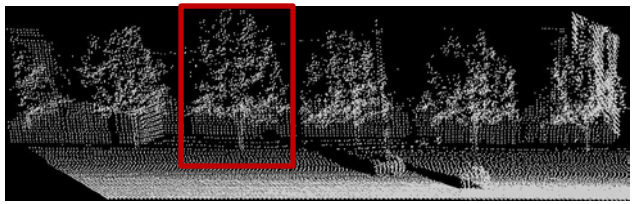


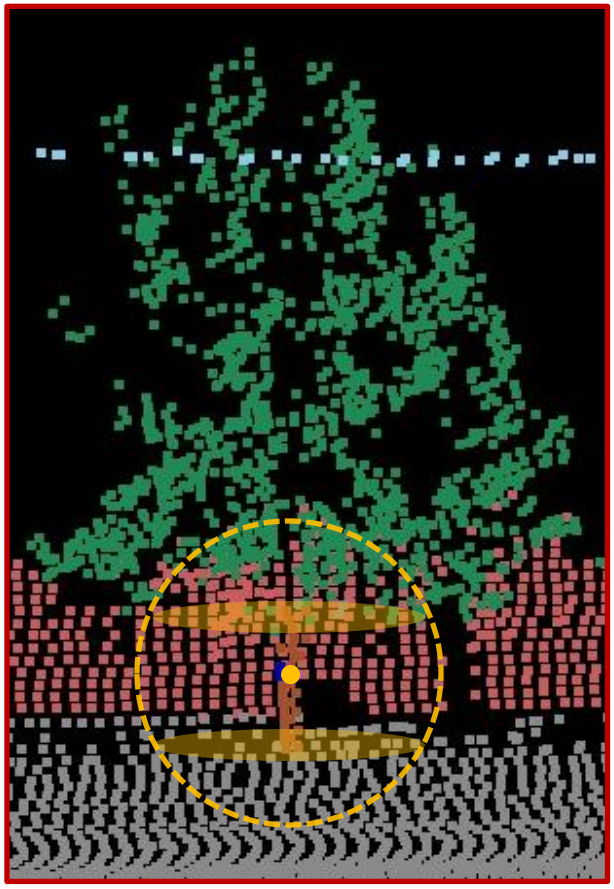
$$\arg\max(\hat{Y}^{(1)})$$

top mid bottom

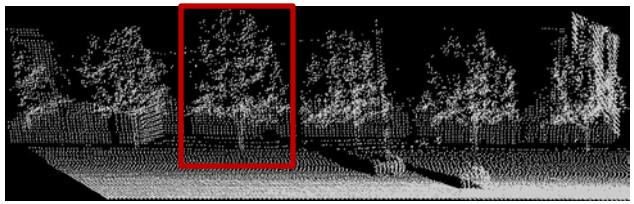
$$\vec{x}_i^{(1)}:$$ point features

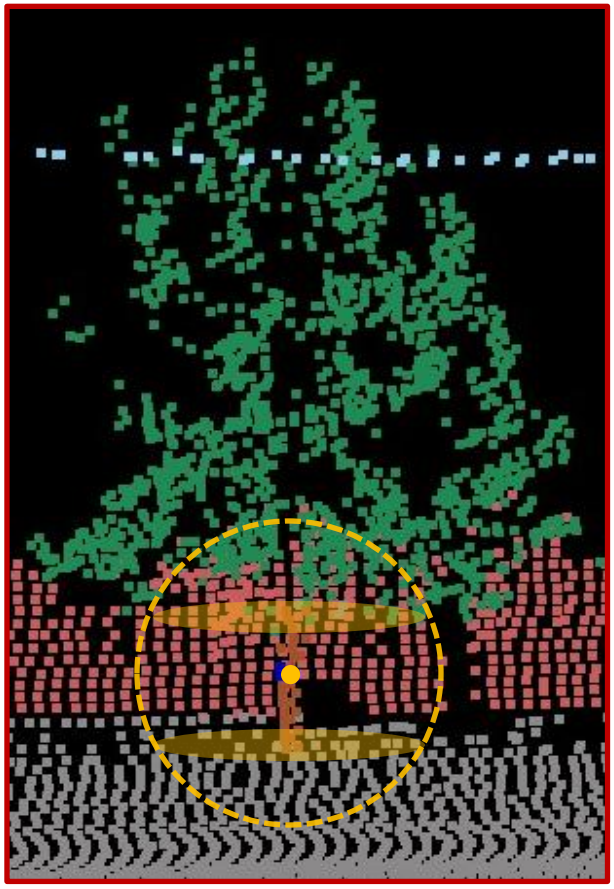
$$\hat{Y}^{(1)} = \text{LogReg}^{(1)}(X^{(1)})$$





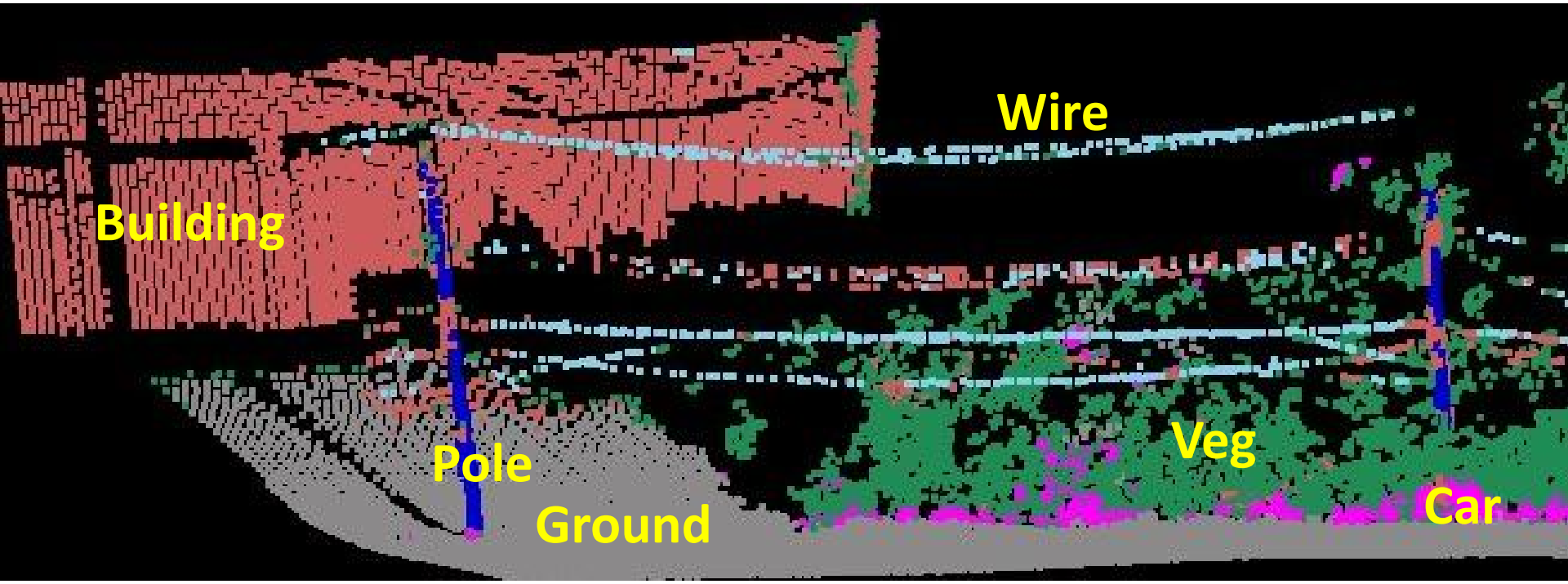
$$\hat{Y}^{(1)} = \text{LogReg}^{(1)}(X^{(1)})$$



