Stacked Hierarchical Labeling

Dan Munoz Drew Bagnell Martial Hebert

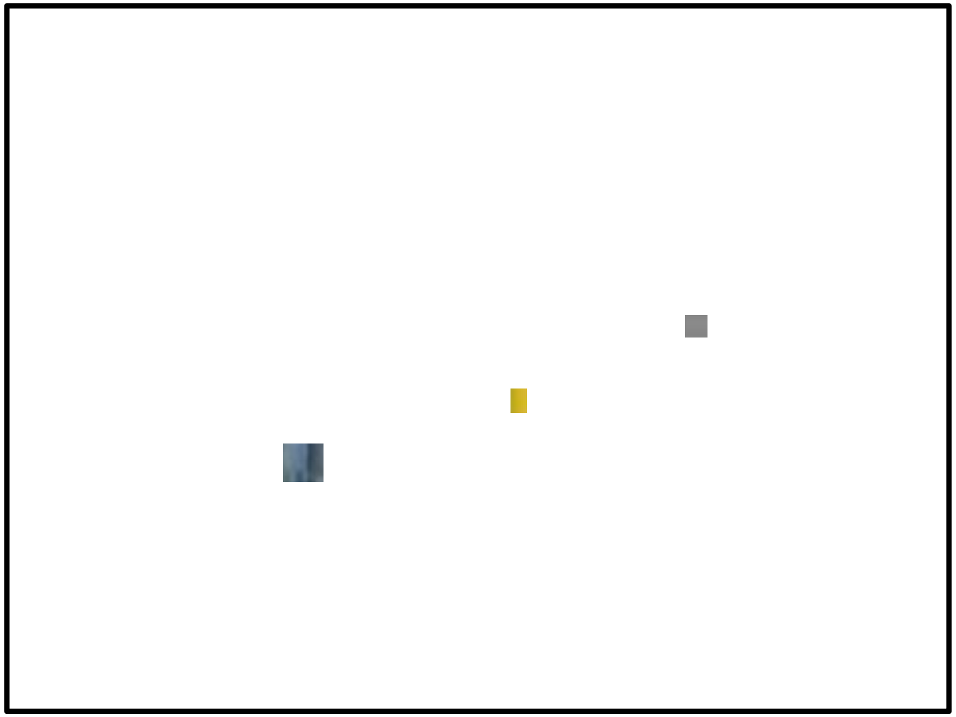
Carnegie Mellon
THE ROBOTICS INSTITUTE

The Labeling Problem

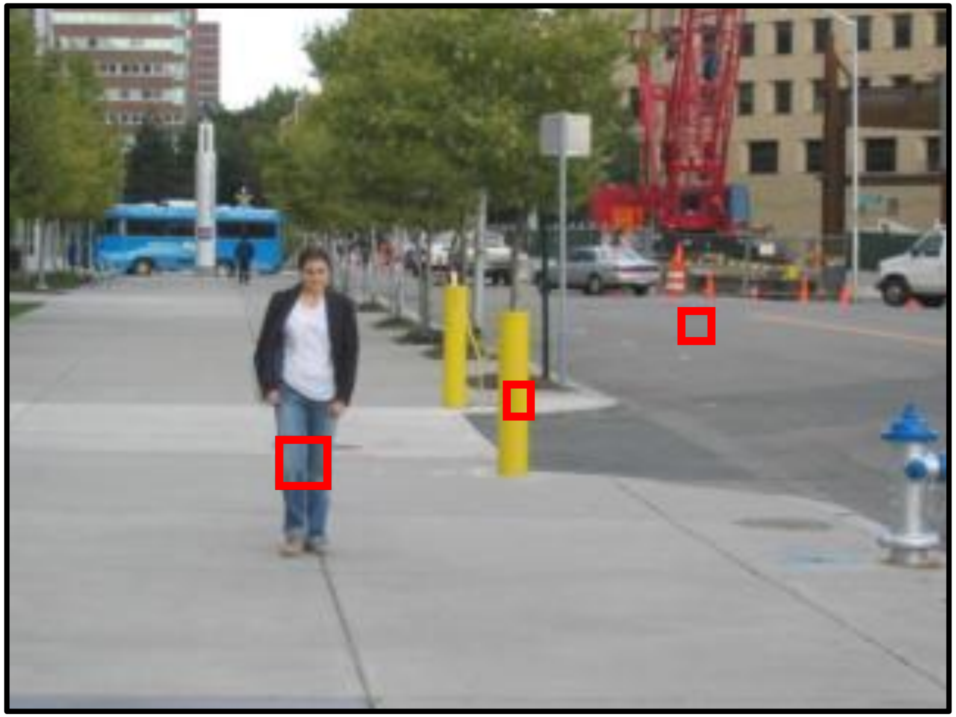
Input

Our Predicted Labels

The Labeling Problem



The Labeling Problem



- Needed: **better representation** & **interactions**
  - *Ohta '78*

Using Regions

Input

Ideal Regions

Slide from T. Malisiewicz

Using Regions

Input **Actual Regions**

Slide from T. Malisiewicz

Using Regions + Interactions

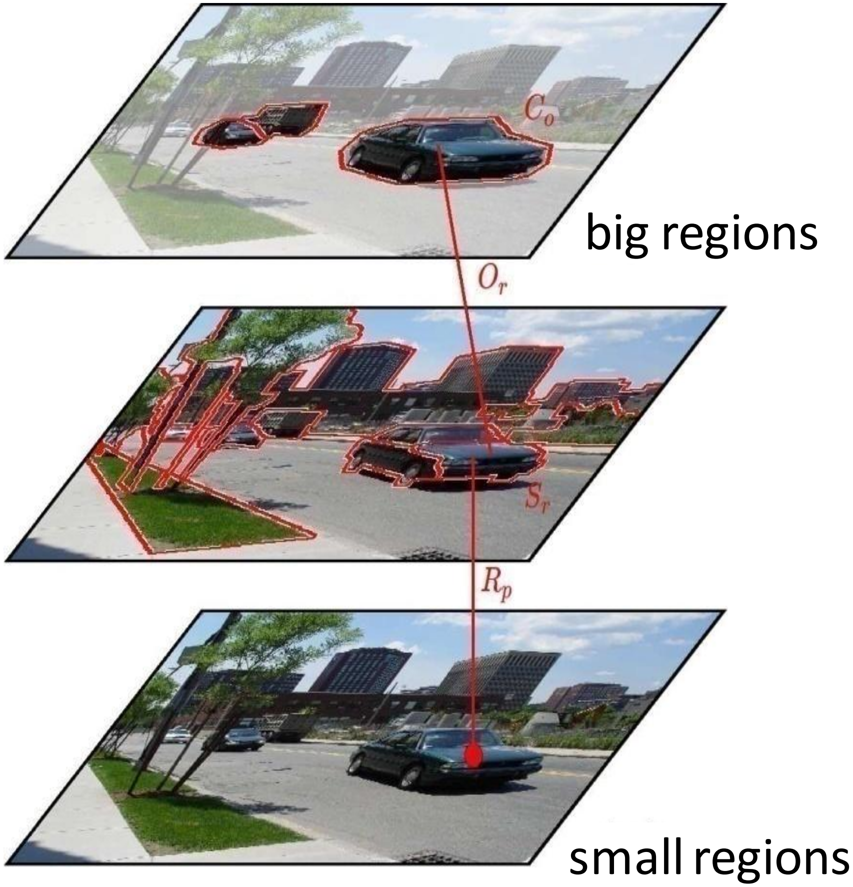
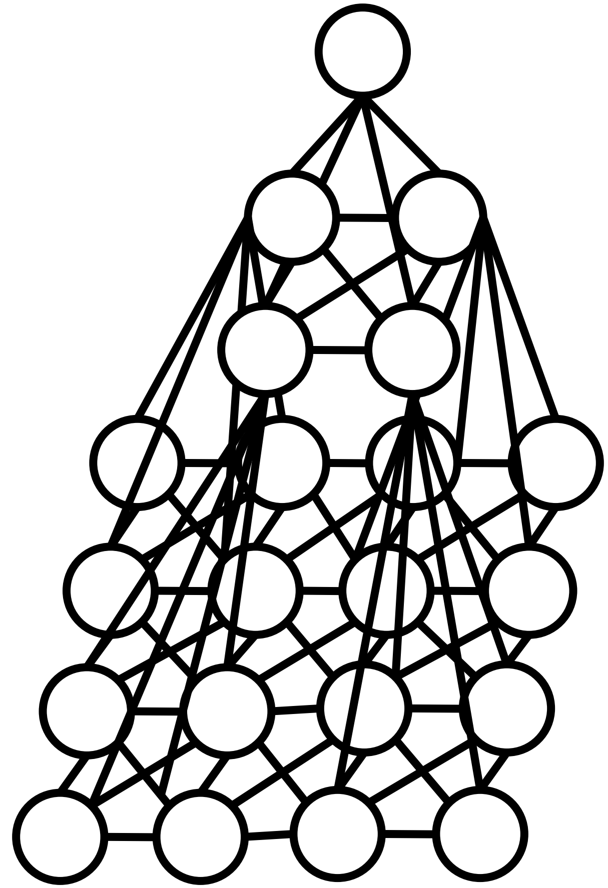


Image Representation



Ideal Prob. Graphical Model

- High-order
- Expressive interactions

Using Regions + Interactions

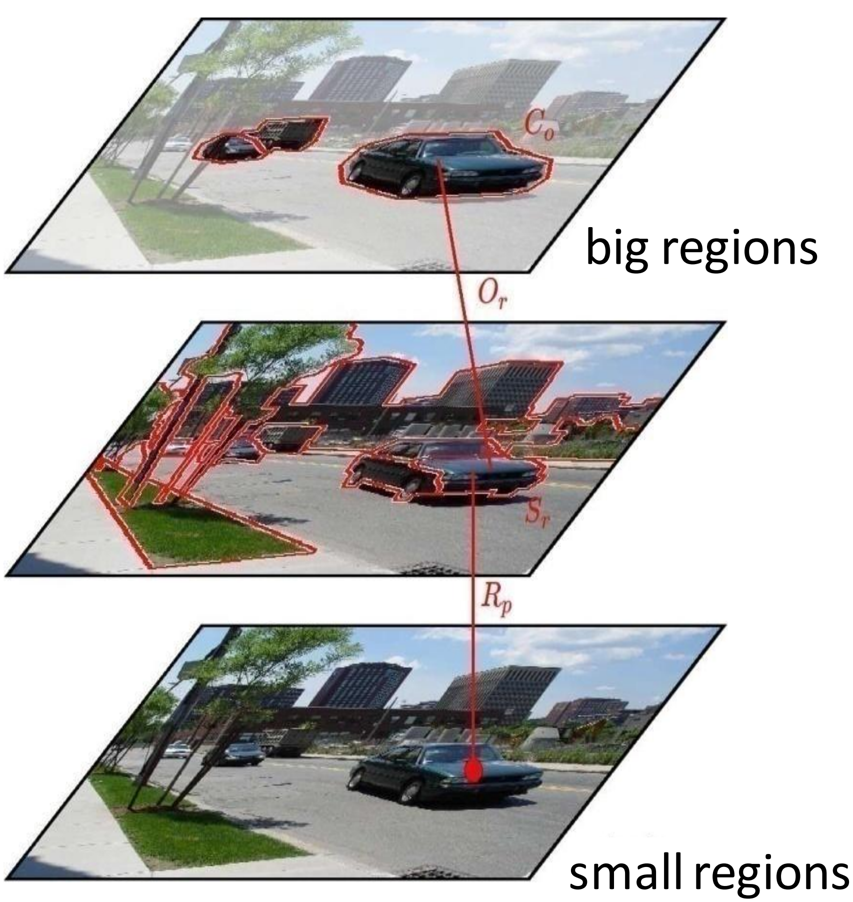
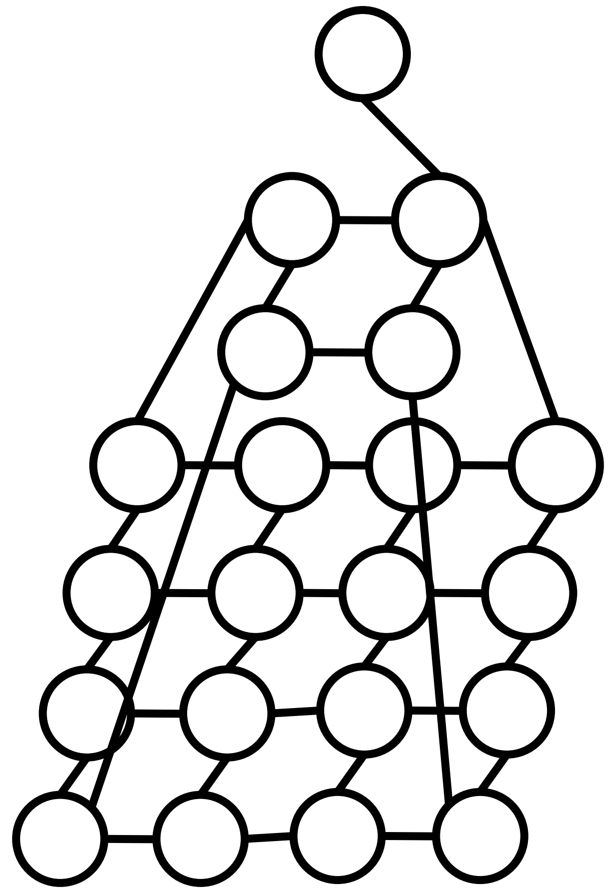


Image Representation

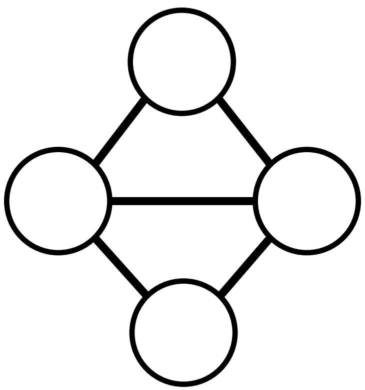


Actual PGM

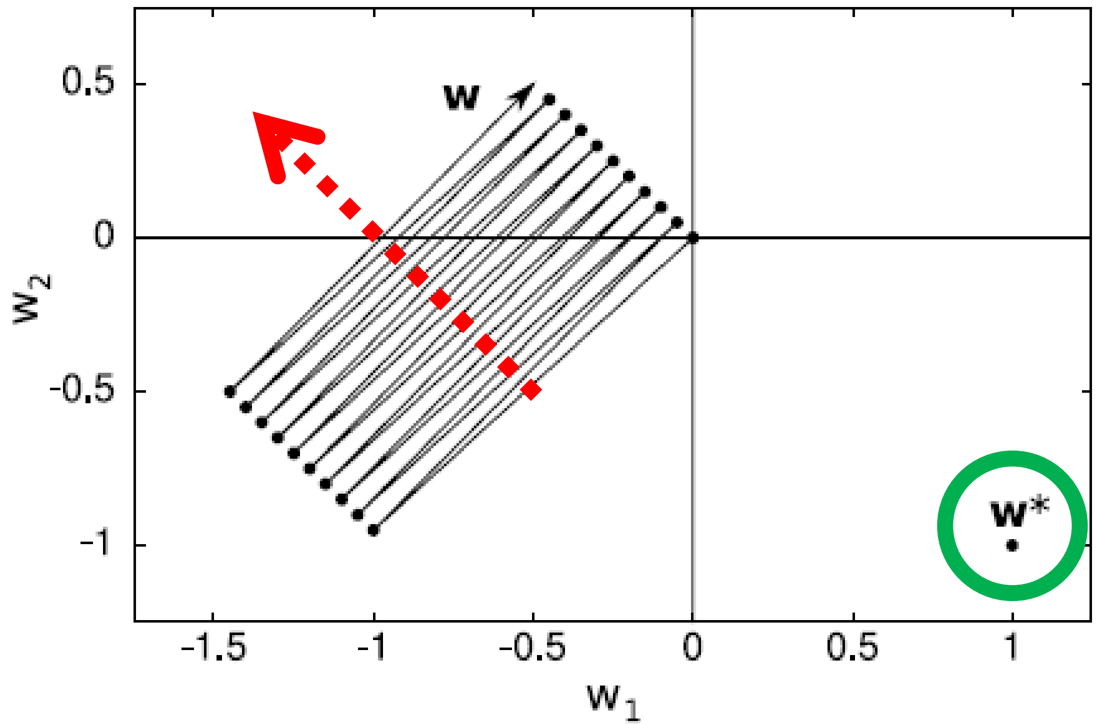
- Restrictive interactions
- <u>Still</u> NP-hard

Learning with Approximate Inference

- PGM learning requires **exact** inference
  - Otherwise, may **diverge** *Kulesza and Pereira '08*

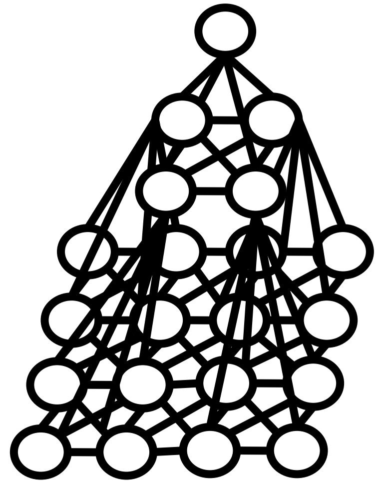


Simple Random Field



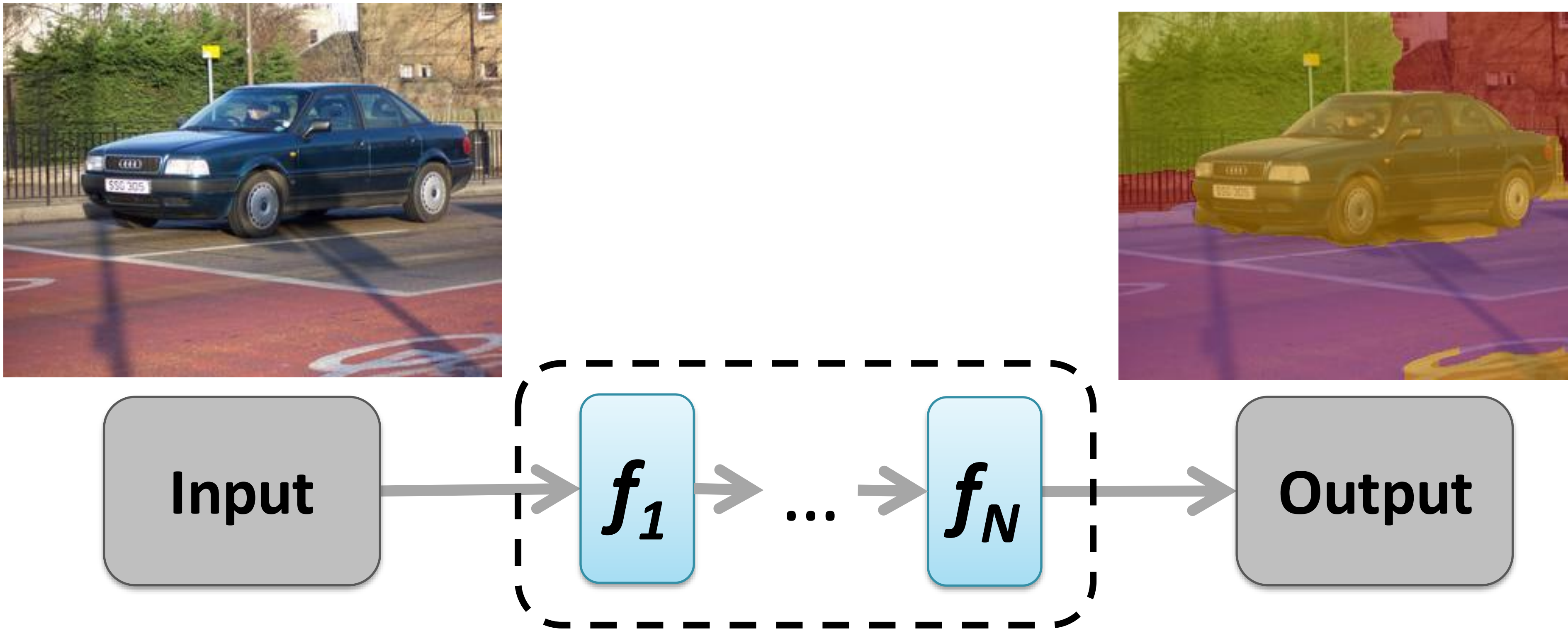
Learning Path

PGM Approach



Input → PGM Inference → Output

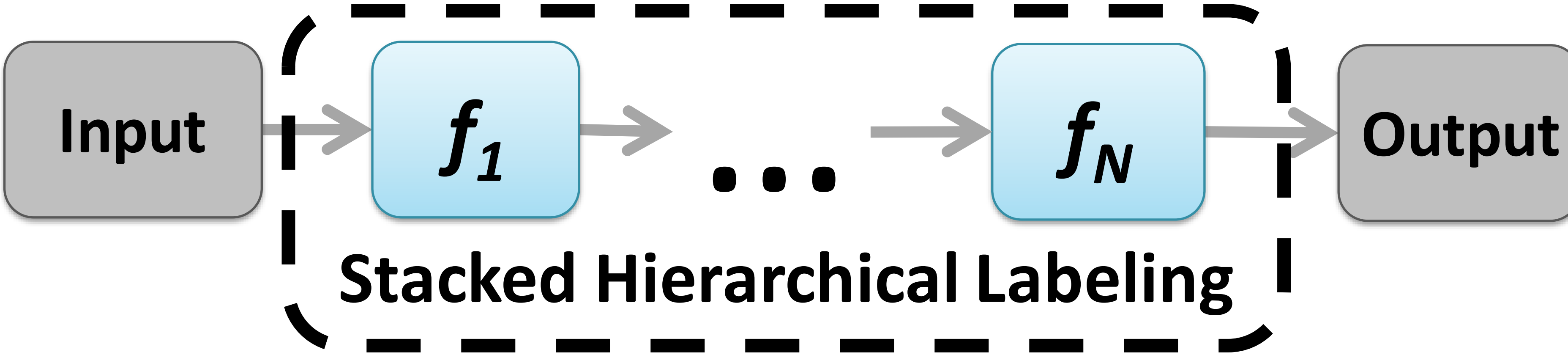
Our Approach



Sequence of simple problems

Cohen ’05, Daume III ’06

A Sequence of Simple Problems



- Training simple modules to net desired output
  - No searching in exponential space
- Not optimizing any joint distribution/energy
  - Not necessarily doing it before! *Kulesza & Pereira '08*