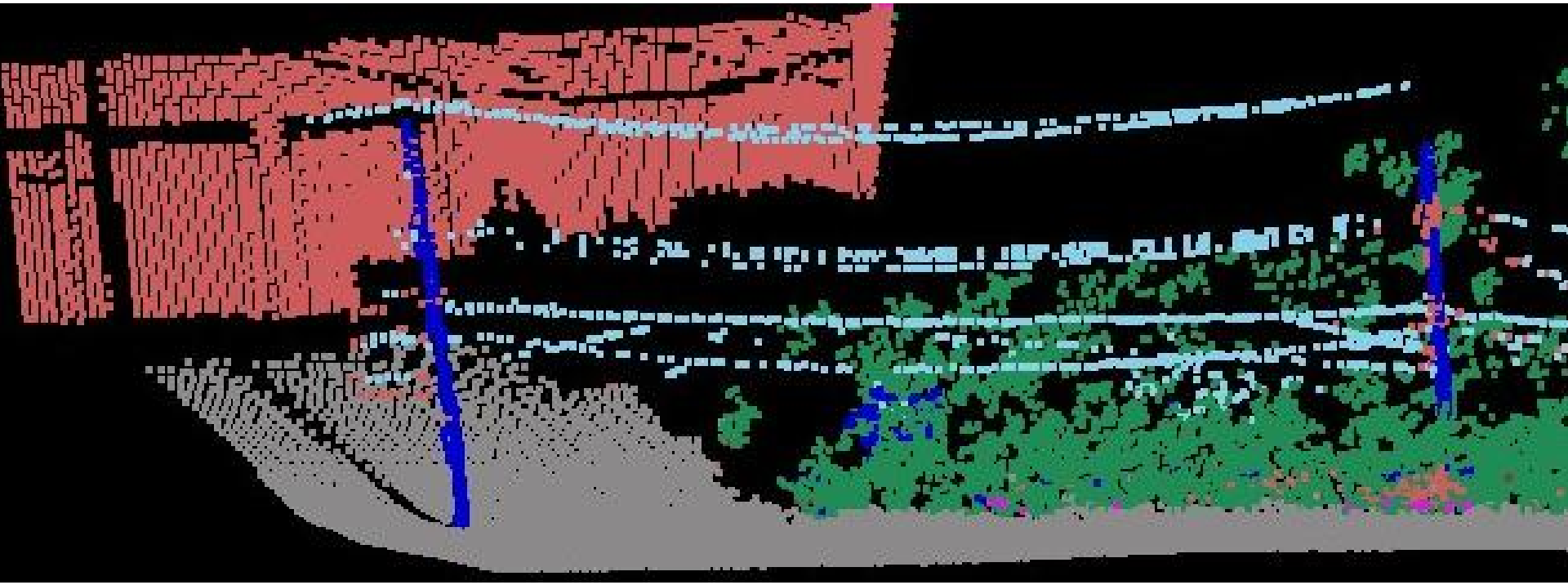


$$\hat{Y}^{(2)} = \mathrm{LogReg}^{(2)}(X^{(2)})$$

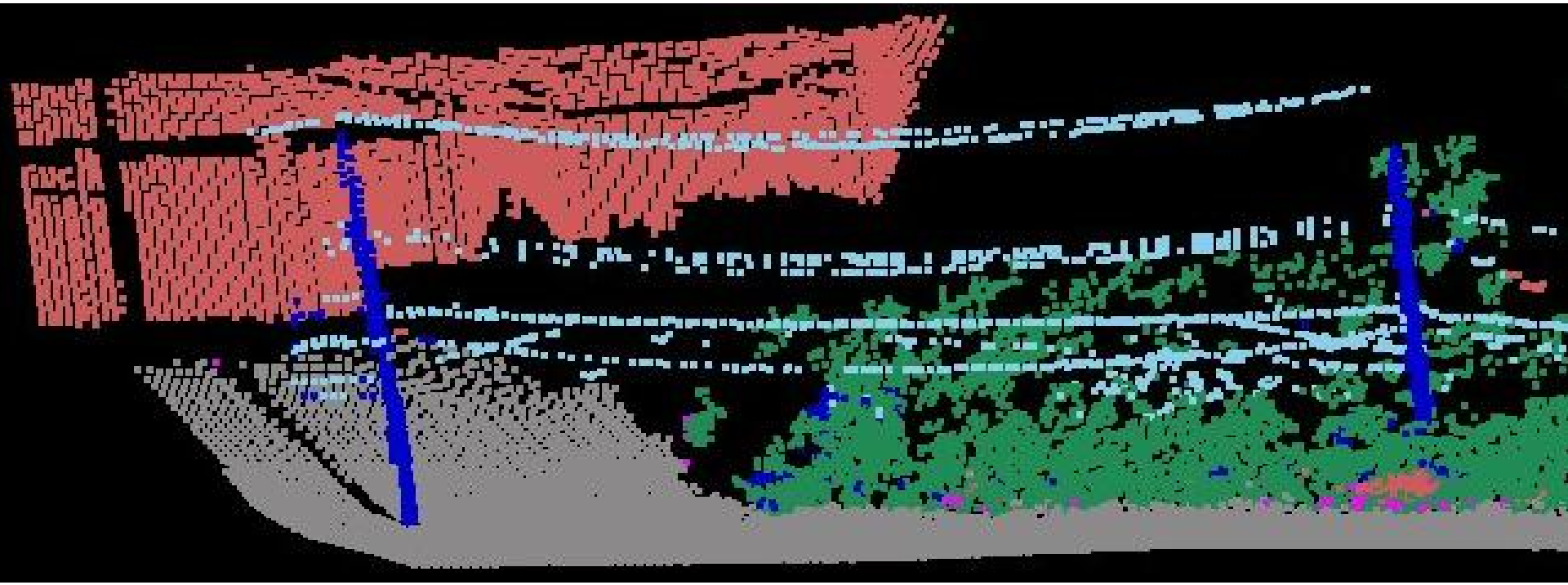
Local features only



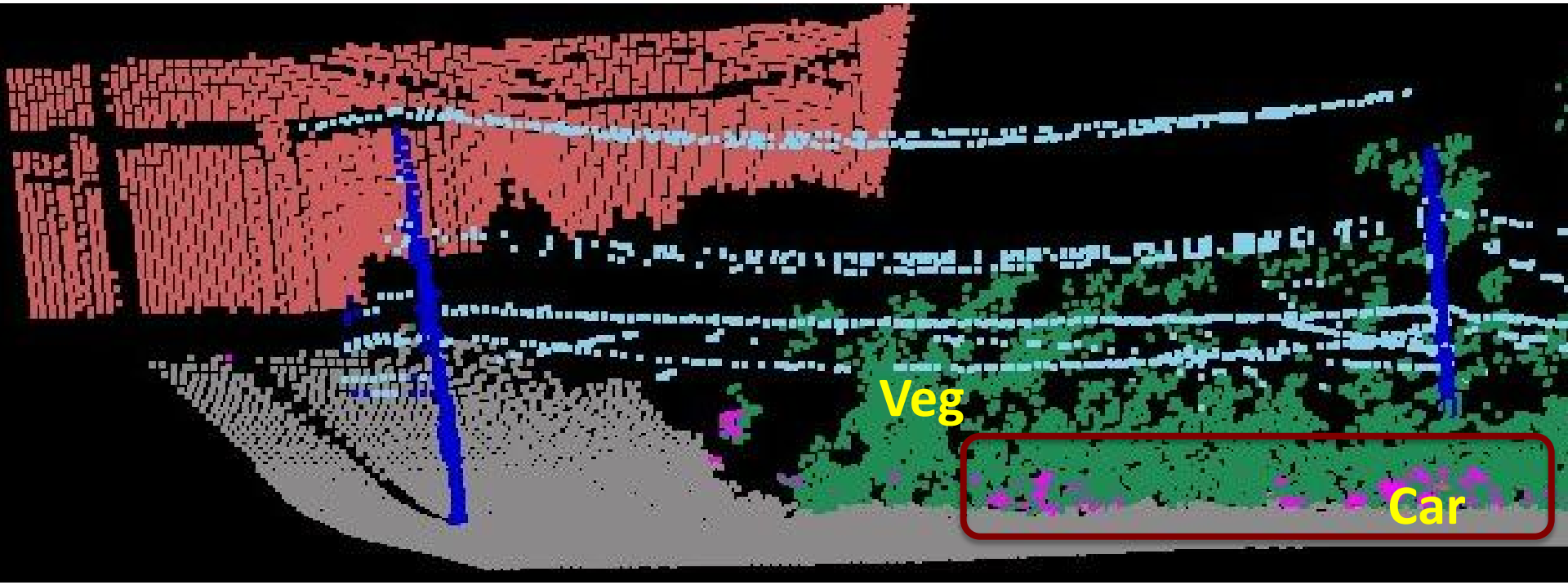
Round 1



Round 2

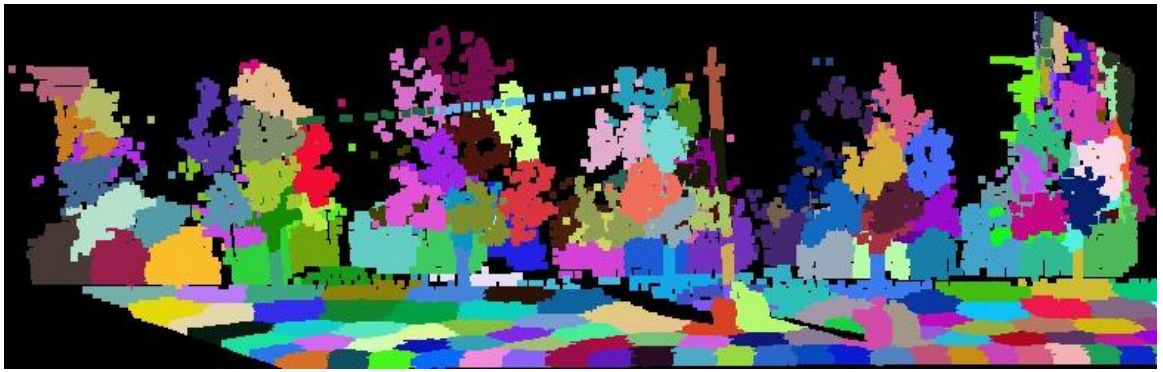


Round 3

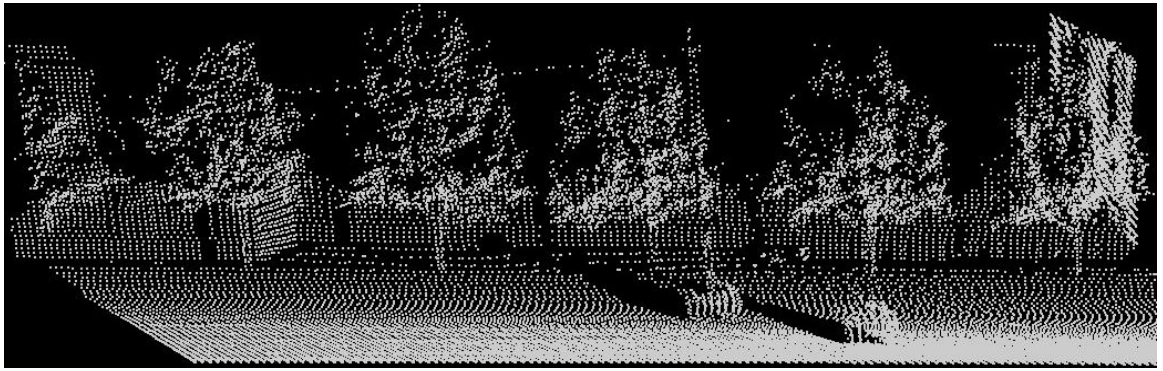


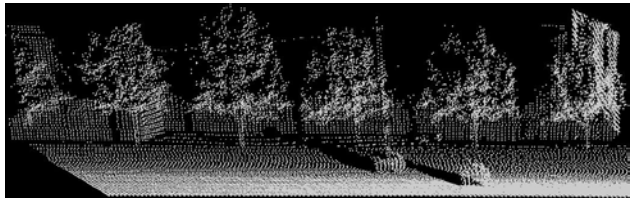
Create regions

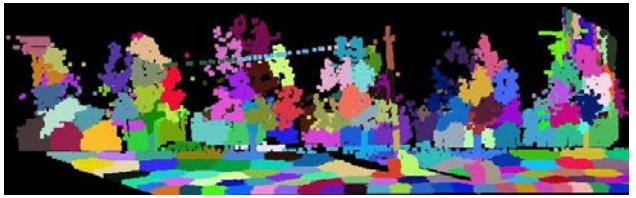
Level 2

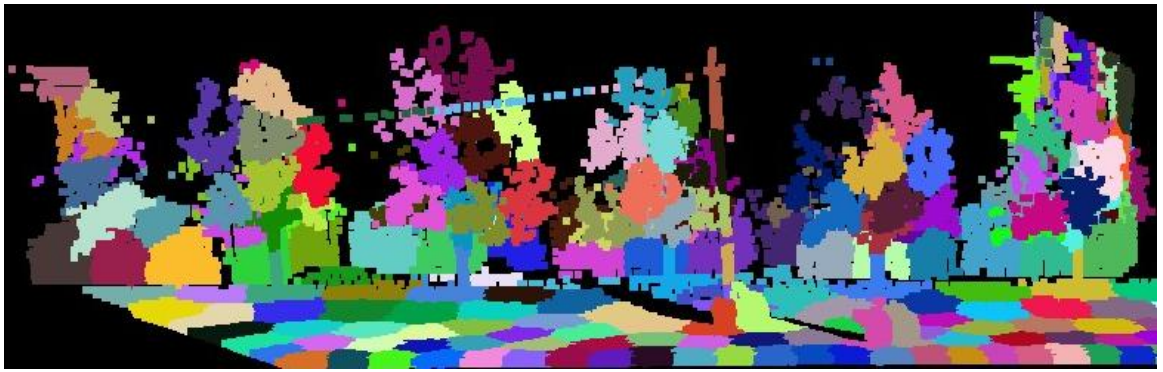


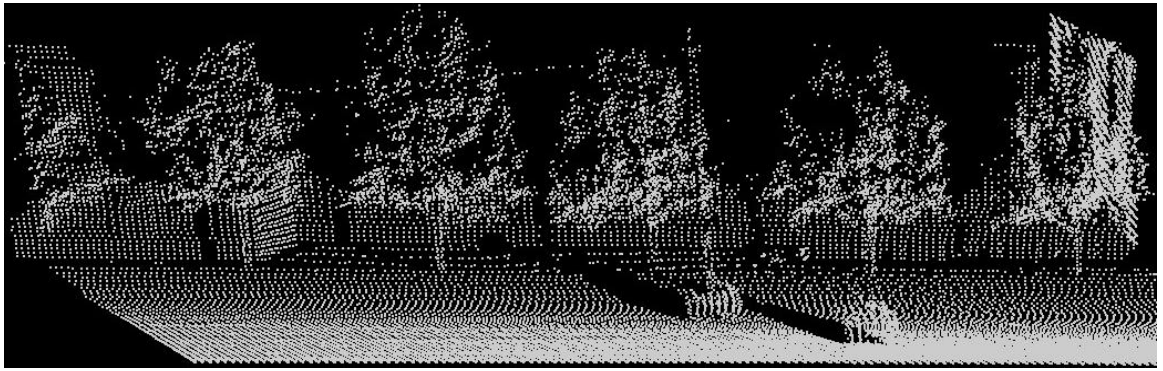
Level 1

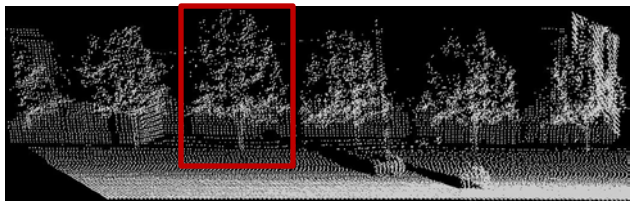


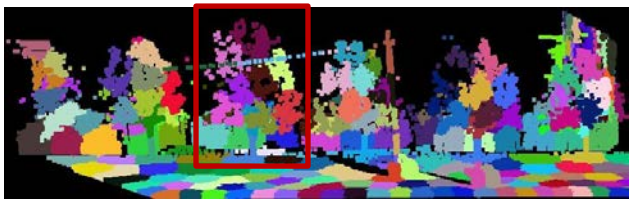


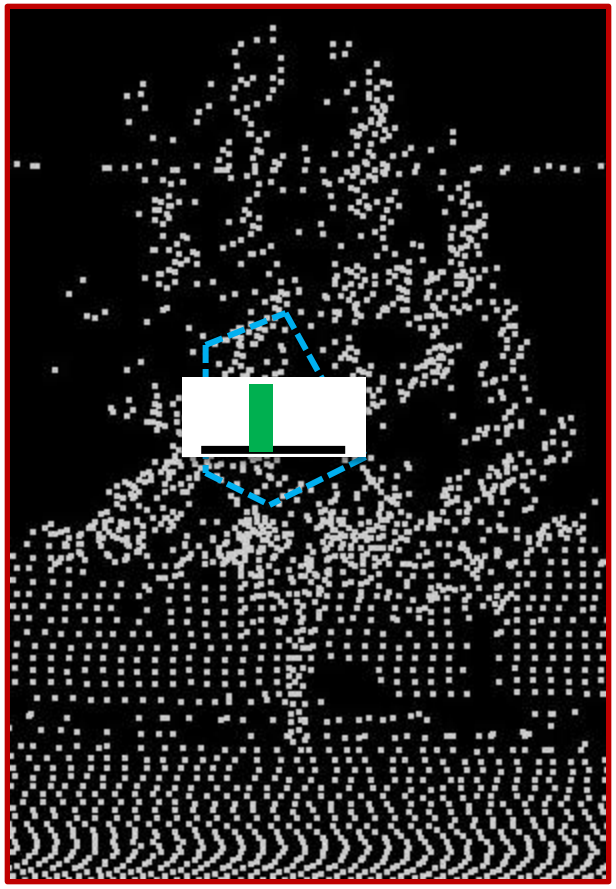