Our Contribution

- **An effective PGM alternative for labeling**
  - Training a **hierarchical** procedure of simple problems
- Naturally analyzes multiple scales
  - Robust to imperfect segmentations
- Enables more expressive interactions
  - Beyond pair-wise smoothing

Related Work

- Learning with multi-scale configurations
  - Joint probability distribution

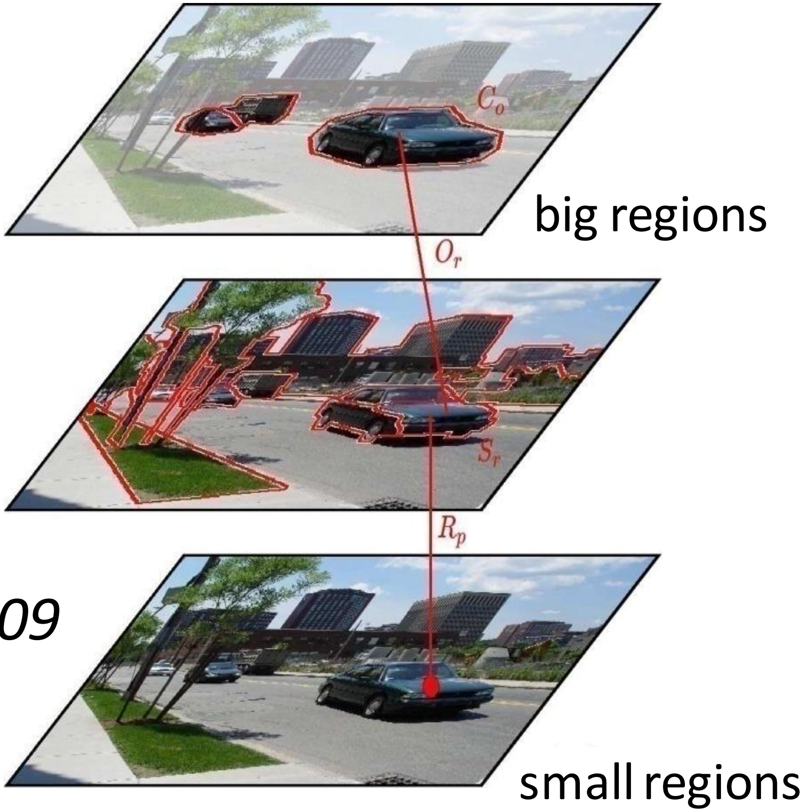
    Bouman '94, Feng '02, He '04

    Borenstein '04, Kumar '05

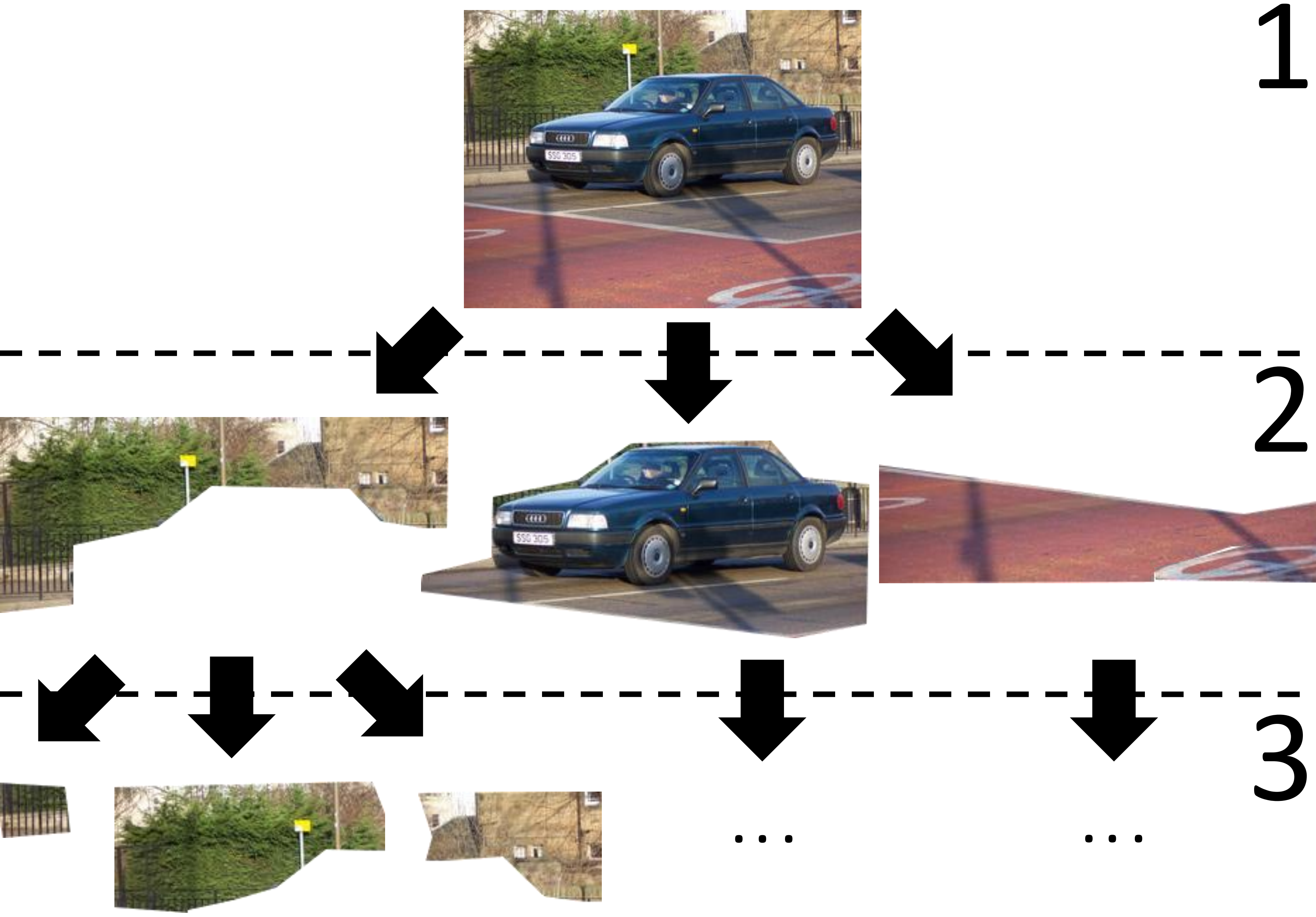
  - Joint score/energy

    Tu '03, S.C. Zhu '06, L. Zhu '08

    Munoz '09, Gould '09, Ladicky '09



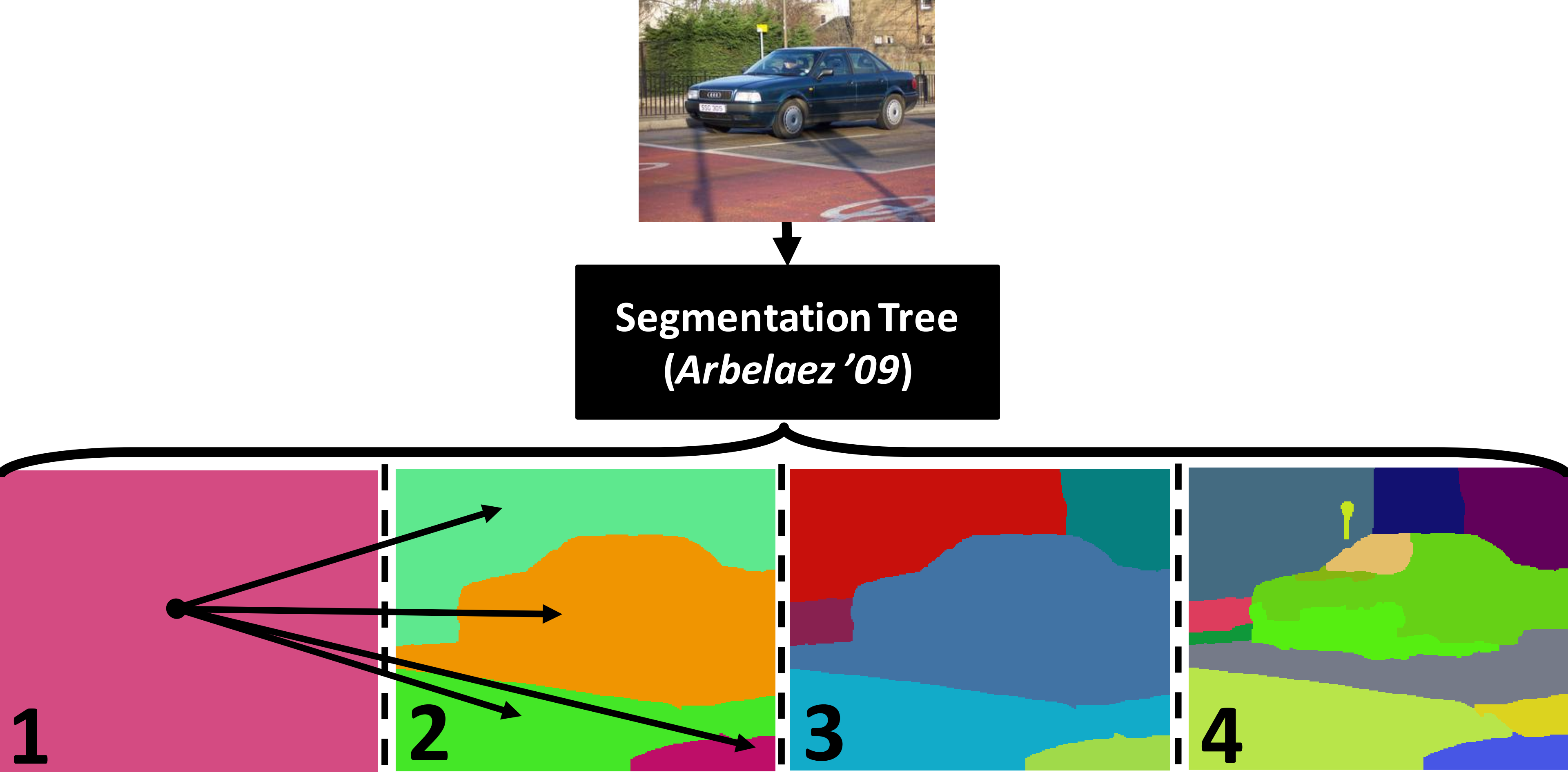
- Mitigating the intractable joint optimization
  - *Cohen '05, Daume III '06, Kou '07,* ***Tu '08****, Ross '10*

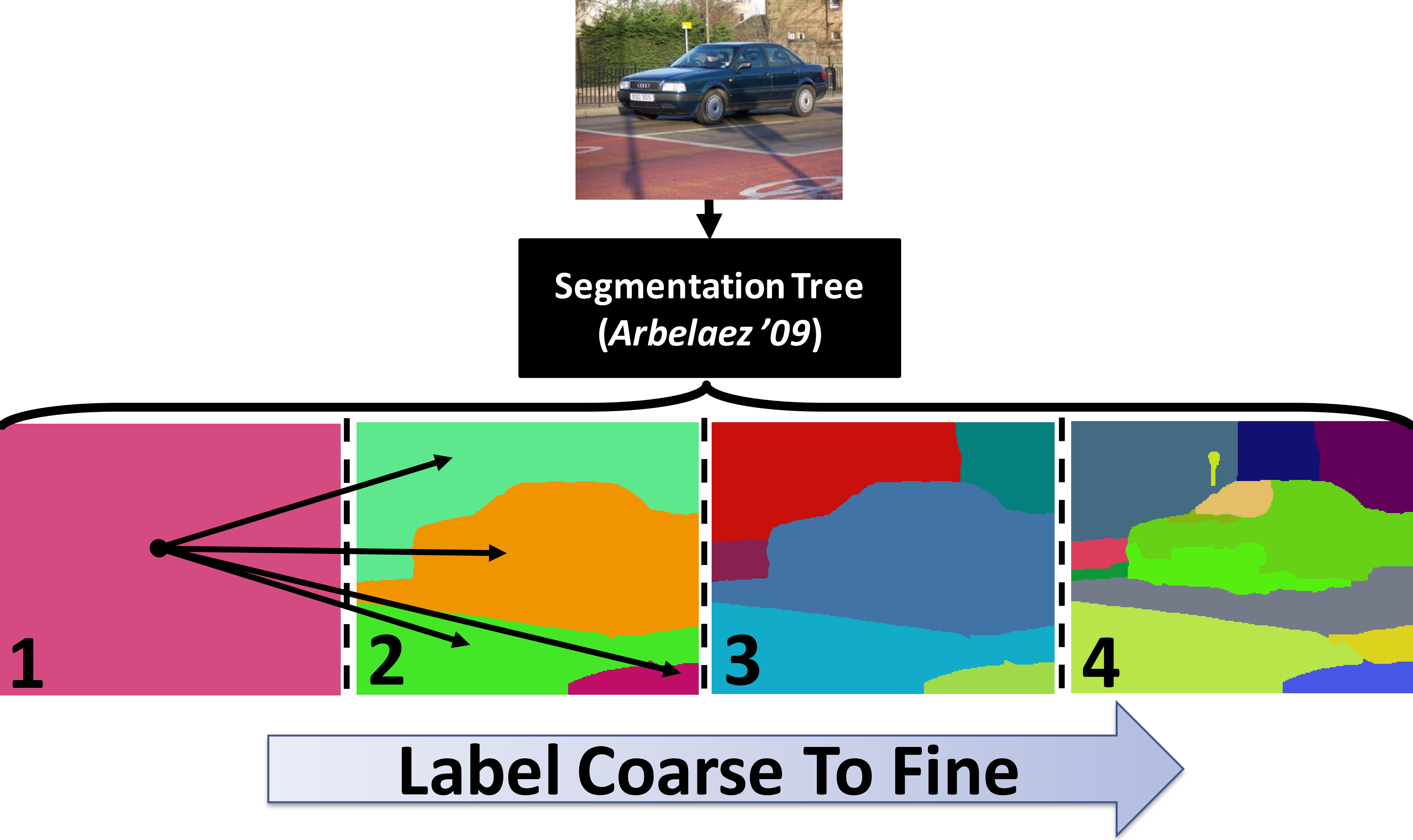

1
2
3
...
...