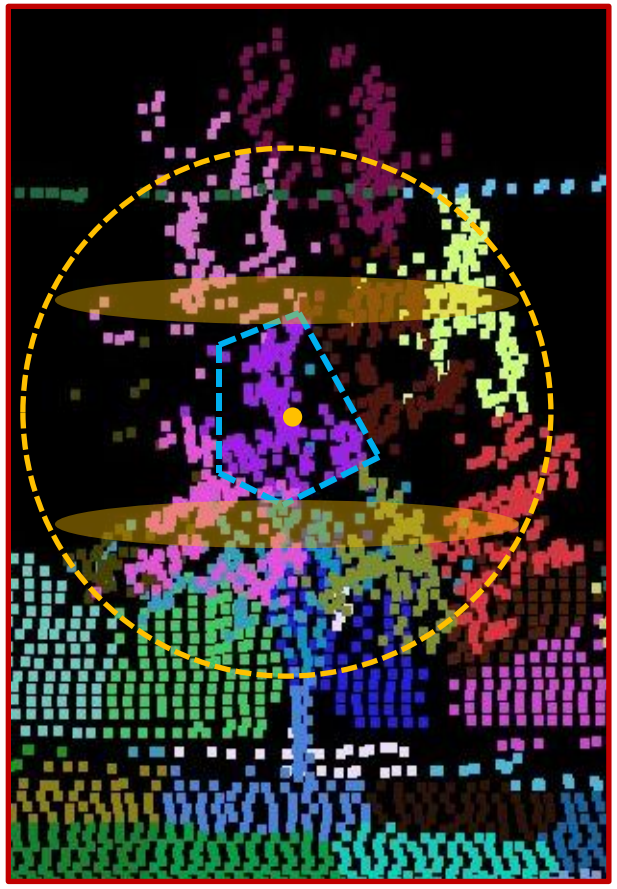


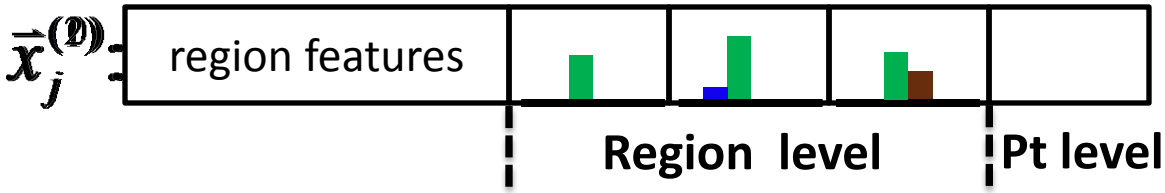




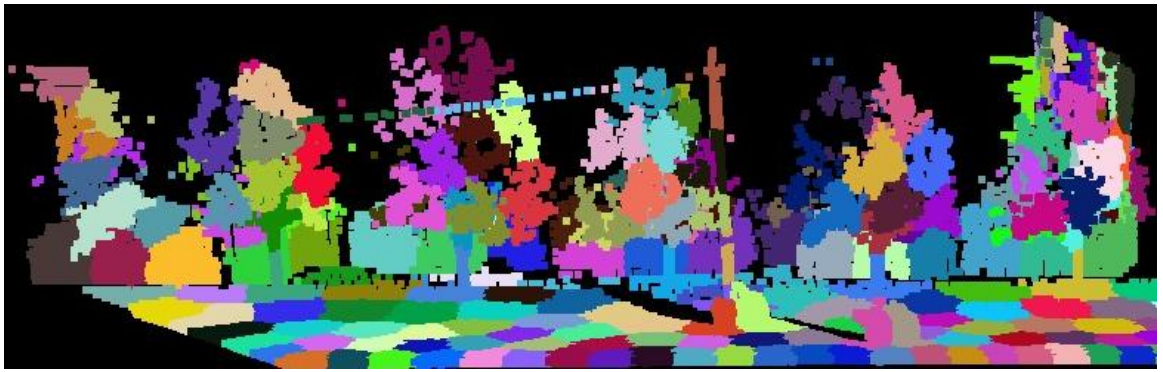




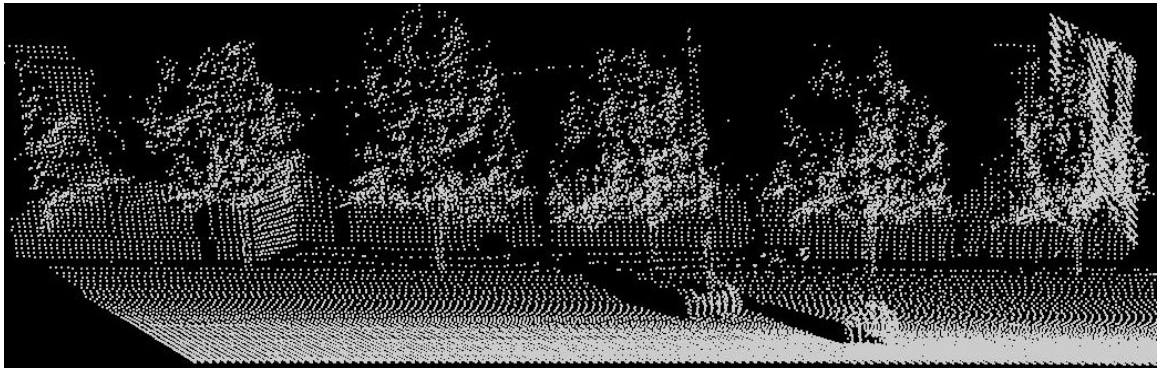


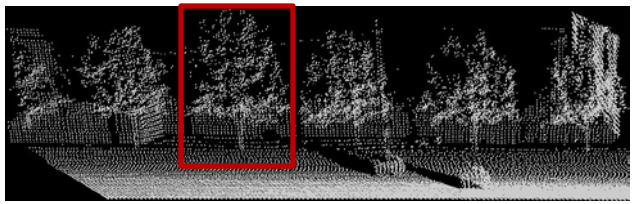

$\vec{x}_j^{(l)}$
region features
Region level
Pt level

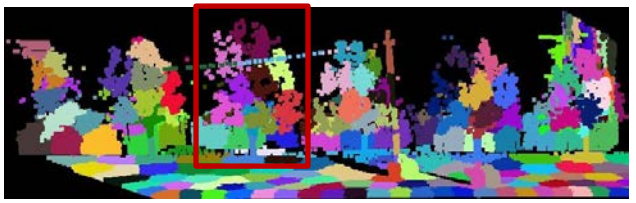
Level 2

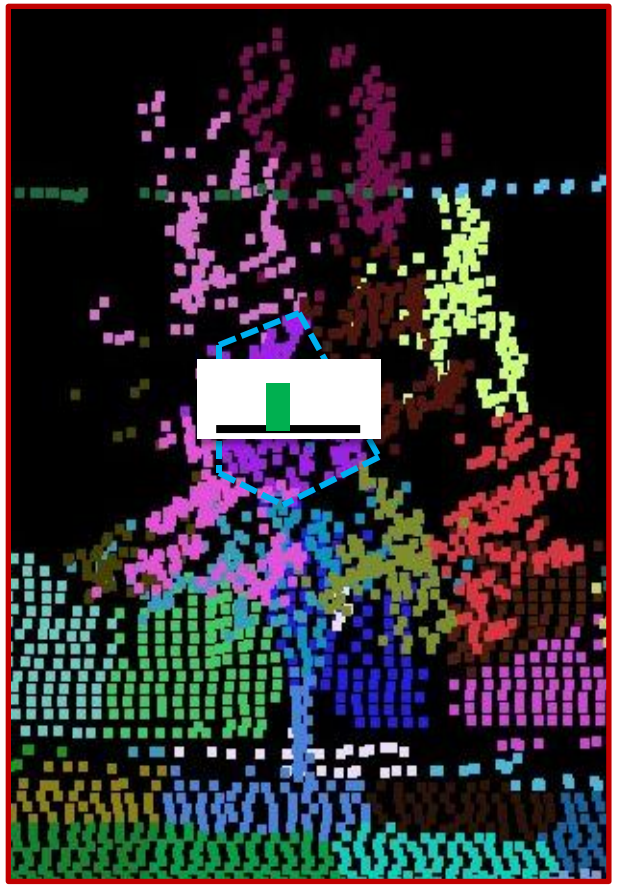


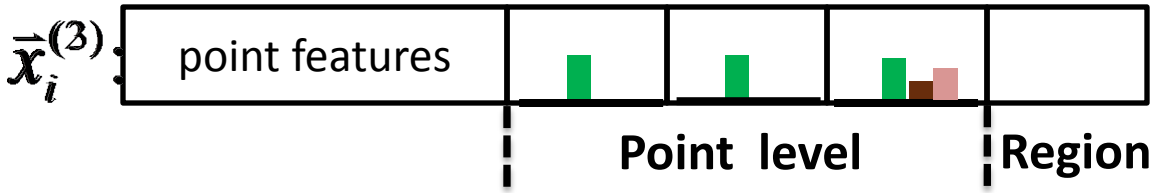
Level 1



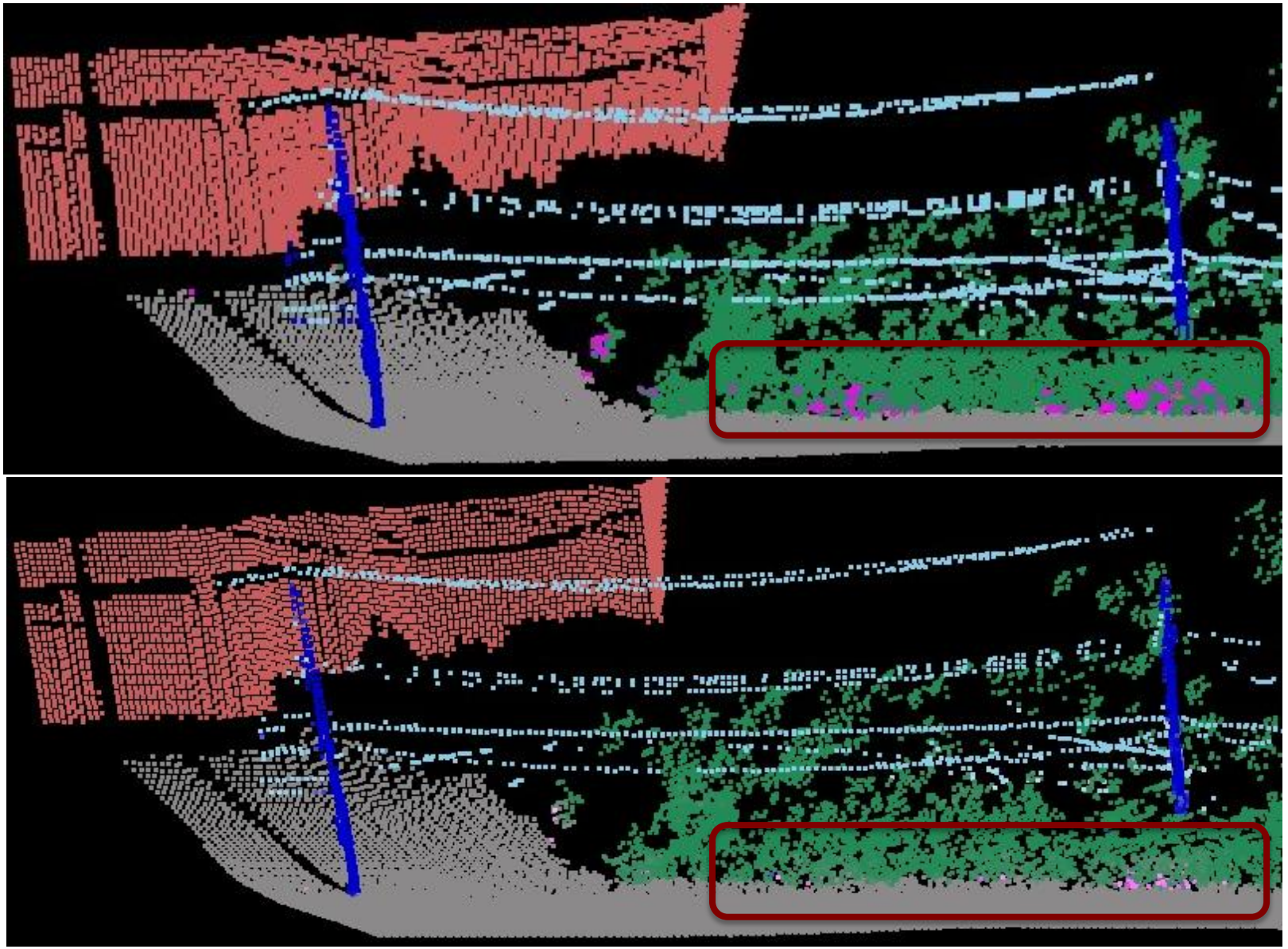






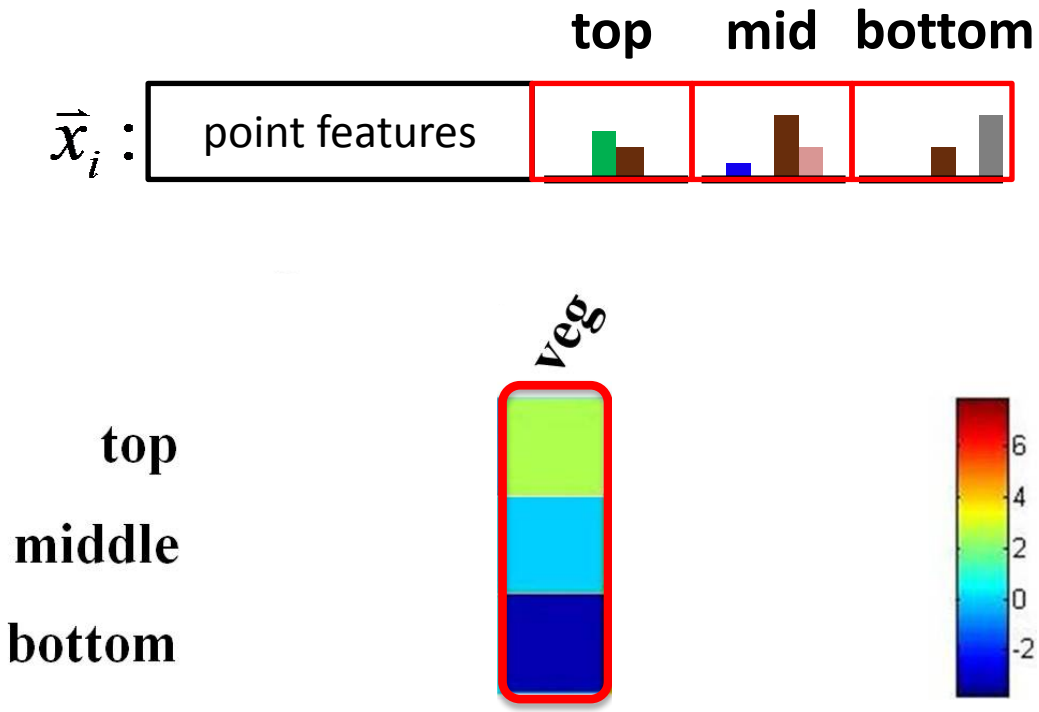

$\vec{x}_i^{(3)}$
point features
Point level
Region level

With Regions



Learned Relationships

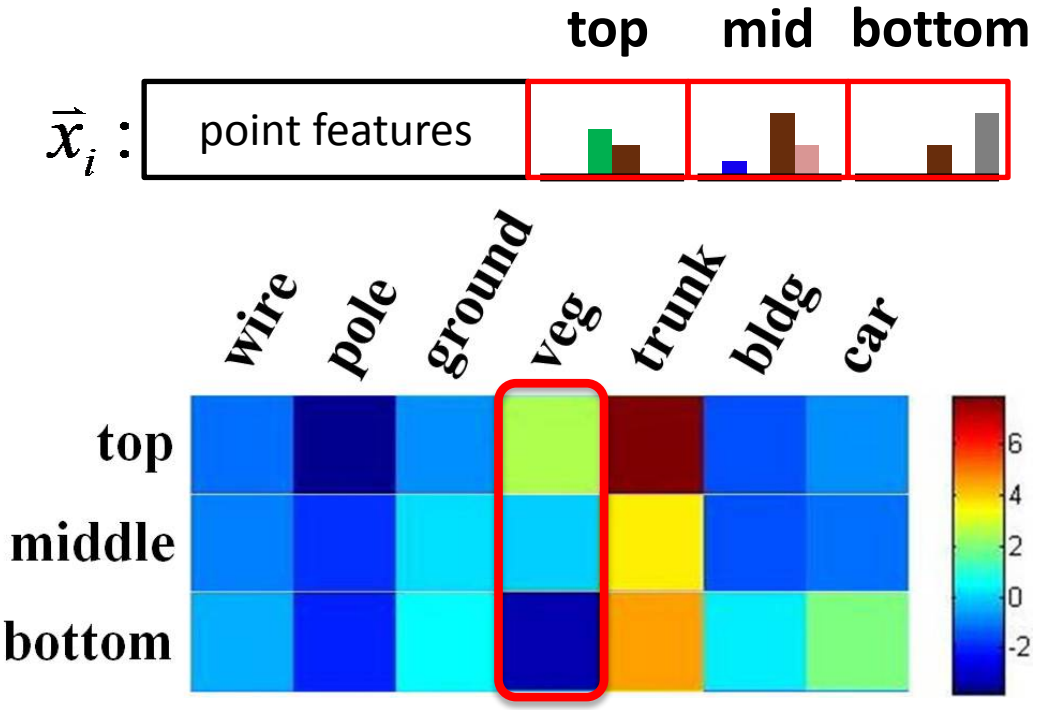