1

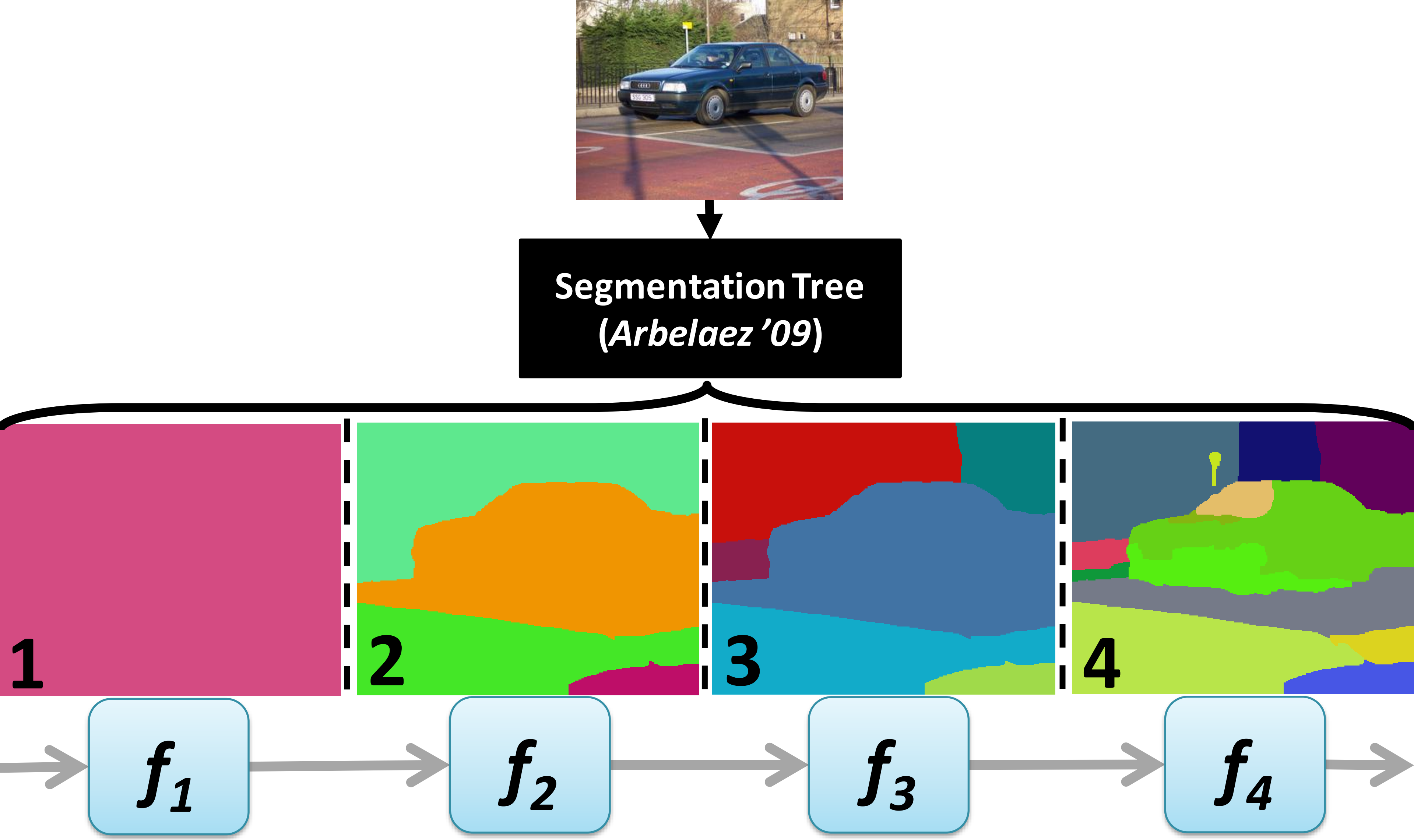
In this work, the **segmentation** tree is **given**

We use the technique from *Arbelaez '09*


Segmentation Tree
(*Arbelaez '09*)
1
2
3
4



- Parent sees big picture
- Naturally handles scales

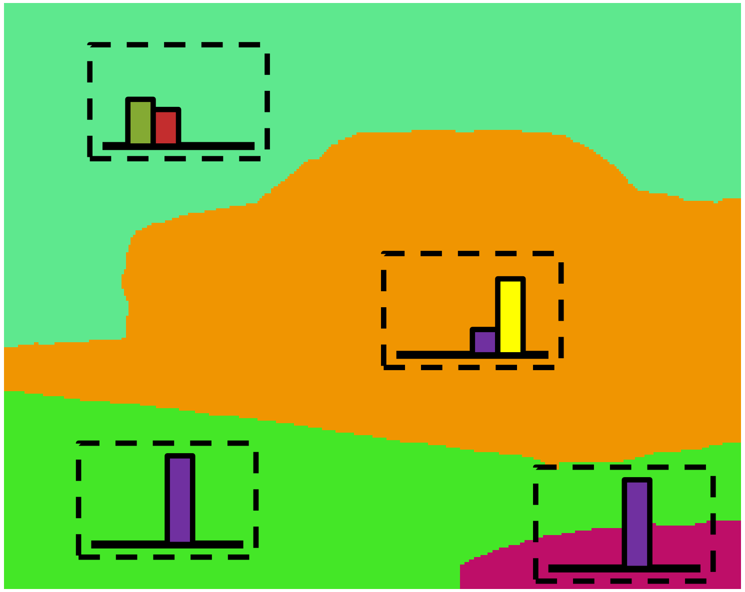


- Parent sees big picture
- Naturally handles scales
- Break into simple tasks
- Predict label **mixtures**

Handling Real Segmentation

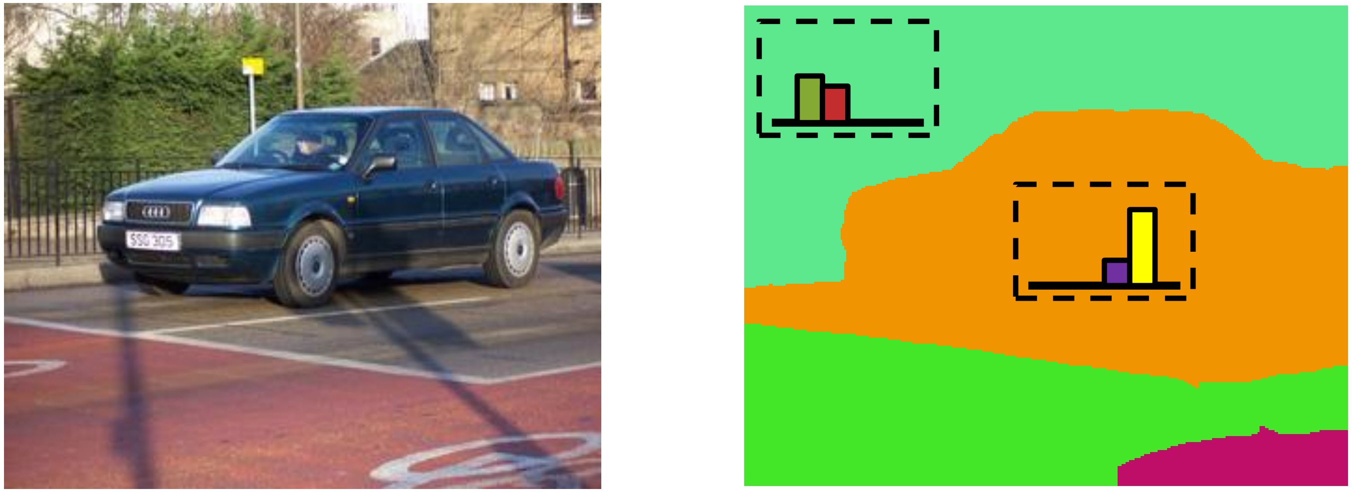
- f_i predicts **mixture** of labels for each region

Input



Segmentation Map

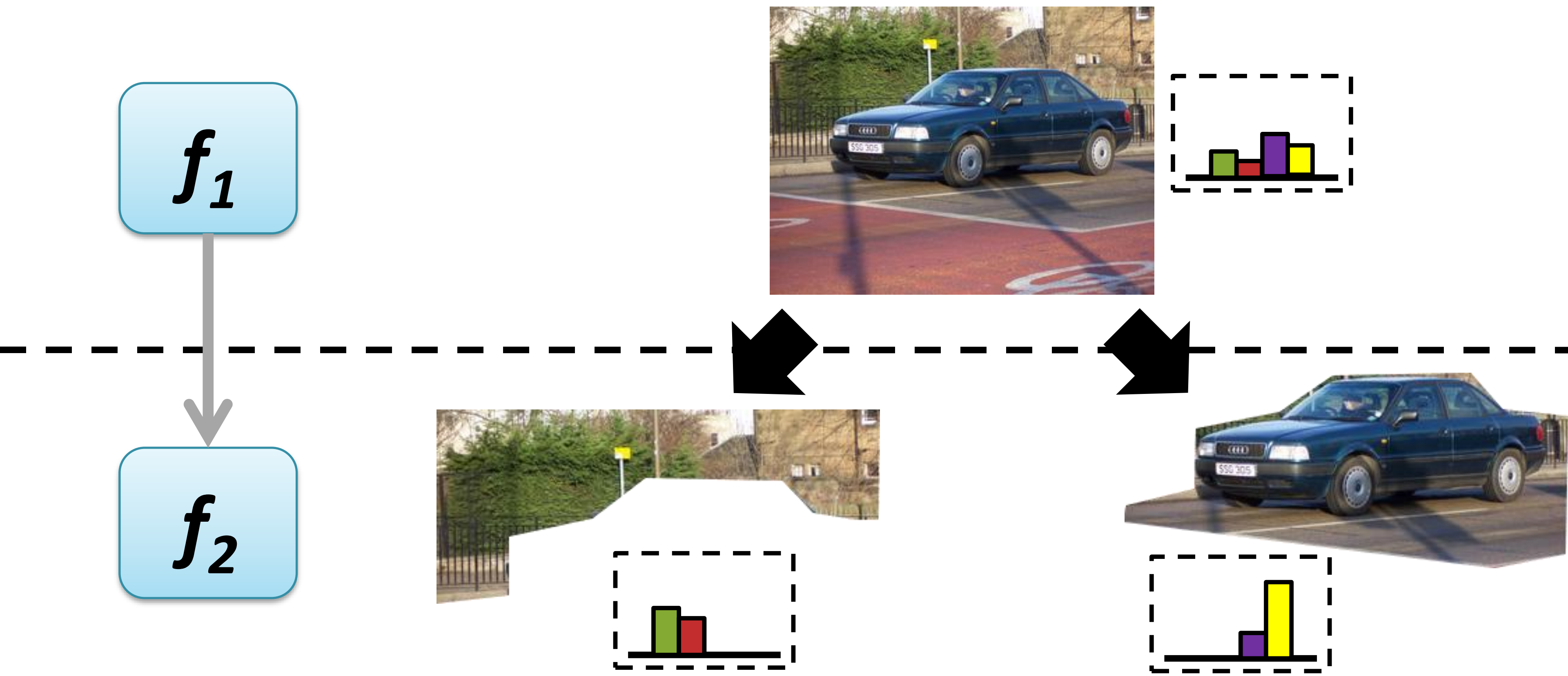
Actual Predicted Mixtures



(brighter → higher probability)

Training Overview

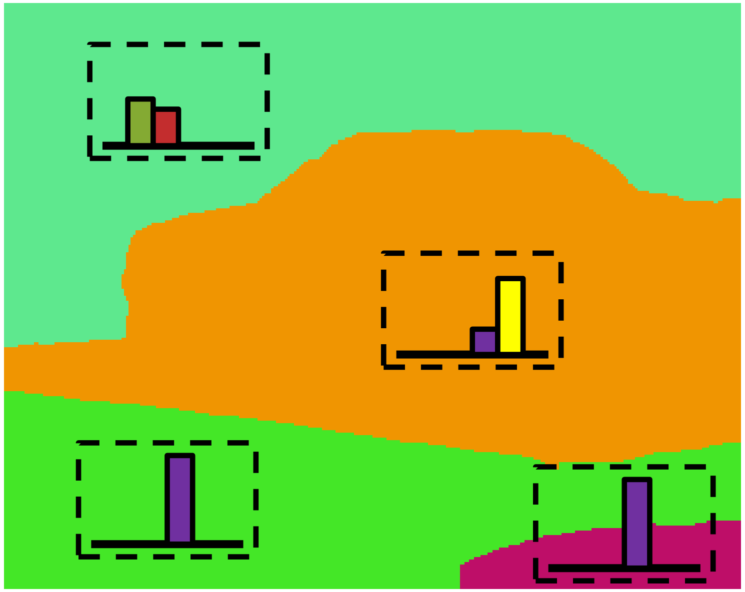
- How to train each module f_i ?
- How to use previous predictions?
- How to train the hierarchical sequence?



Training Overview

- **How to train each module f_i ?**
- How to use previous predictions?
- How to train the hierarchical sequence?

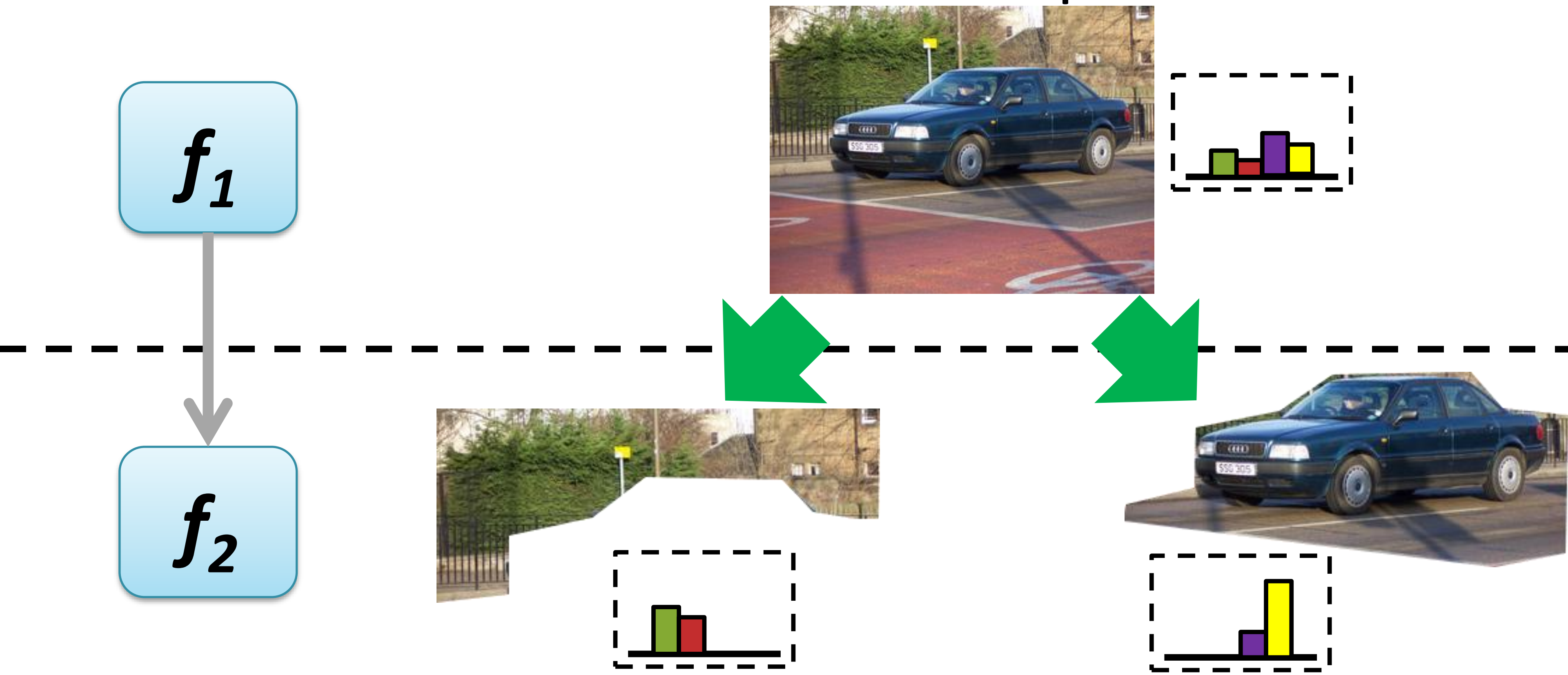
Modeling Heterogeneous Regions



- Count **true labels P_r** present in each region r
- Train a **model Q** to match each P_r
  - Logistic Regression
- $min_Q\ H(P,Q)$ → Weighted Logistic Regression
  - Image features: texture, color, etc. (*Gould '08*)

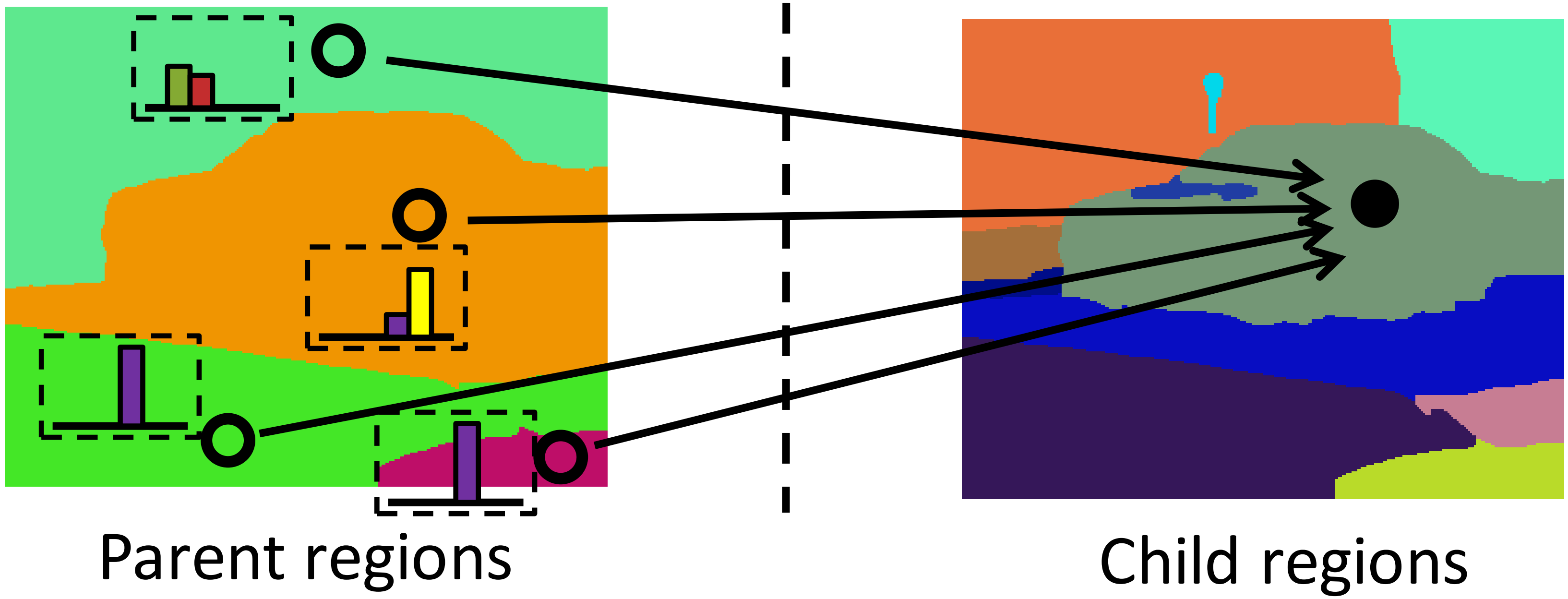
Training Overview

- How to train each module f_i ?
- **How to use previous predictions?**
- How to train the hierarchical sequence?



Using Parent Predictions

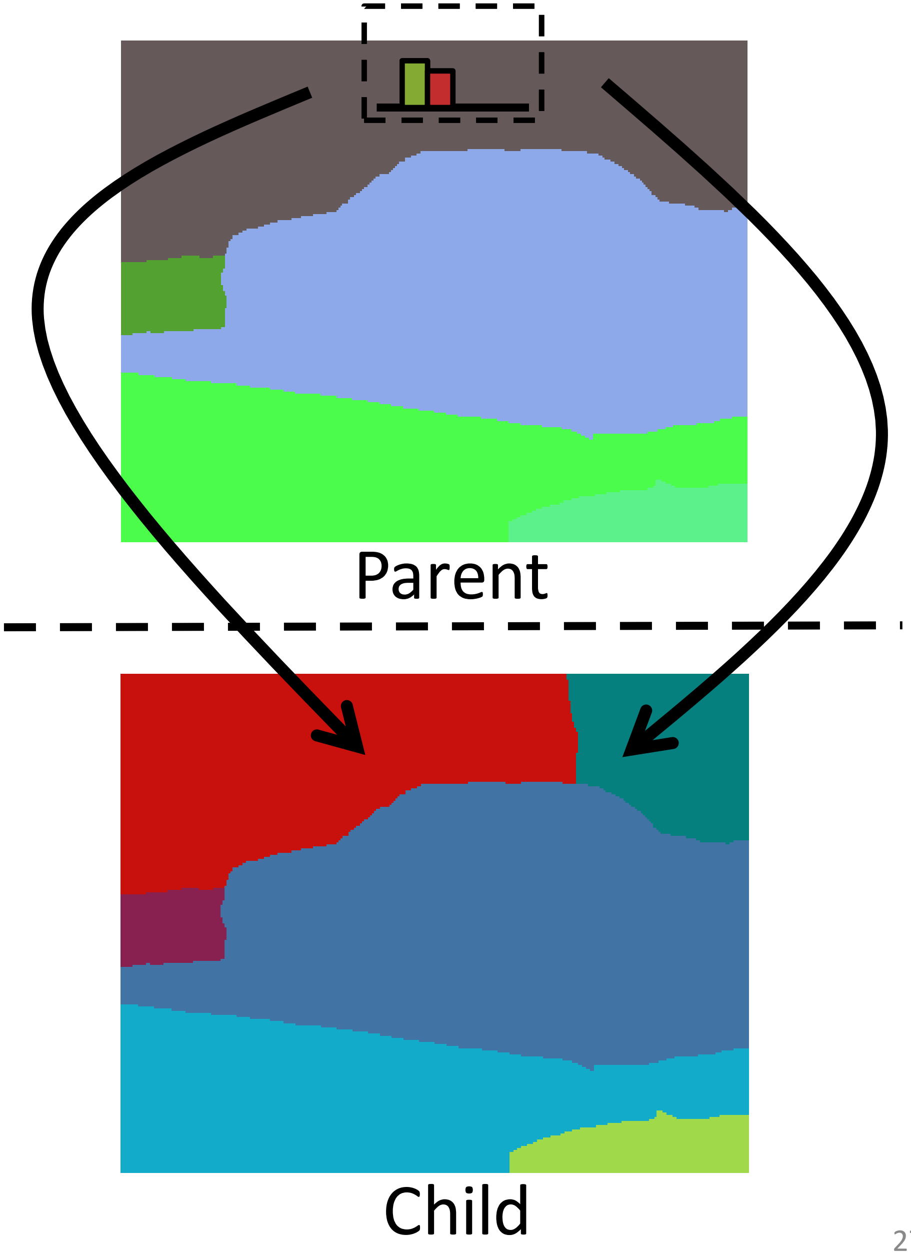
- Use broader context in the finer regions



  - Allow finer regions access to **all parent predictions**
- Create & **append** 3 types of context features
  - *Kumar '05, Sofman '06, Shotton '06, Tu '08*

Parent Context

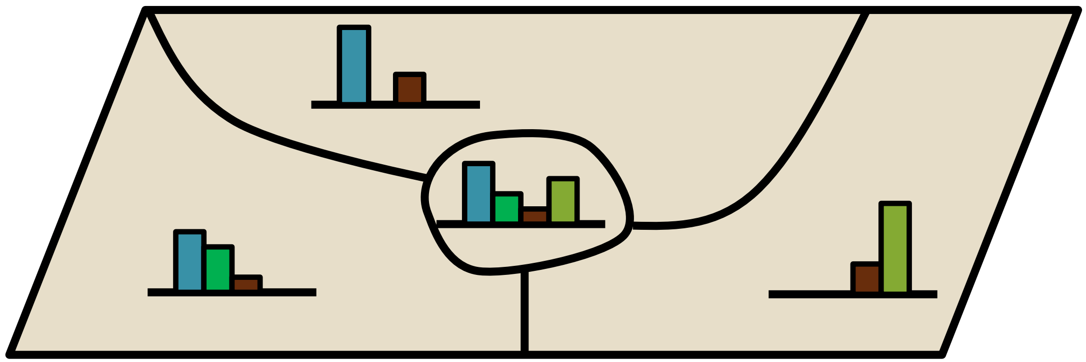
- Refining the parent



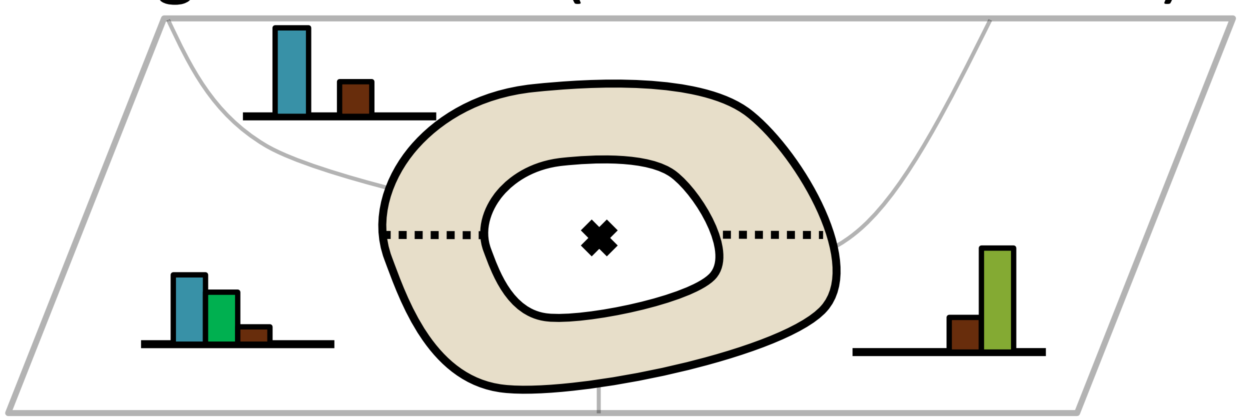
Detailed In Paper

- Image-wise (co-occurrence)

$$\sum_{\textit{regions}}$$

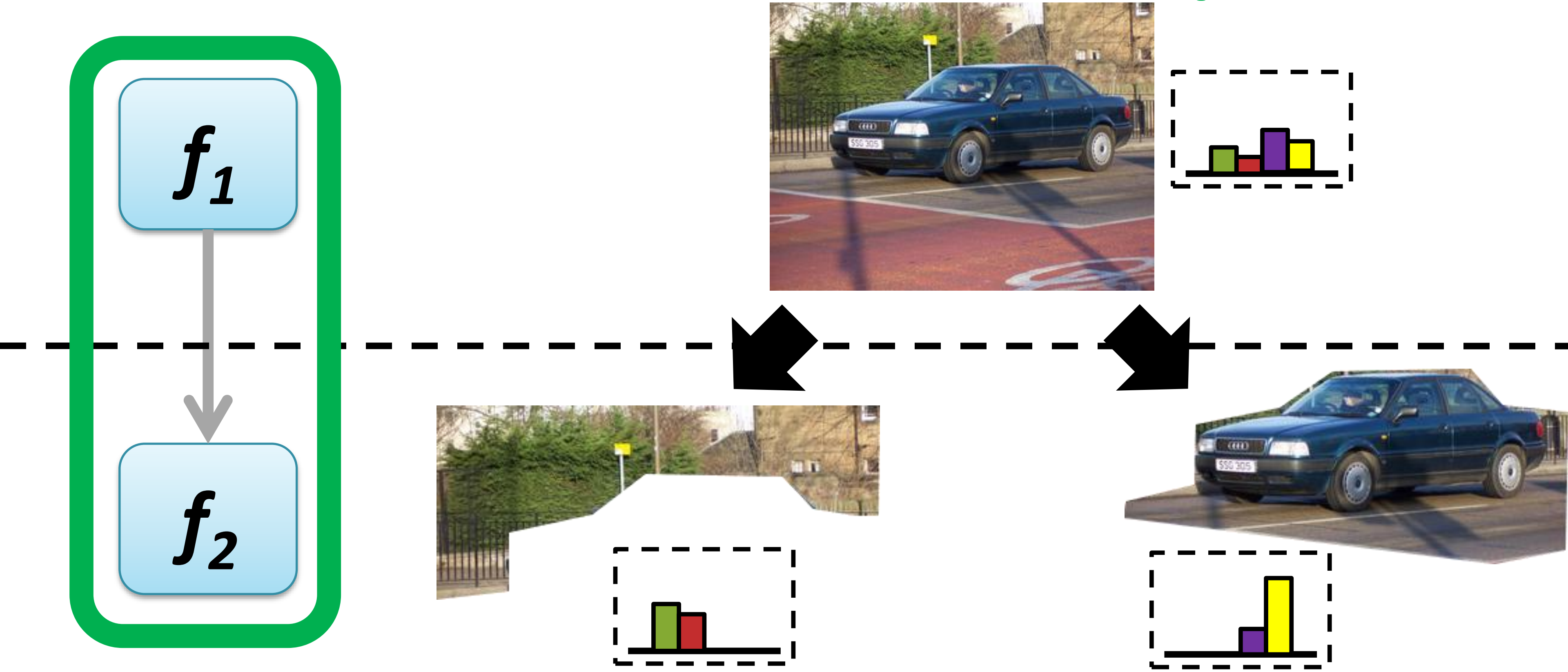


- Spatial Neighborhood (center-surround)



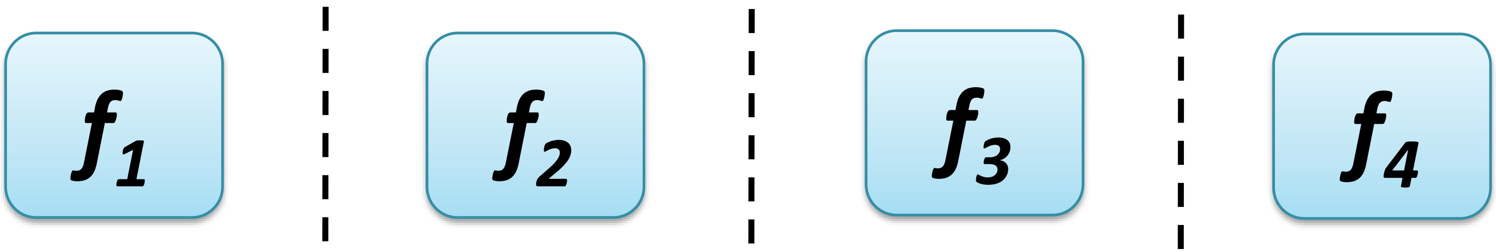
Training Overview

- How to train each module f_i ?
- How to use previous predictions?
- **How to train the hierarchical sequence?**



Approach #1

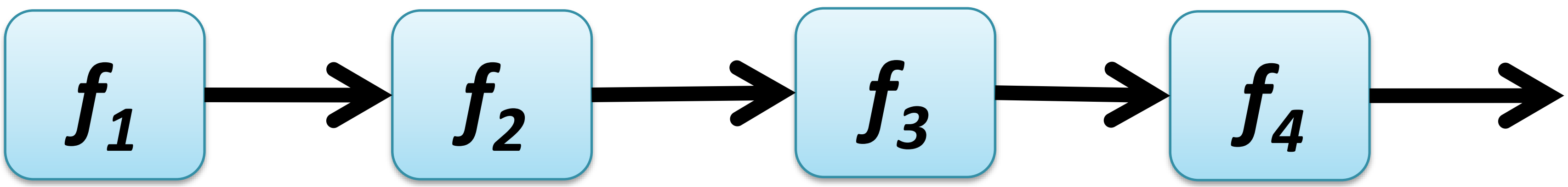
- Train each module independently
  - Use ground truth context features



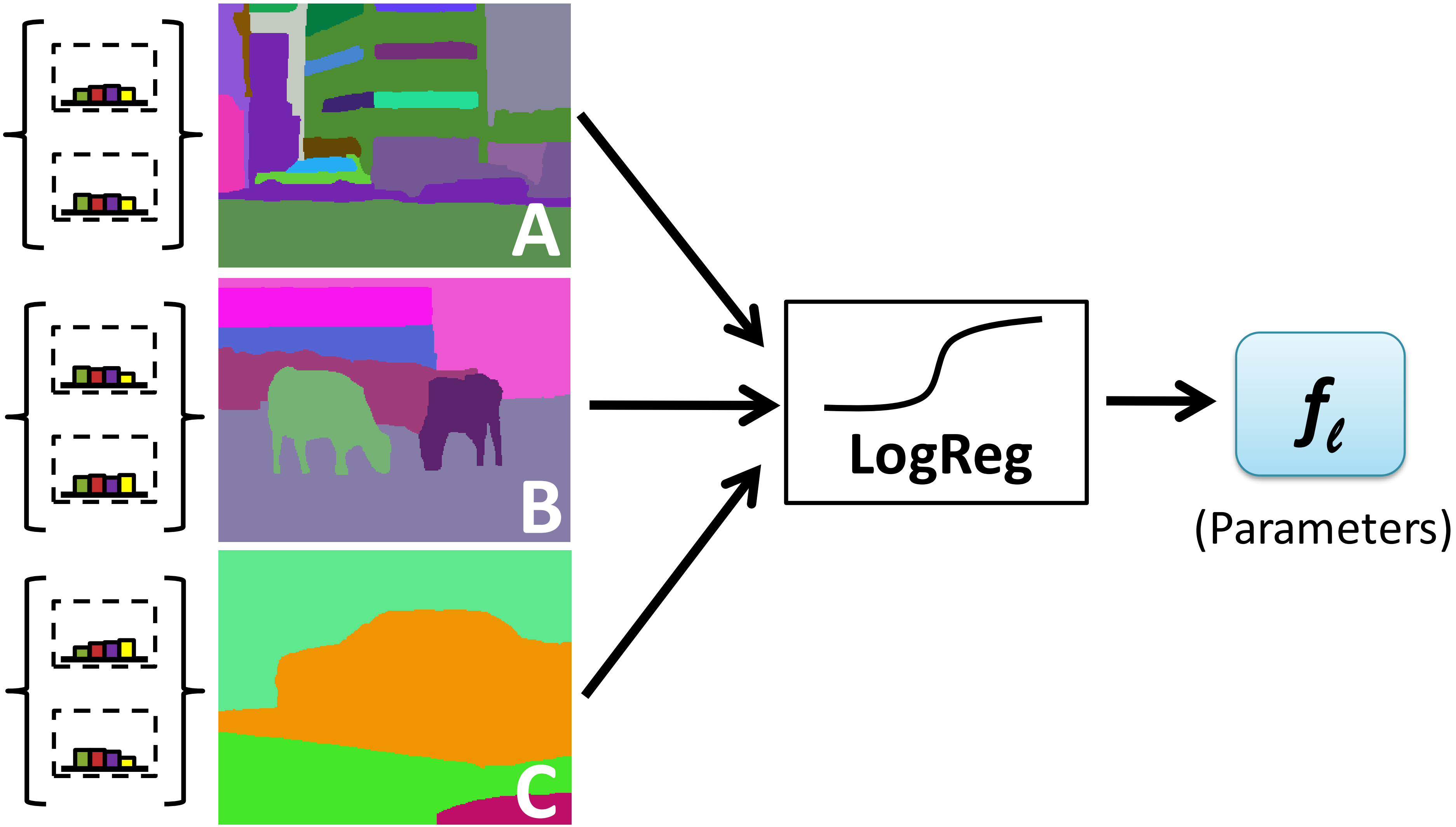
- **Problem: Cascades of Errors**
  - Modules depend on **perfect** context features
  - Observe no mistakes during training
    → Propagate mistakes during testing

Approach #2

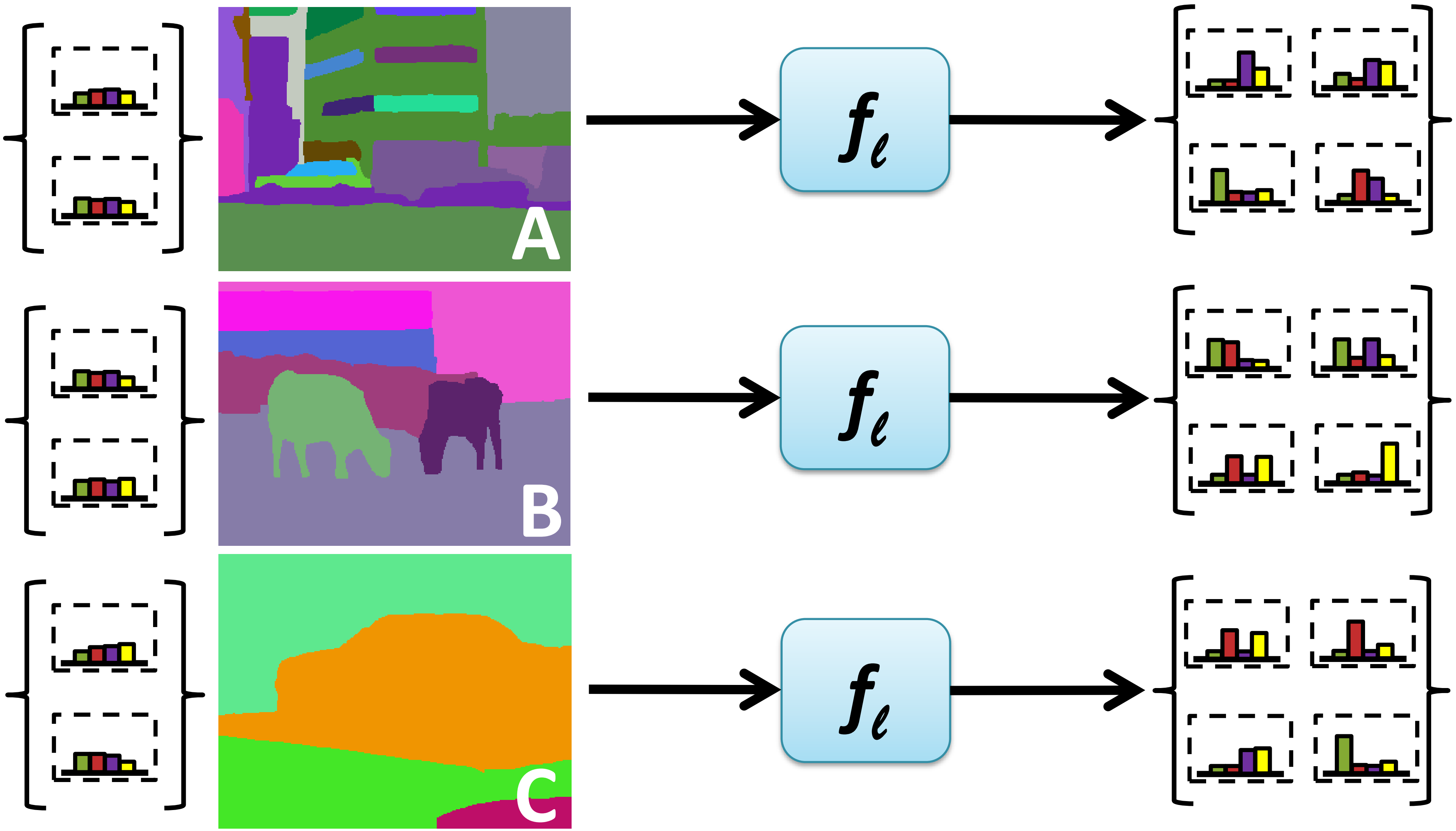
- **Solution: Train in feed-forward manner**
  - *Viola-Jones '01, Kumar '05, Wainwright '06, Ross '10*