Neighbor contextual feature



Learned weights

Learned Relationships

Neighbor contextual feature

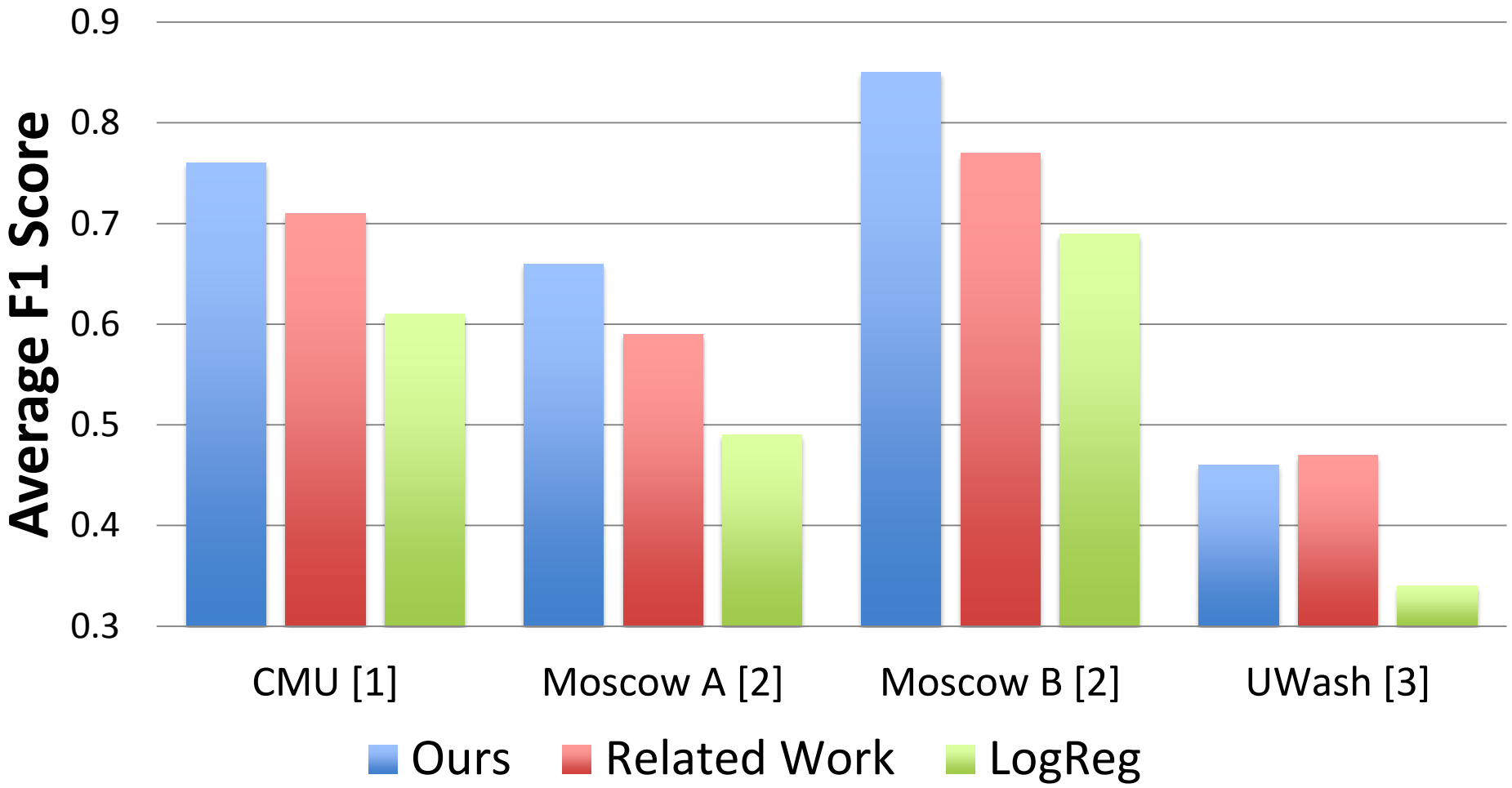


Learned weights

Experiments

- 3 large-scale datasets
  - CMU (26M), Moscow State (10M), Univ. Wash (10M)
- Multiple classes (4 to 8)
  - car, building, veg, wire, fence, people, trunk, pole, ground, street sign
- Different sensors
  - SICK (ground), ALTM 2050 (aerial), Velodyne (ground)
- Comparisons
  - Graphical models, exemplar based

Quantitative Results



[1] Munoz CVPR 2009 [2] Shapovalov PCV 2010 [3] Lai RSS 2010 *

* Use additional semi-supervised data not leveraged by other methods.

[1] Munoz, et. al. CVPR 2009

CMU Dataset

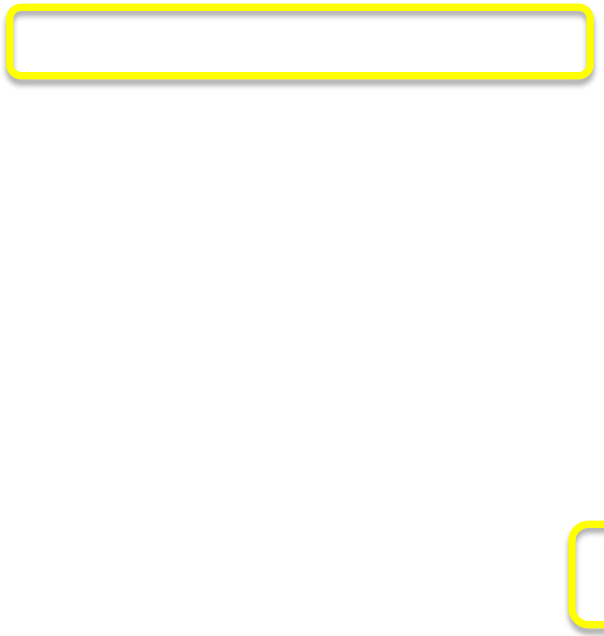
Ours

Max Margin CRF [1]

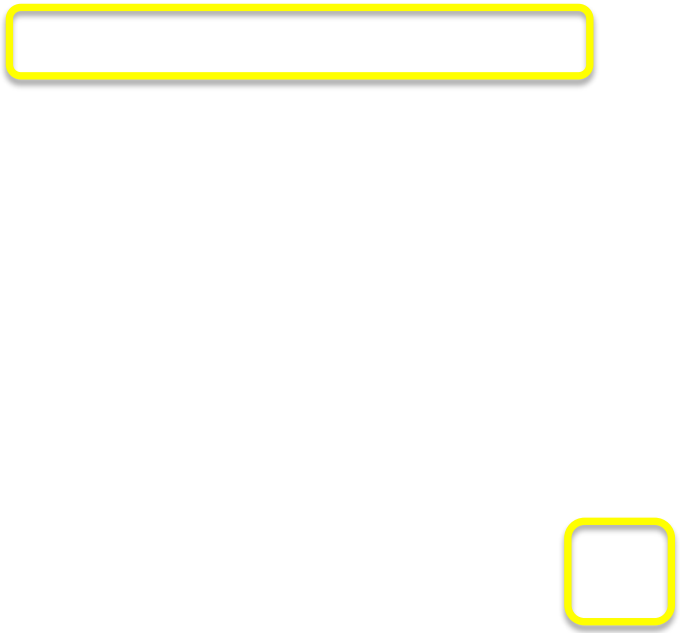
[1] Munoz, et. al. CVPR 2009

CMU Dataset

Ours



Max Margin CRF [1]

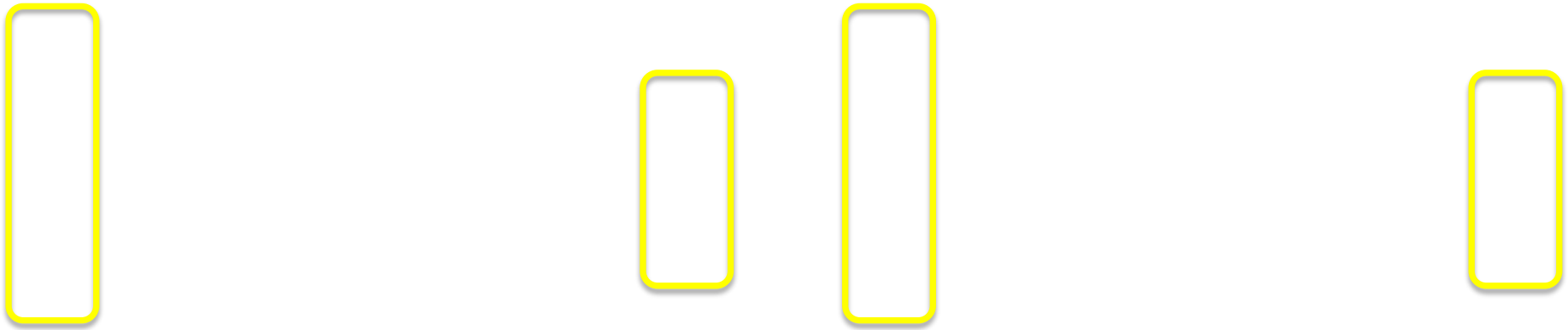


[1] Munoz, et. al. CVPR 2009

CMU Dataset

Ours

Max Margin CRF [1]

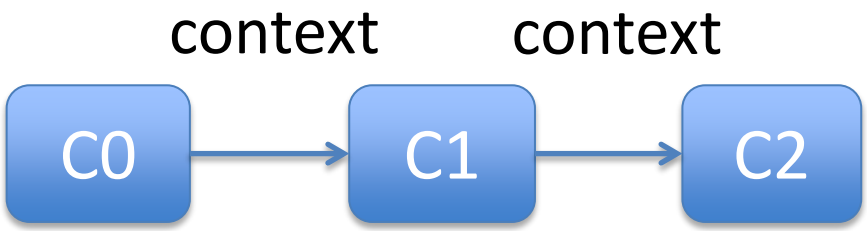


Moscow State Dataset

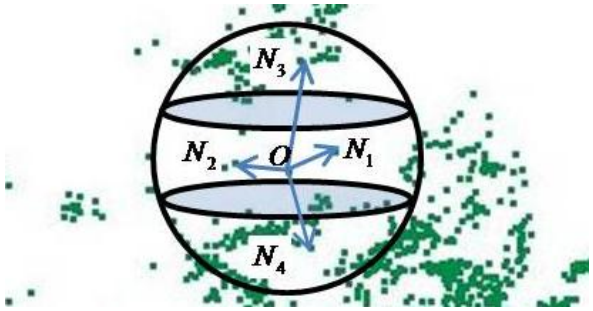
Ours **Logistic regression**

Conclusion

- Simple and fast approach for scene labeling
  - No graphical model
  - Labeling via 5x logistic regression predictions



- Support flexible contextual features
  - Learning rich relationships



Thank you! And Questions?

- Acknowledgements
  - US Army Research Laboratory, Collaborative Technology Alliance
  - QinetiQ North America Robotics Fellowship