Training Feed-Forward

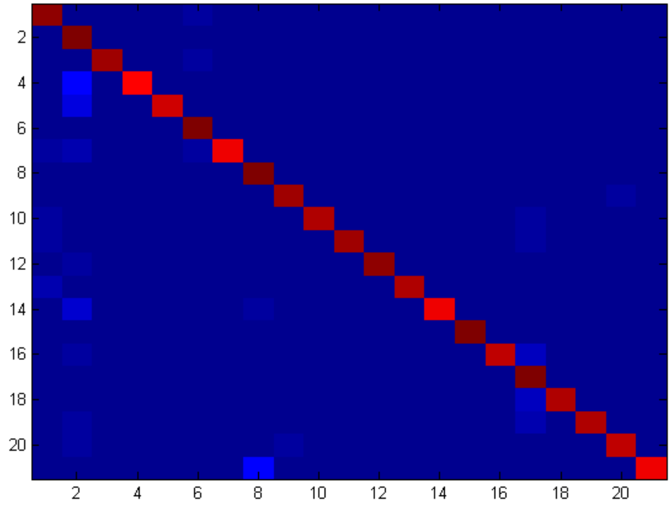


Training Feed-Forward

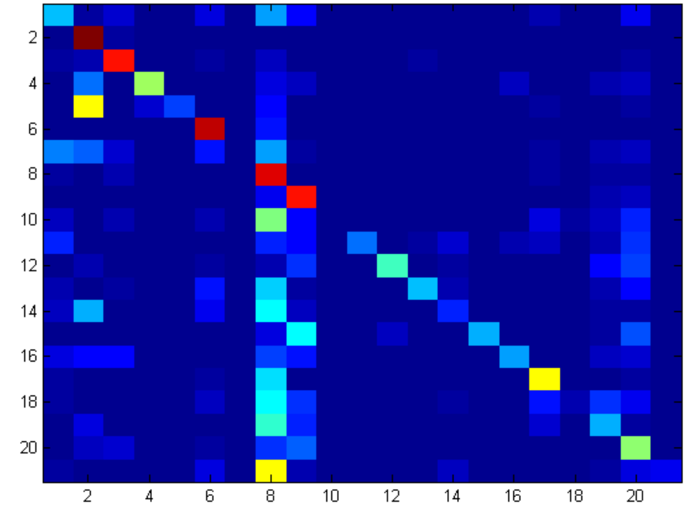


Cascades of Overfitting

F.F. Train Confusions

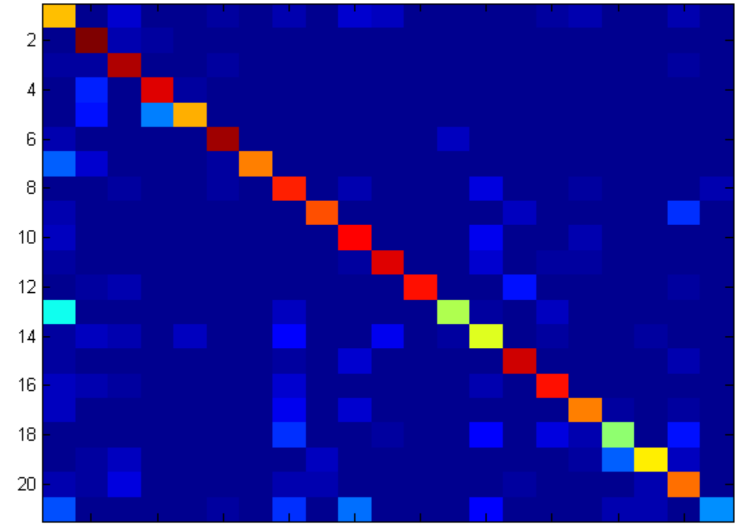


F.F. Test Confusions

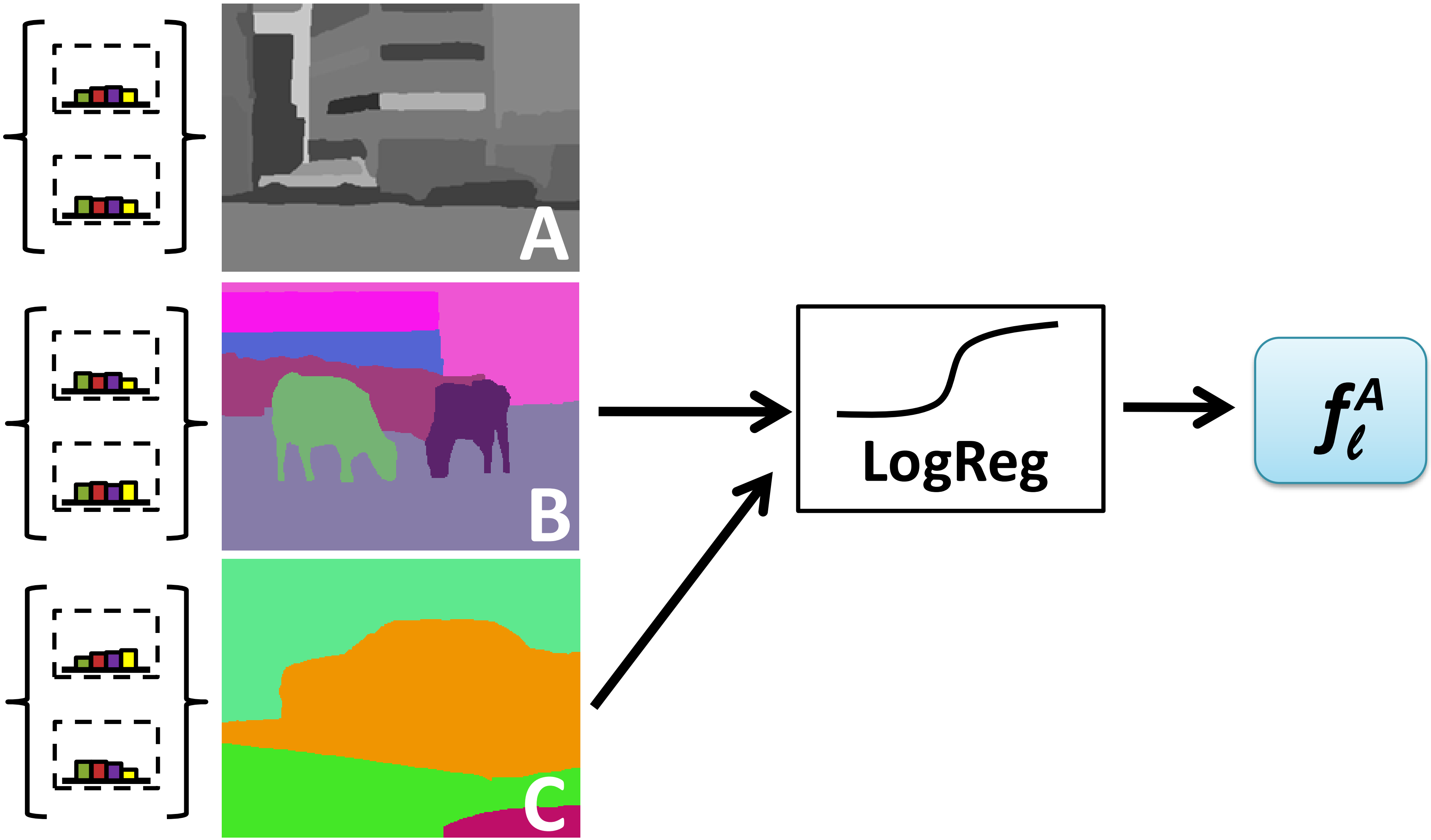


- **Solution: Stacking**
  - ***Wolpert '92***, *Cohen '05*
  - Similar to x-validation
  - Don't predict on data used for training

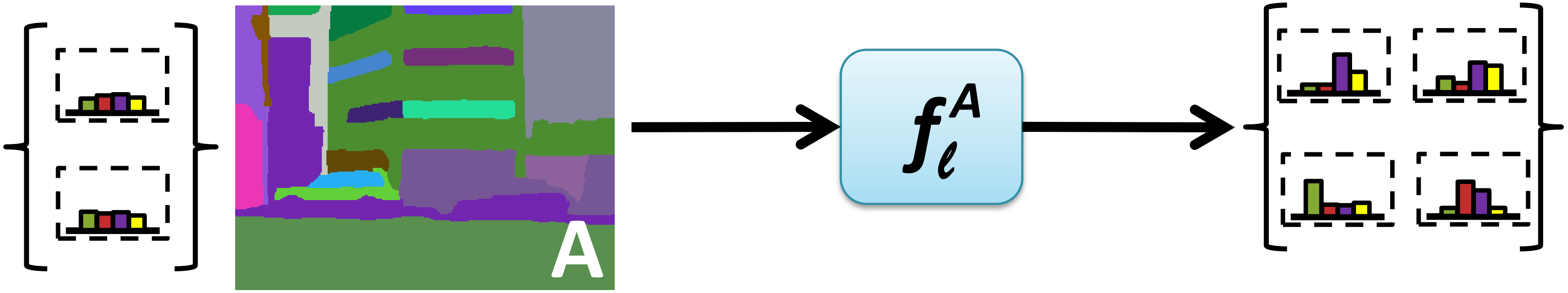
Stacking Test Confusions



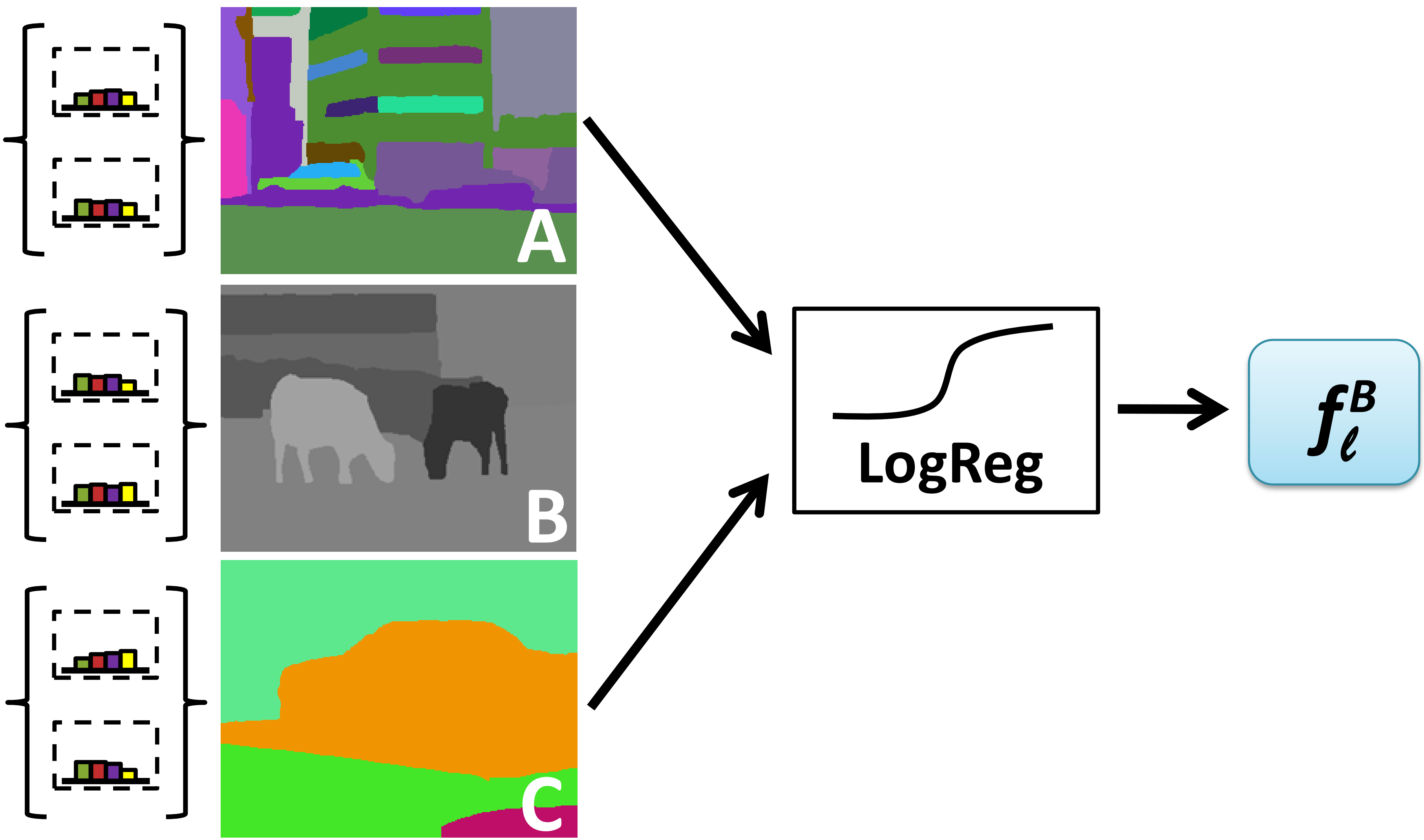
Stacking



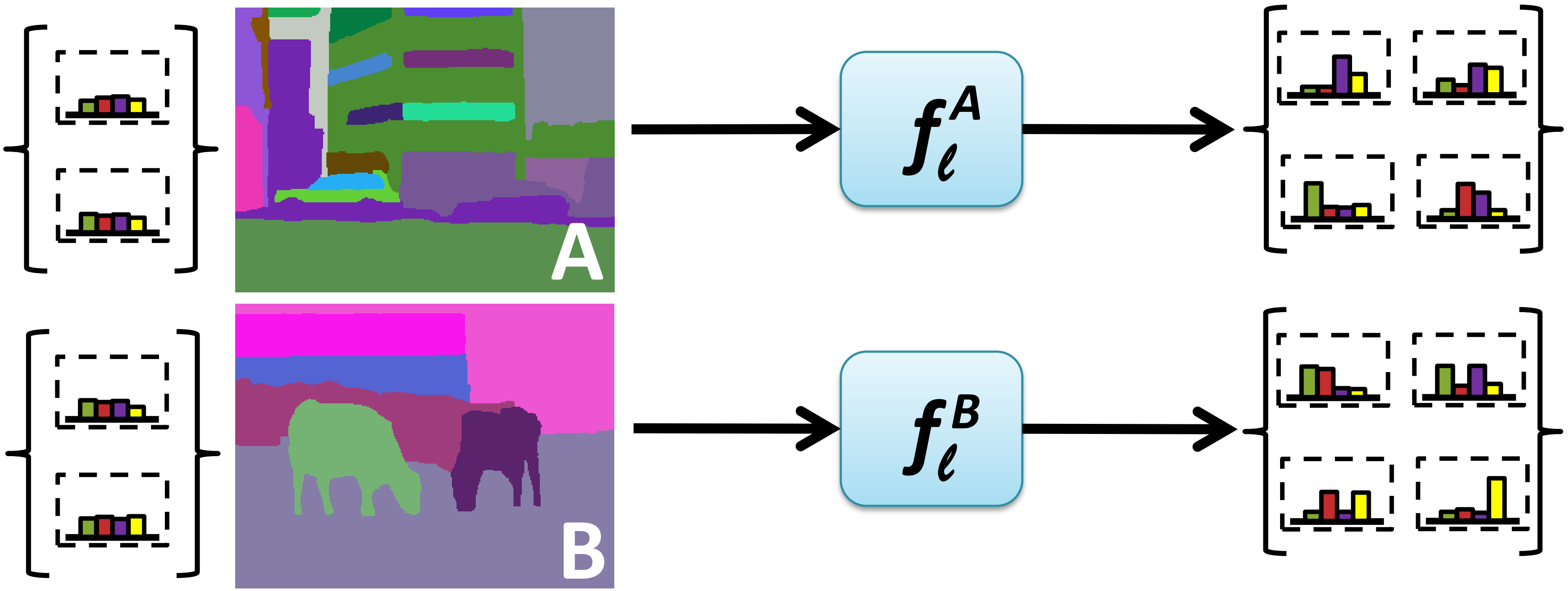
Stacking



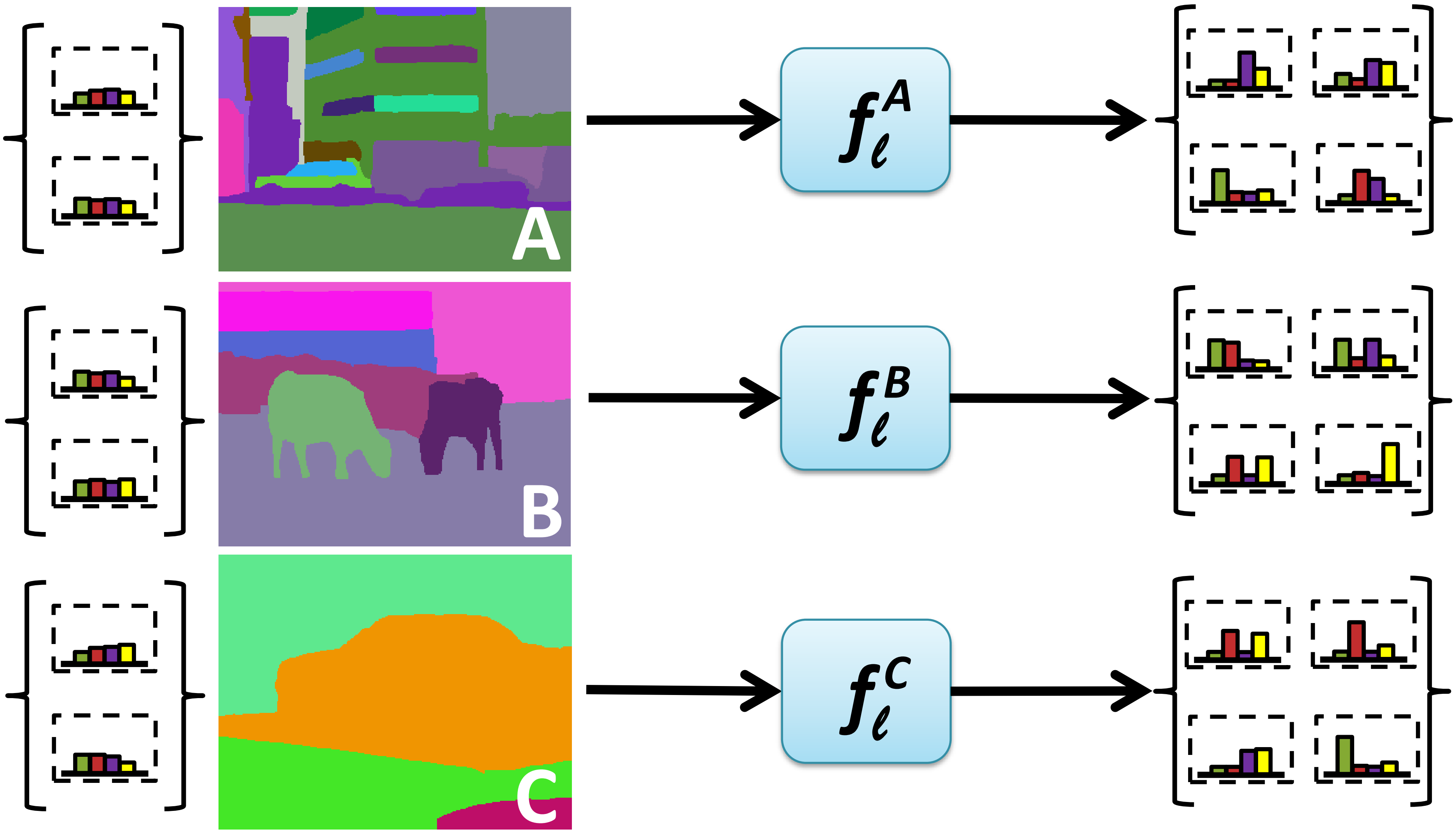
Stacking



Stacking



Stacking



Learning to Fix Mistakes

Person part of incorrect segment

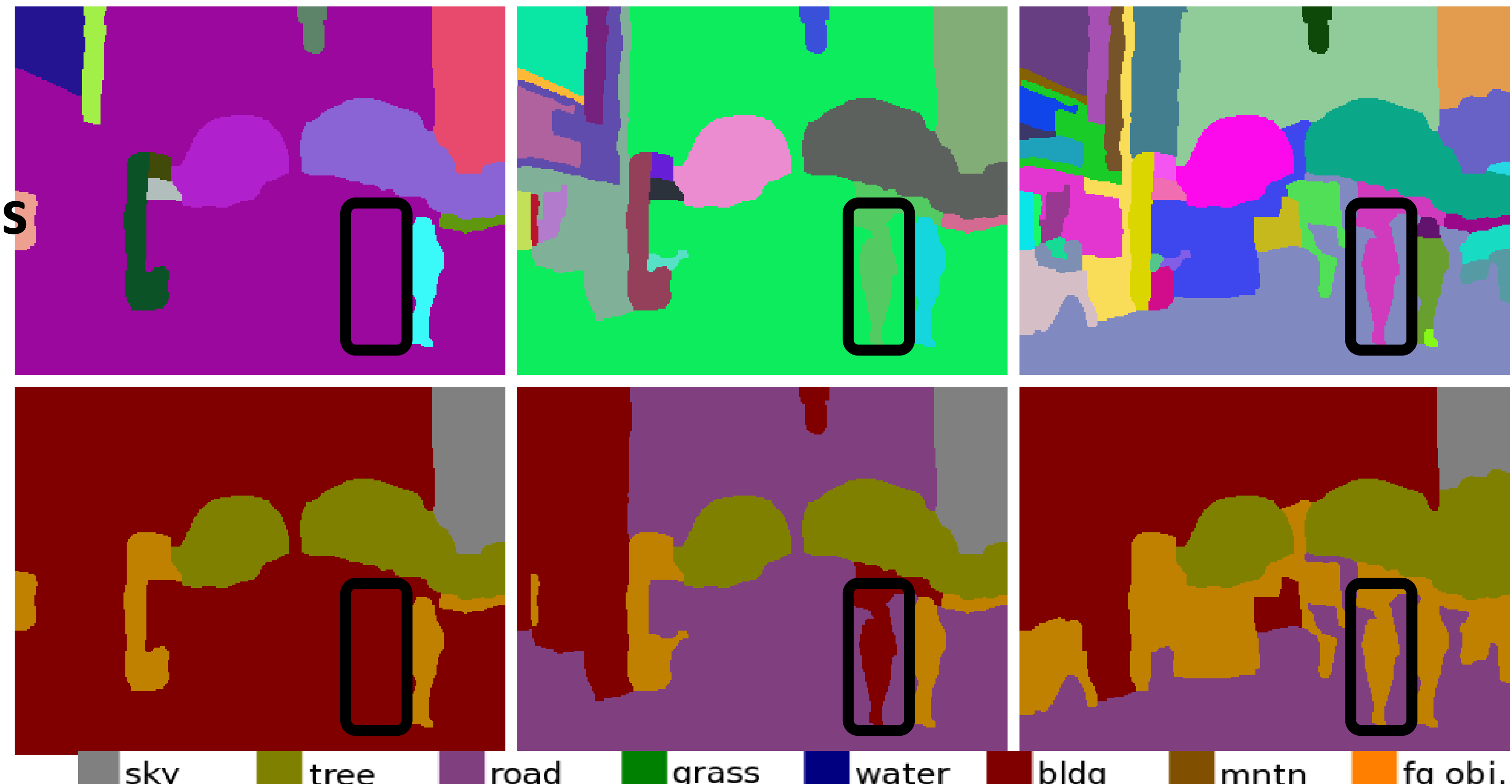
Person segmented, but relies on parent

Person fixes previous mistake

Level 5 | Level 6 | Level 7

Segments

Current Output



Level 1/8 Predictions

Segmentation

Level 1/8 Predictions

Segmentation

P(Foreground)

15%

P(Tree)

18%

P(Building)

12%

P(Road)

31%

Level 1/8 Predictions

Current Output

Segmentation

P(Foreground)

15%

P(Tree)

18%

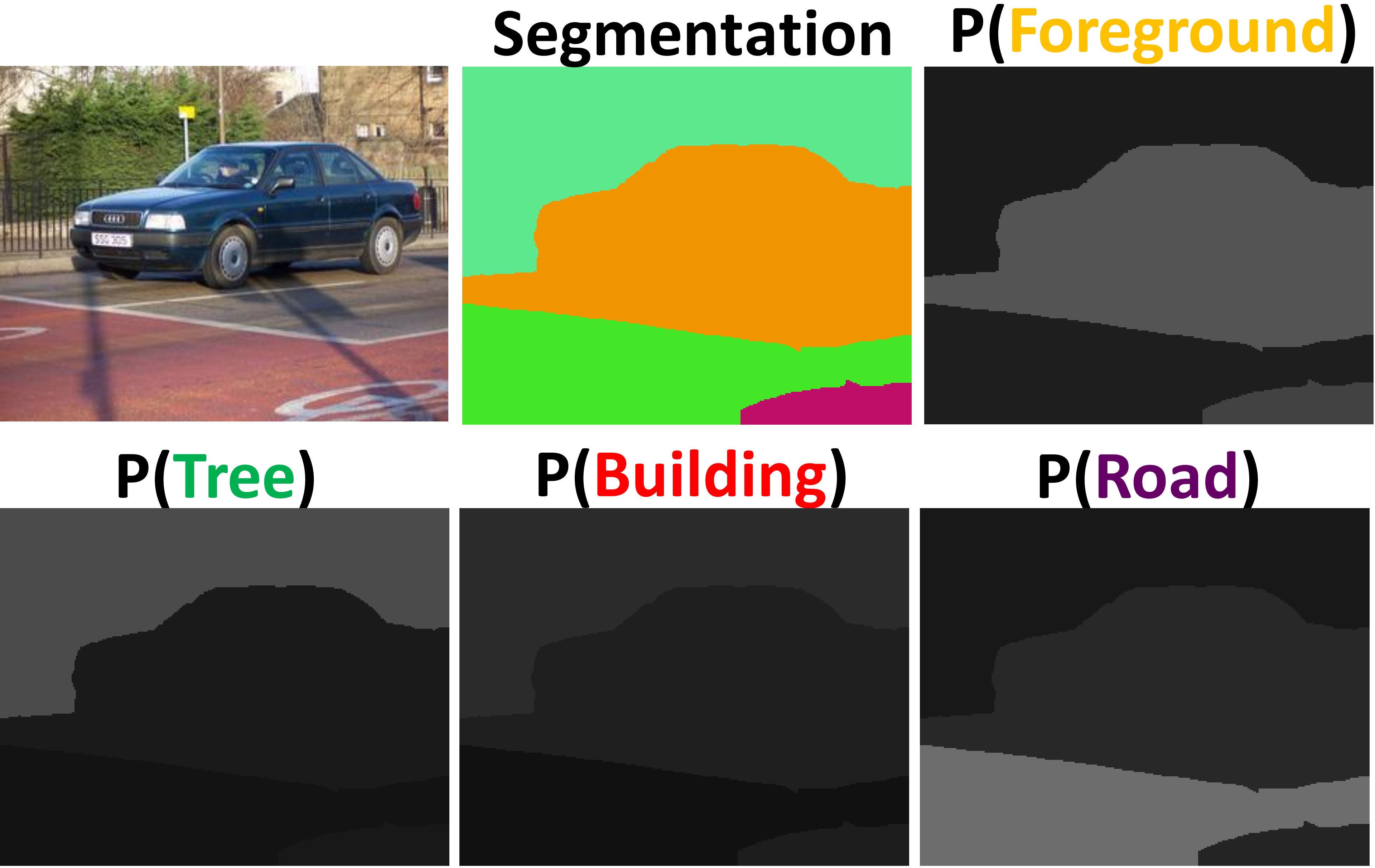
P(Building)

12%

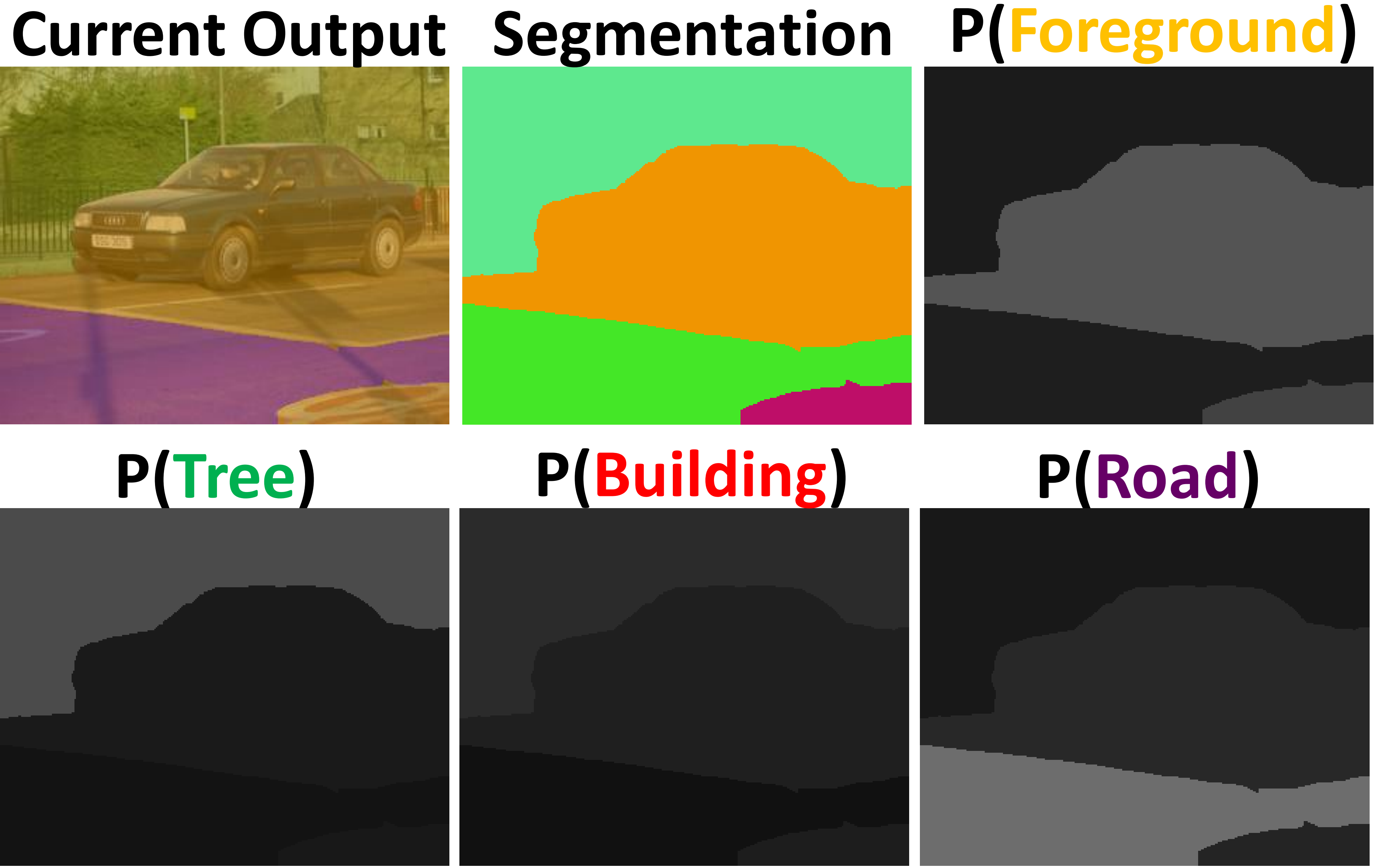
P(Road)

31%

Level 2/8 Predictions



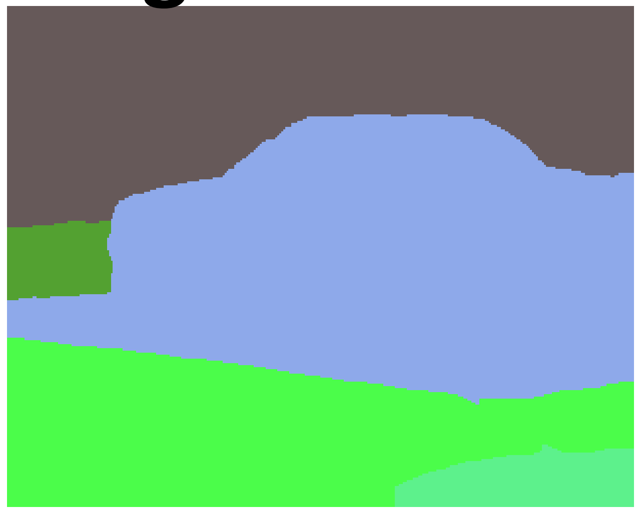
Level 2/8 Predictions



Level 3/8 Predictions

Current Output

Segmentation



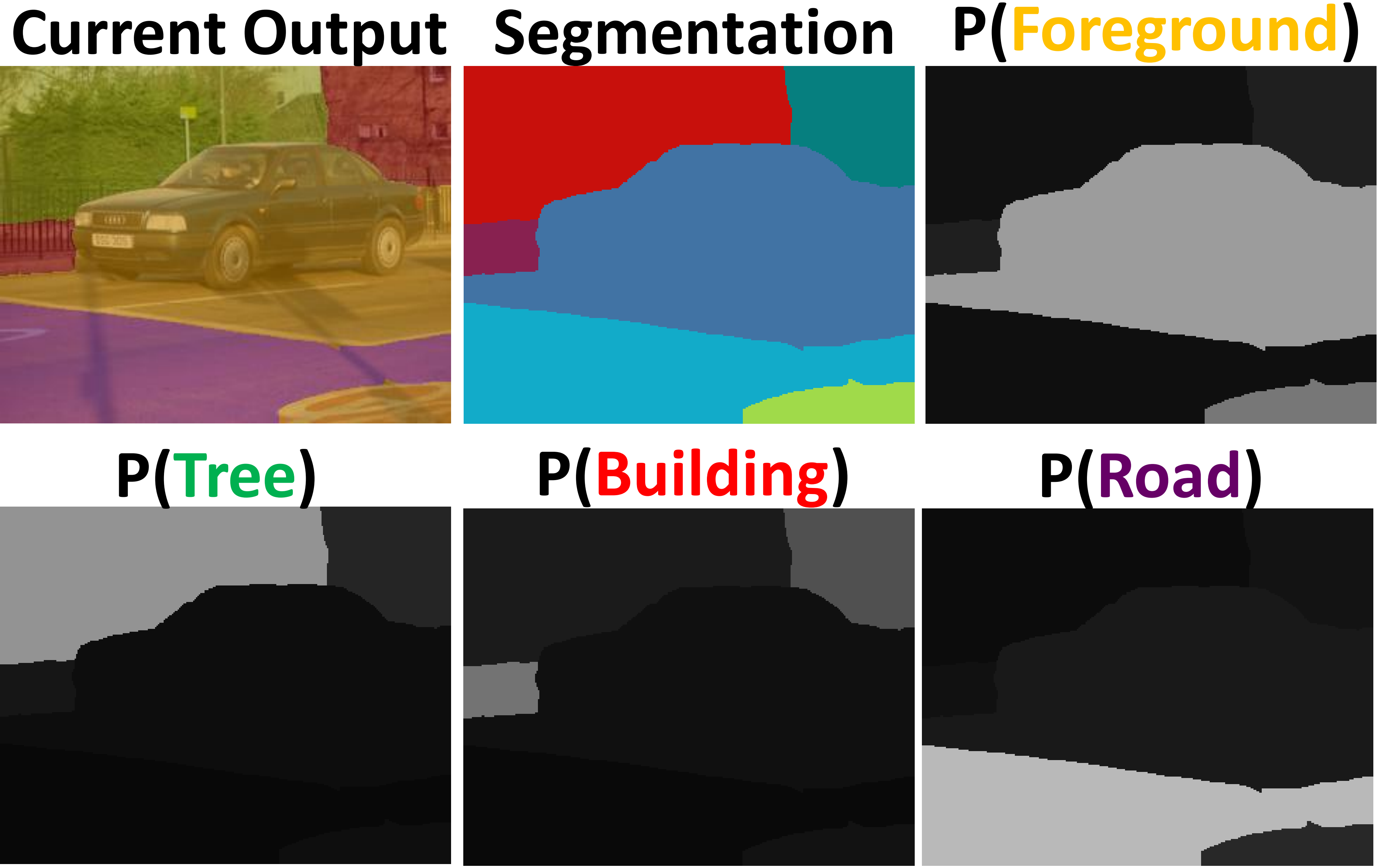
P(Foreground)

P(Tree)

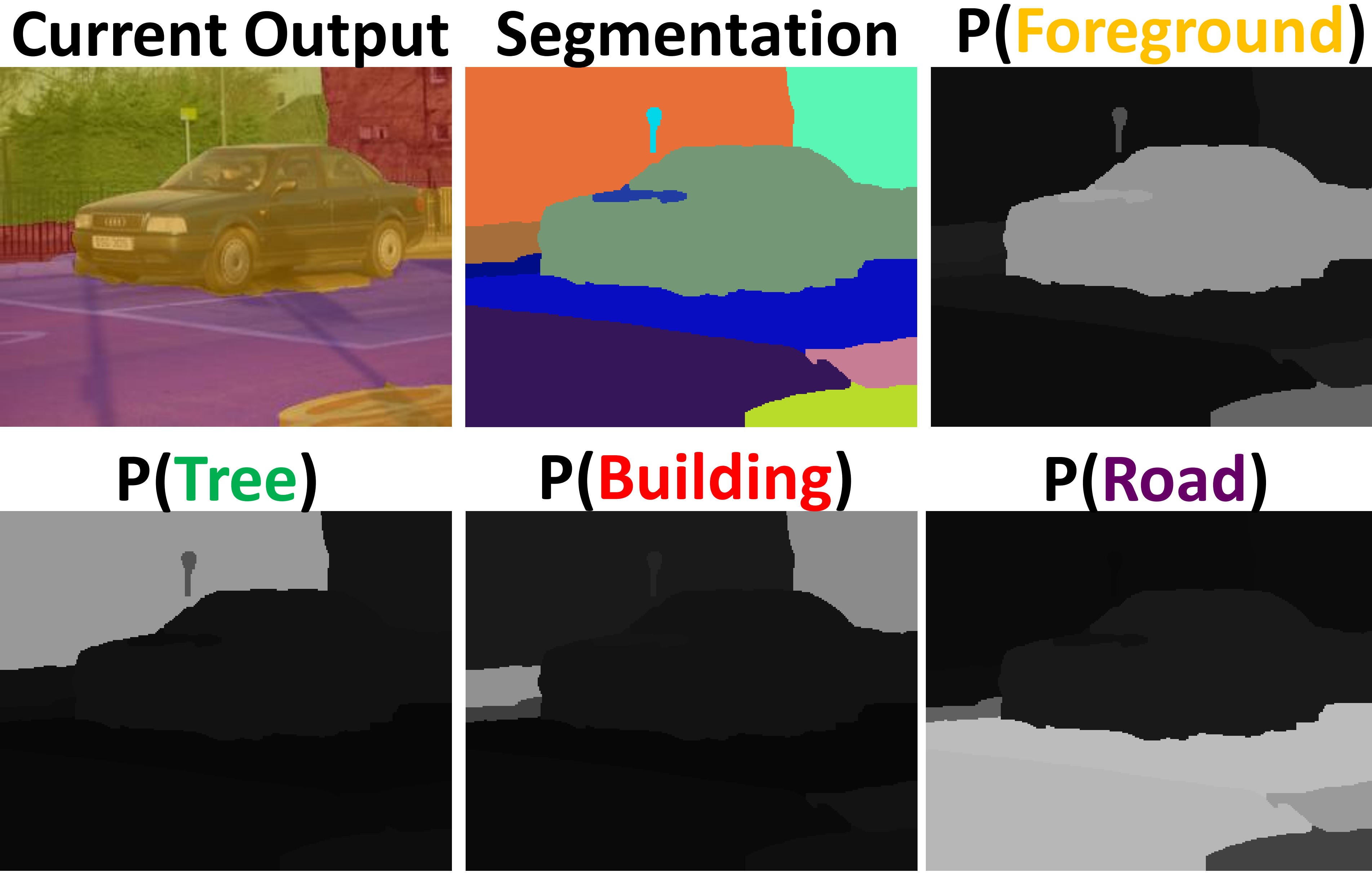
P(Building)

P(Road)

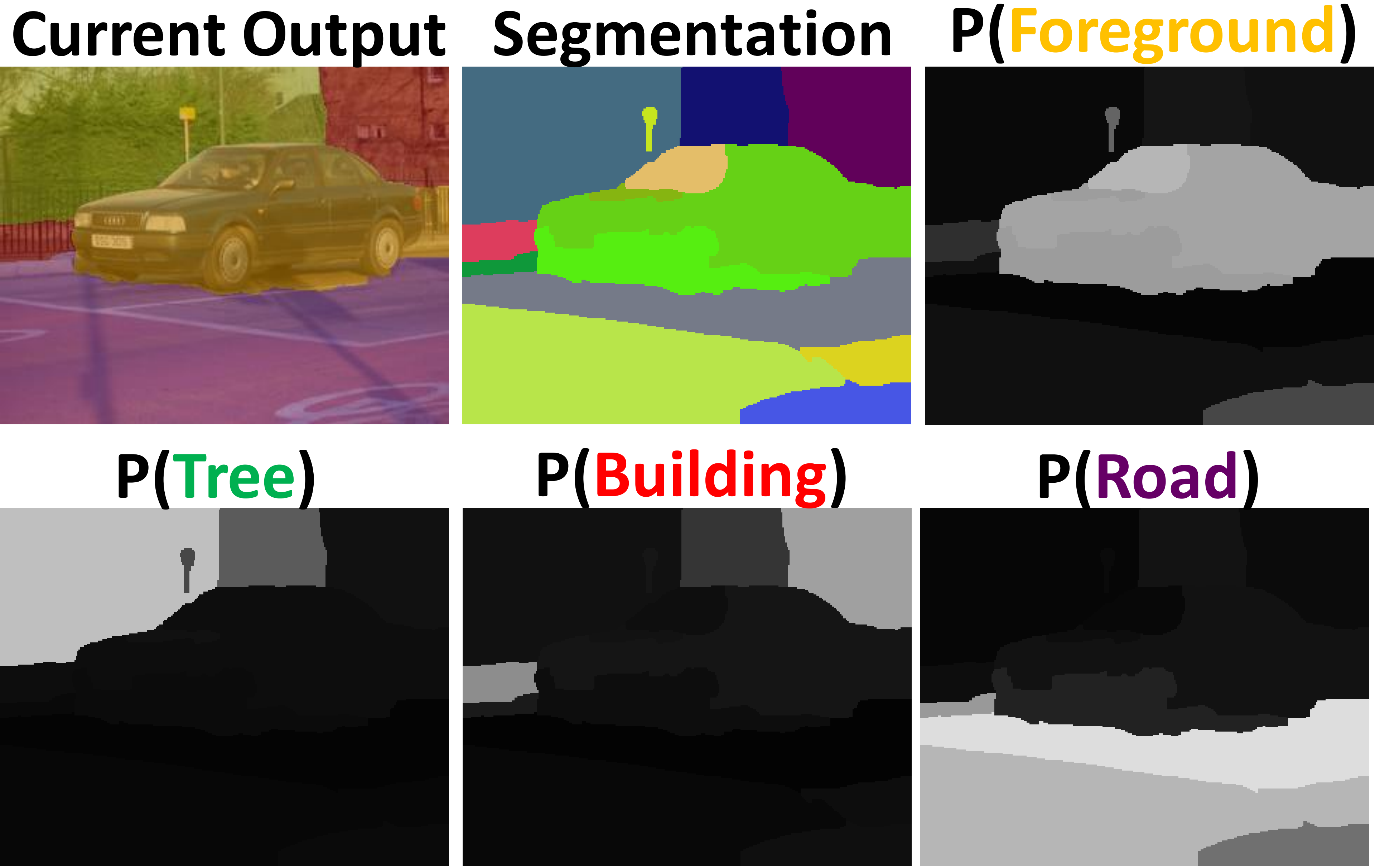
Level 4/8 Predictions



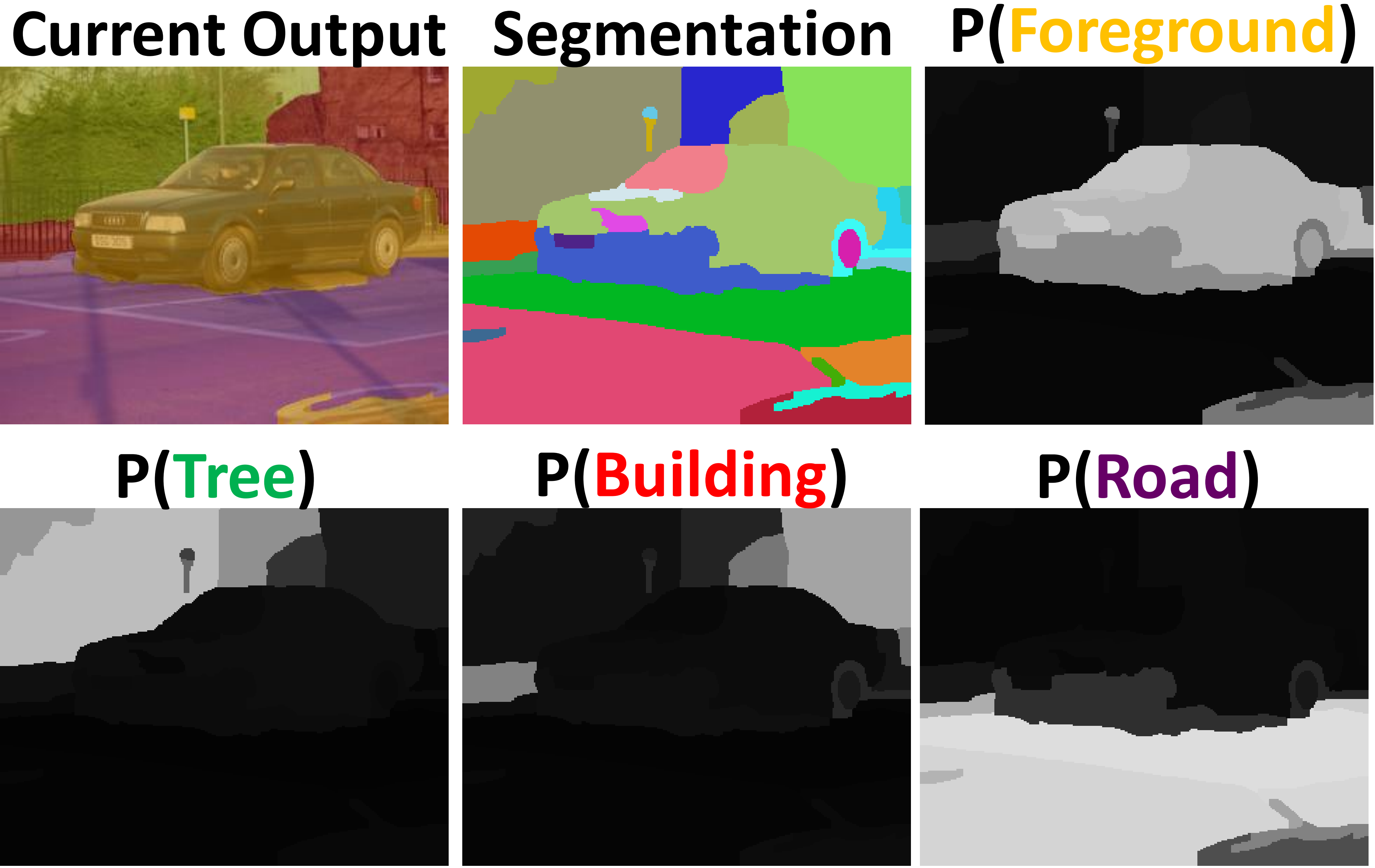
Level 5/8 Predictions



Level 6/8 Predictions



Level 7/8 Predictions



Level 8/8 Predictions

Current Output

Segmentation

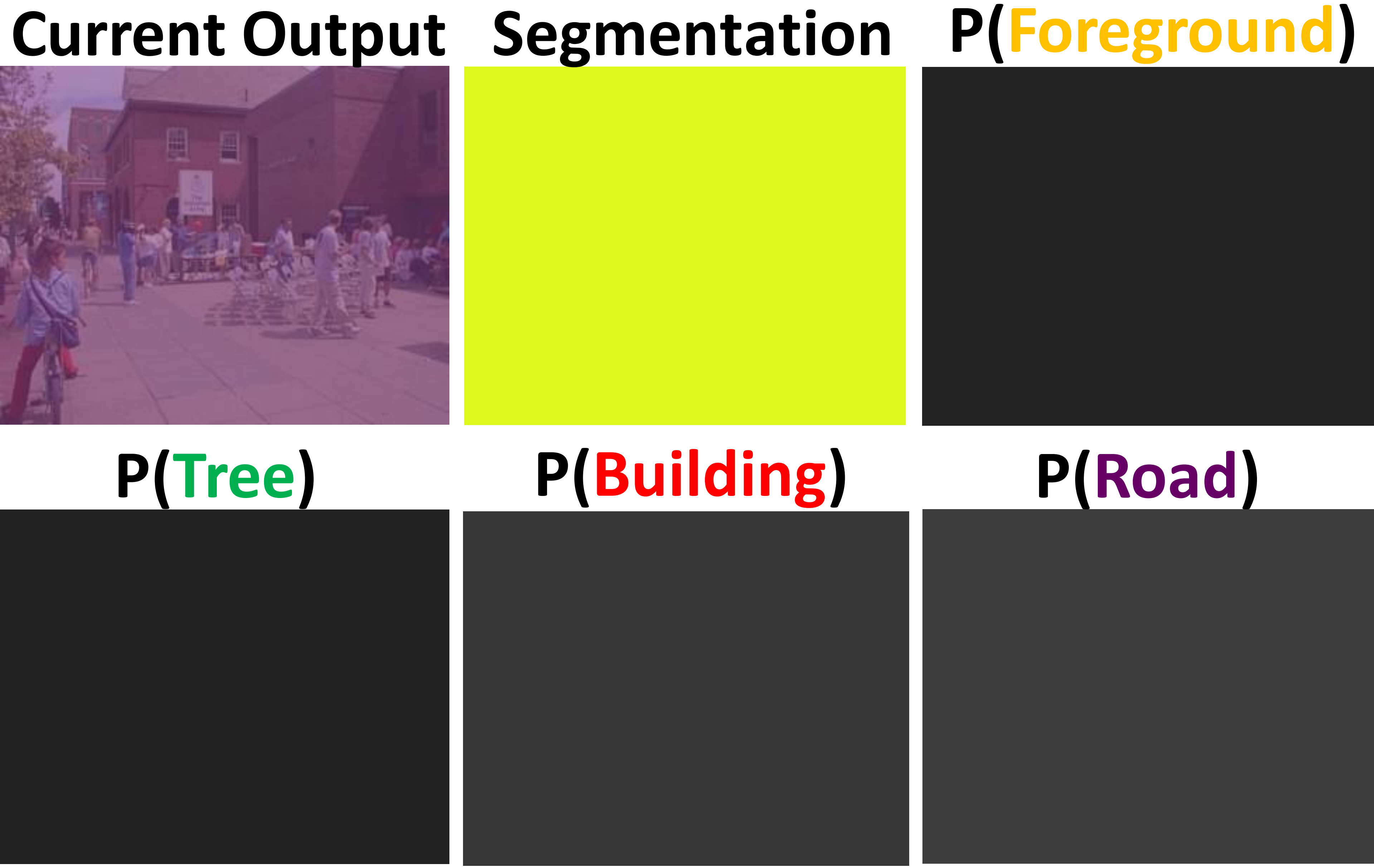
P(Foreground)

P(Tree)

P(Building)

P(Road)

Level 1/8 Predictions



Level 2/8 Predictions

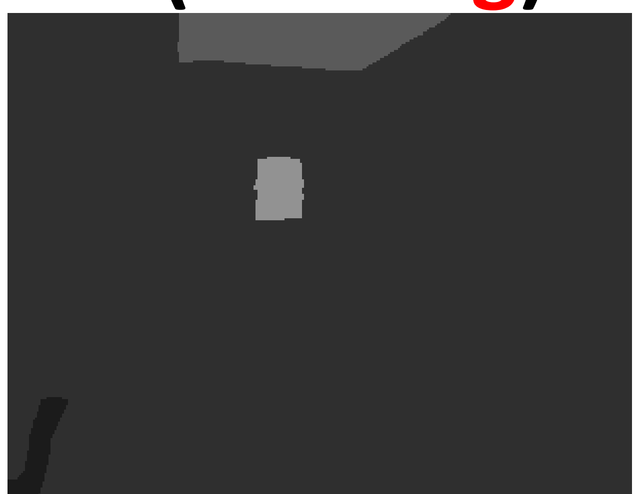
Current Output

Segmentation

P(Foreground)

P(Tree)

P(Building)



P(Road)

Level 3/8 Predictions

Current Output

Segmentation

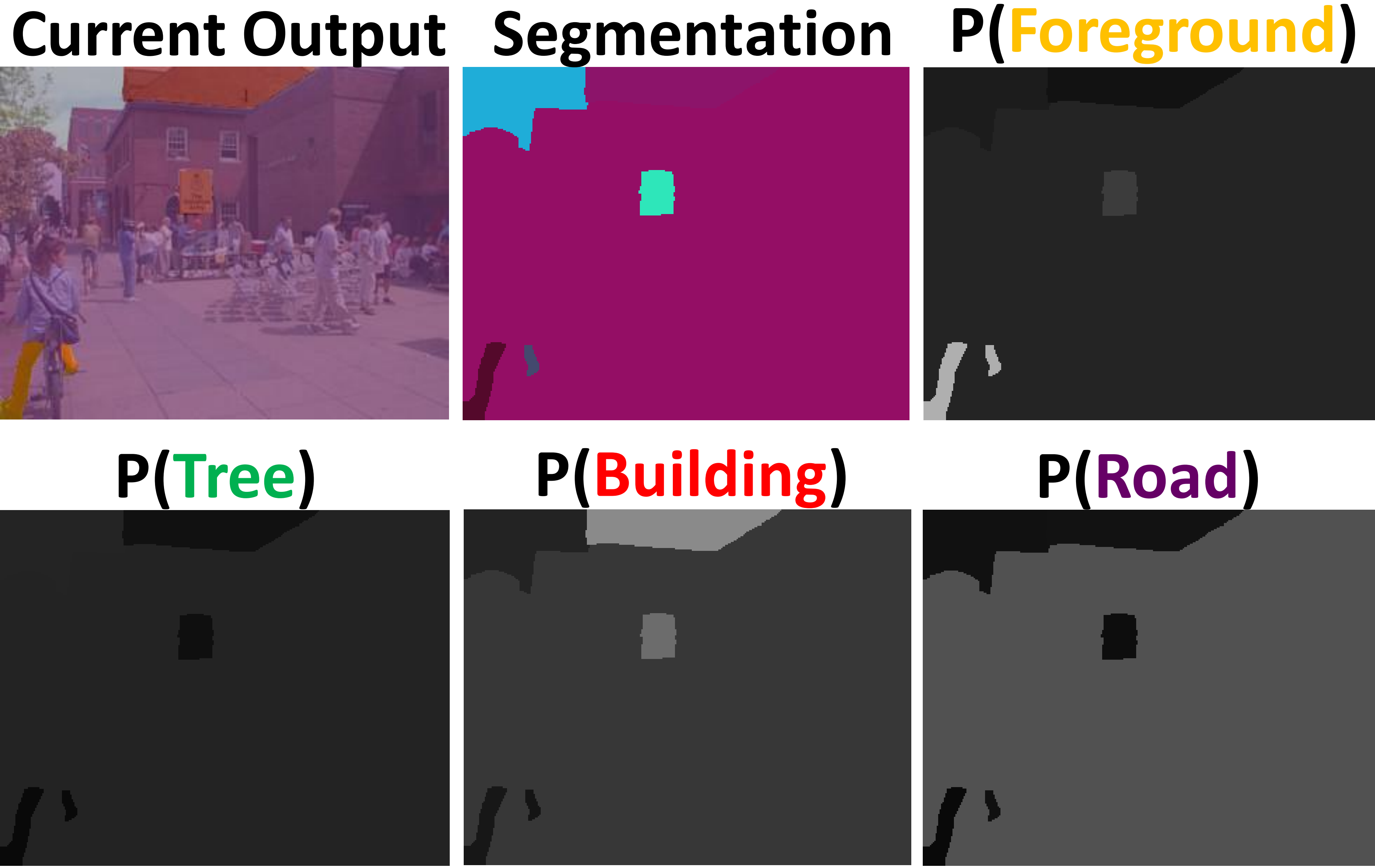
P(Foreground)

P(Tree)

P(Building)

P(Road)

Level 4/8 Predictions



Level 5/8 Predictions

Current Output

Segmentation

P(Foreground)

P(Tree)

P(Building)

P(Road)

Level 6/8 Predictions

Current Output

Segmentation

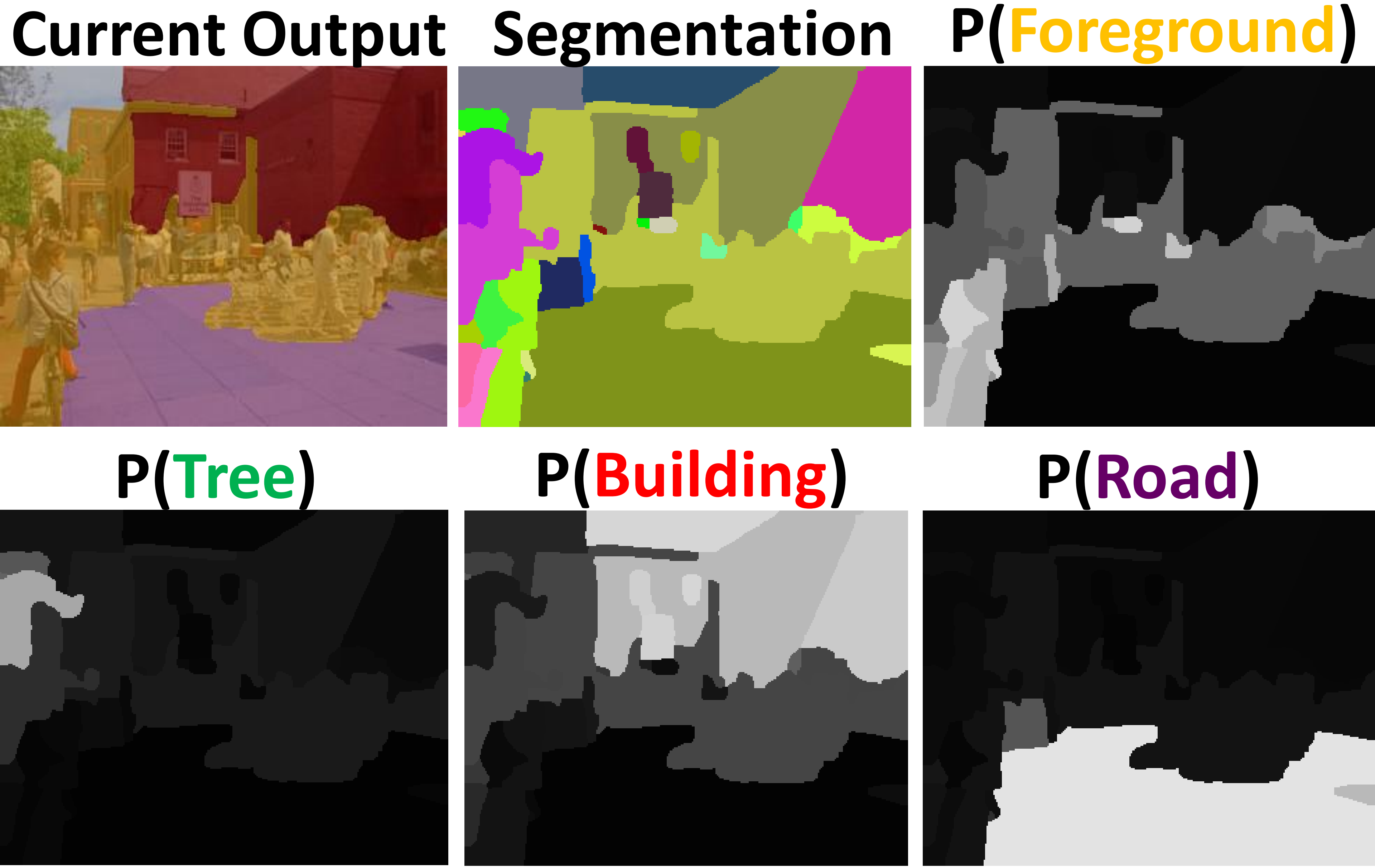
P(Foreground)

P(Tree)

P(Building)

P(Road)

Level 7/8 Predictions



Level 8/8 Predictions

Current Output

Segmentation

P(Foreground)

P(Tree)

P(Building)

P(Road)

Stanford Background Dataset

- 8 Classes
- 715 Images

Method	Avg Class Accuracy
Gould ICCV '09	**65.5**
LogReg (Baseline)	58.0
SHL (Proposed)	**66.2**

- Inference time
  - Segmentation & image features held constant

Method	sec/image
Gould ICCV '09	30 - 600
SHL (Proposed)	**10 - 12**

MSRC-21

- 21 Classes
- 591 Images

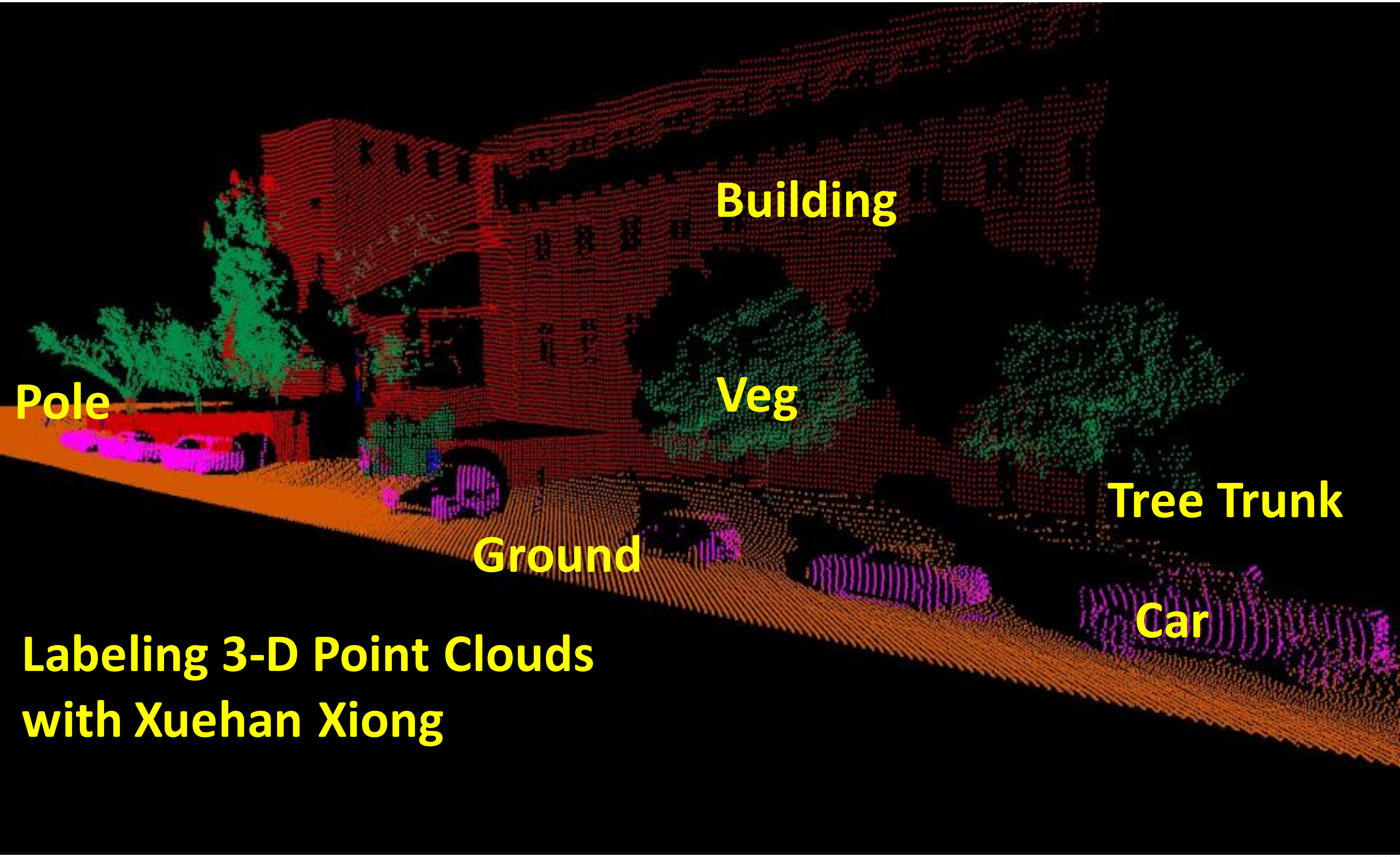
Method	Avg Class Accuracy
Gould IJCV '08	64
LogReg (Baseline)	60
SHL (Proposed)	**71**
Lim ICCV'09	67
Tu PAMI'09	69
Zhu NIPS'08	**74**
Ladicky ICCV '09	**75**

MSRC-21

- 21 Classes
- 591 Images

Method	Avg Class Accuracy
Gould IJCV '08	64
LogReg (Baseline)	60
SHL (Proposed)	**71**
Lim ICCV'09	67
Tu PAMI'09	69
Zhu NIPS'08	**74**
Ladicky ICCV '09	**75**
LogReg (Baseline)	69
SHL (Proposed)	**75**

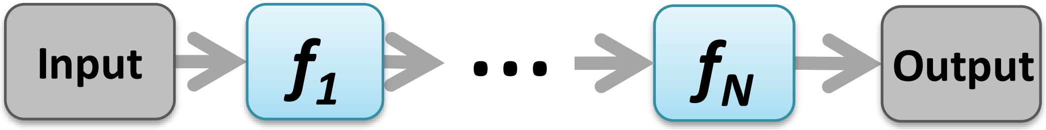
Ongoing Work



Conclusion

- **An effective structured prediction alternative**
  - High performance with no graphical model
- **Beyond site-wise representations**
  - Robust to imperfect segmentations & multiple scales

- **Prediction is a series of simple problems**
  - Stacked to avoid cascading errors and overfitting

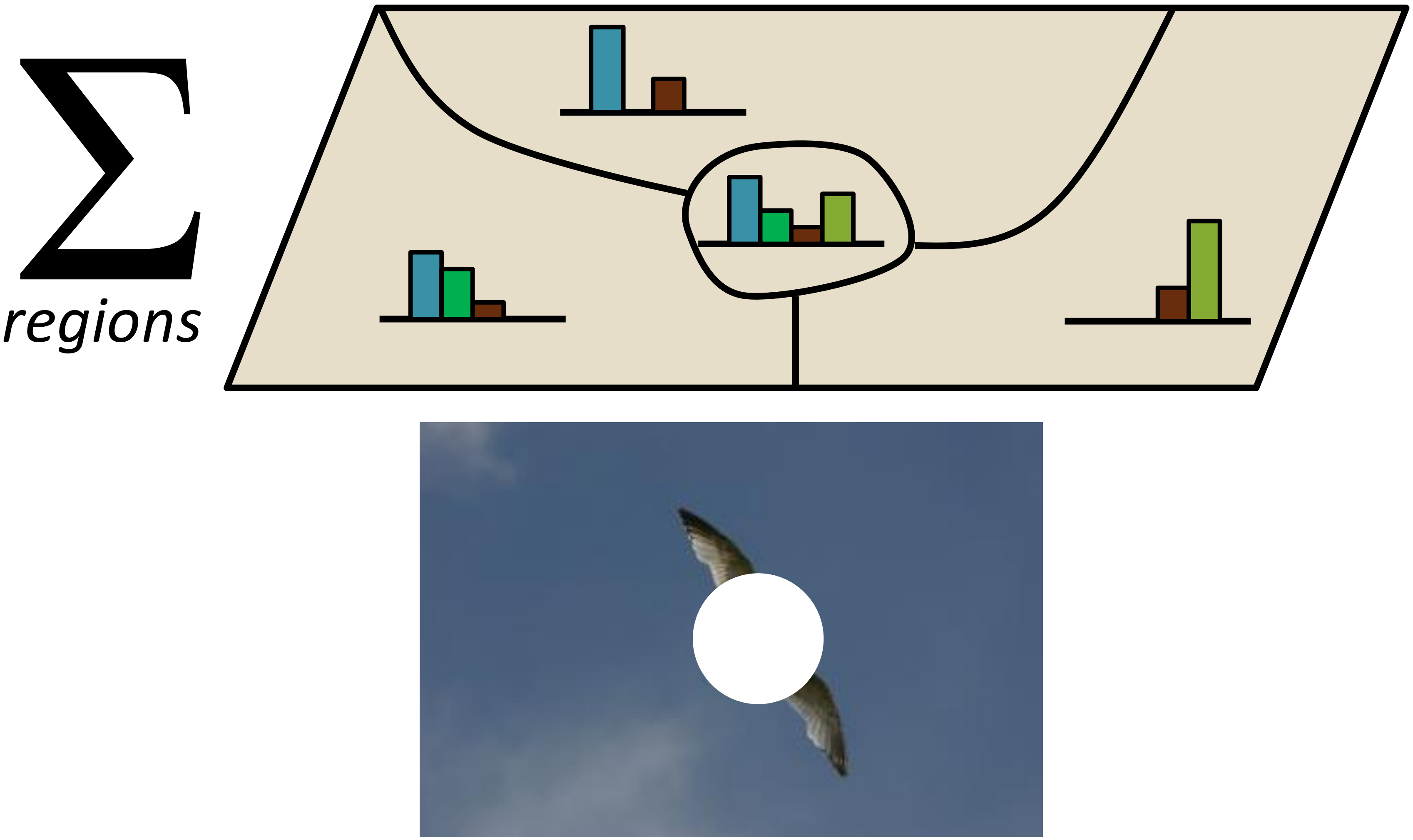


Thank You

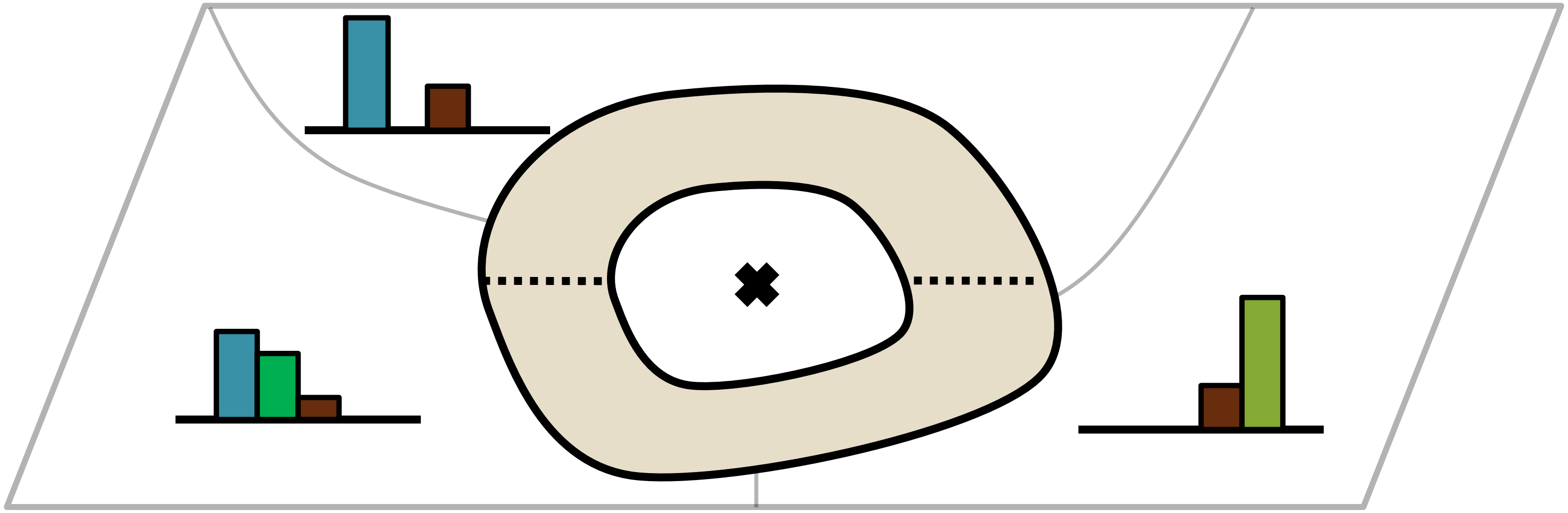
- Acknowledgements
  – QinetiQ North America Robotics Fellowship
  – ONR MURI: Reasoning in Reduced Information Spaces
  – Reviewers, S. Ross, A. Grubb, B. Becker, J.-F. Lalonde

- Questions?

Image-wise

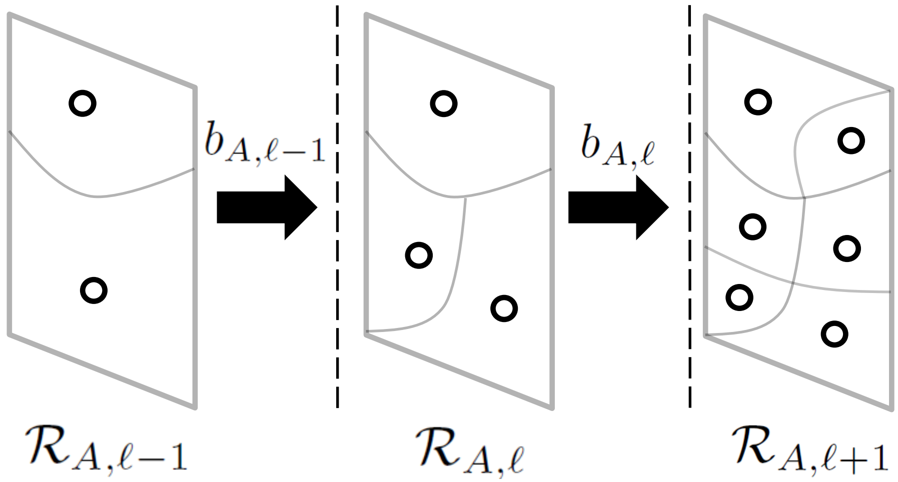


Spatial neighborhood

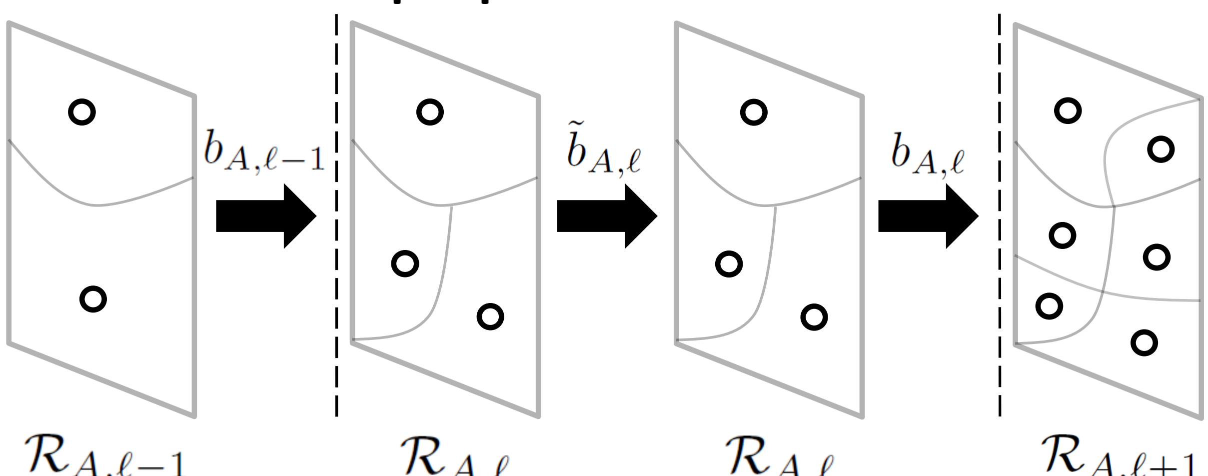


Interactions

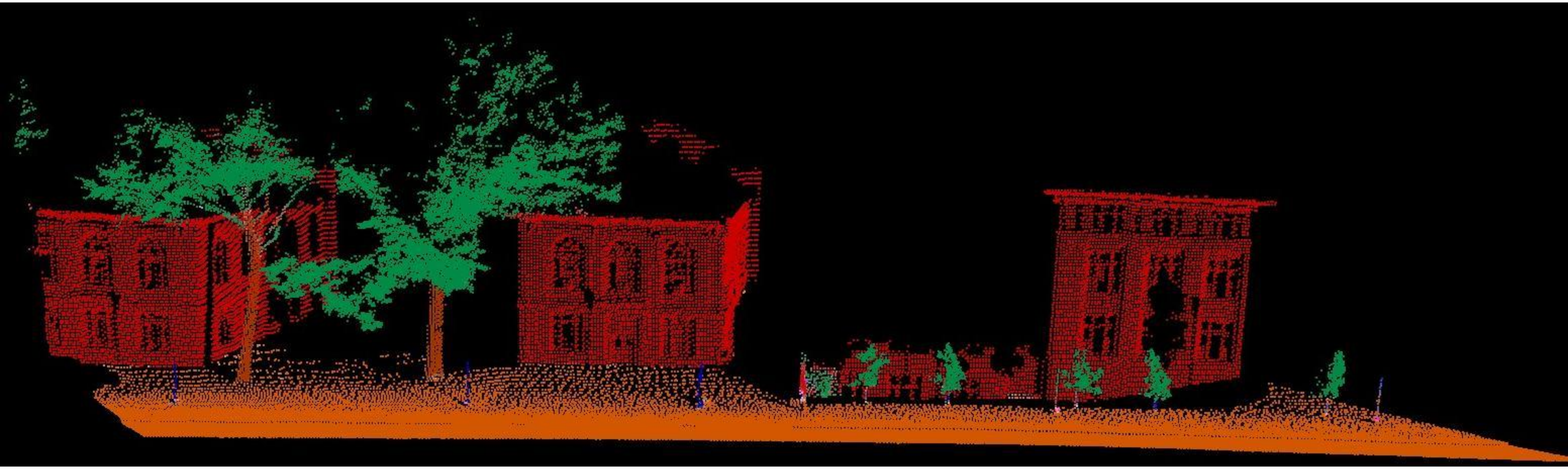
- Described in this talk

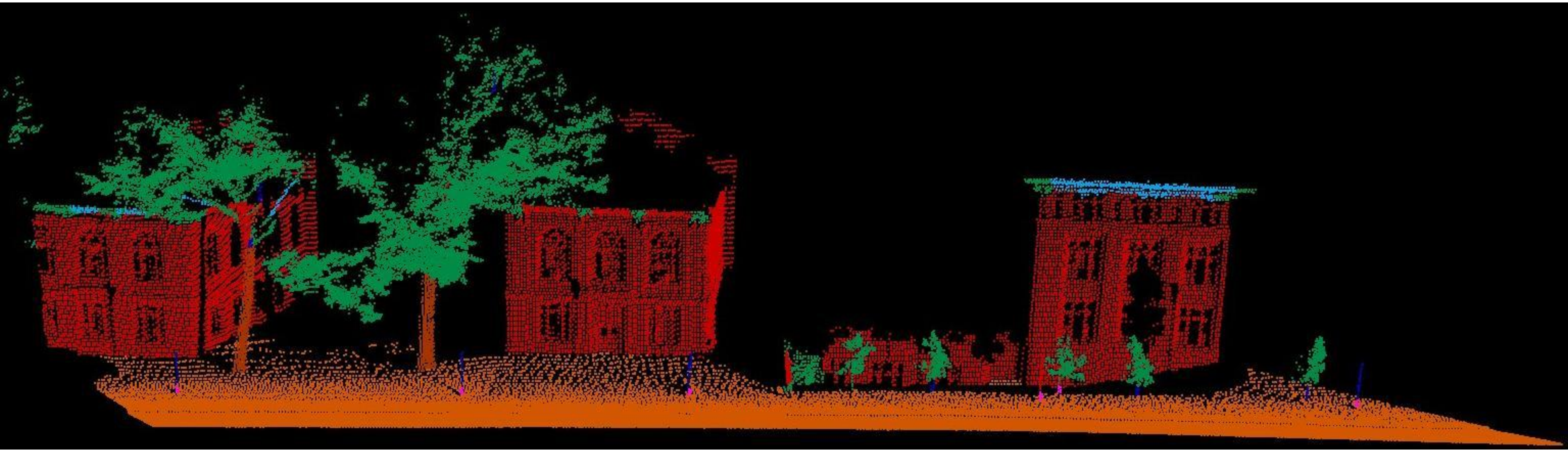


- Described in the paper



SHL vs. M3N





SHL vs. M3N