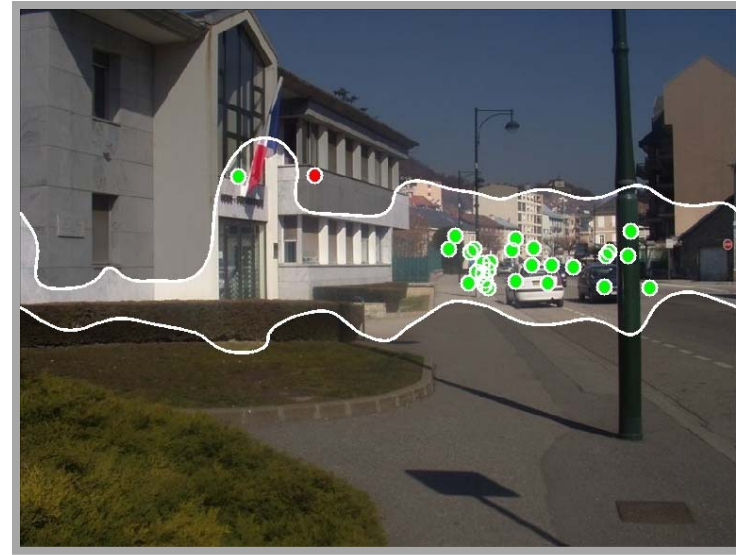
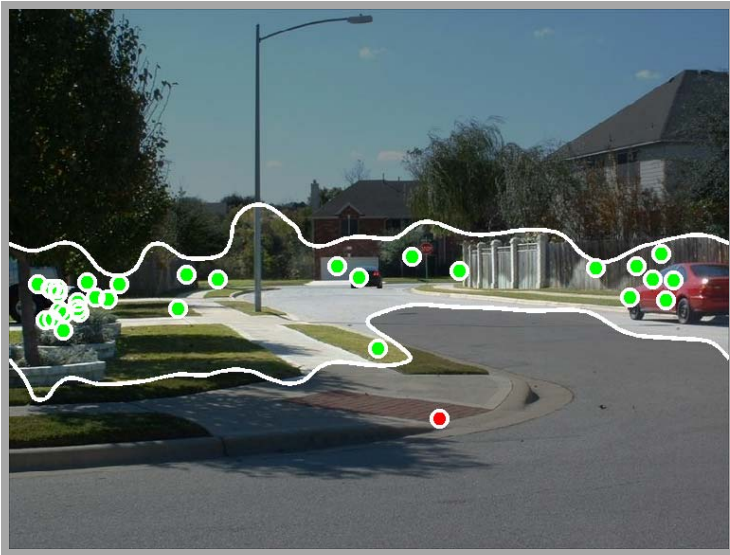
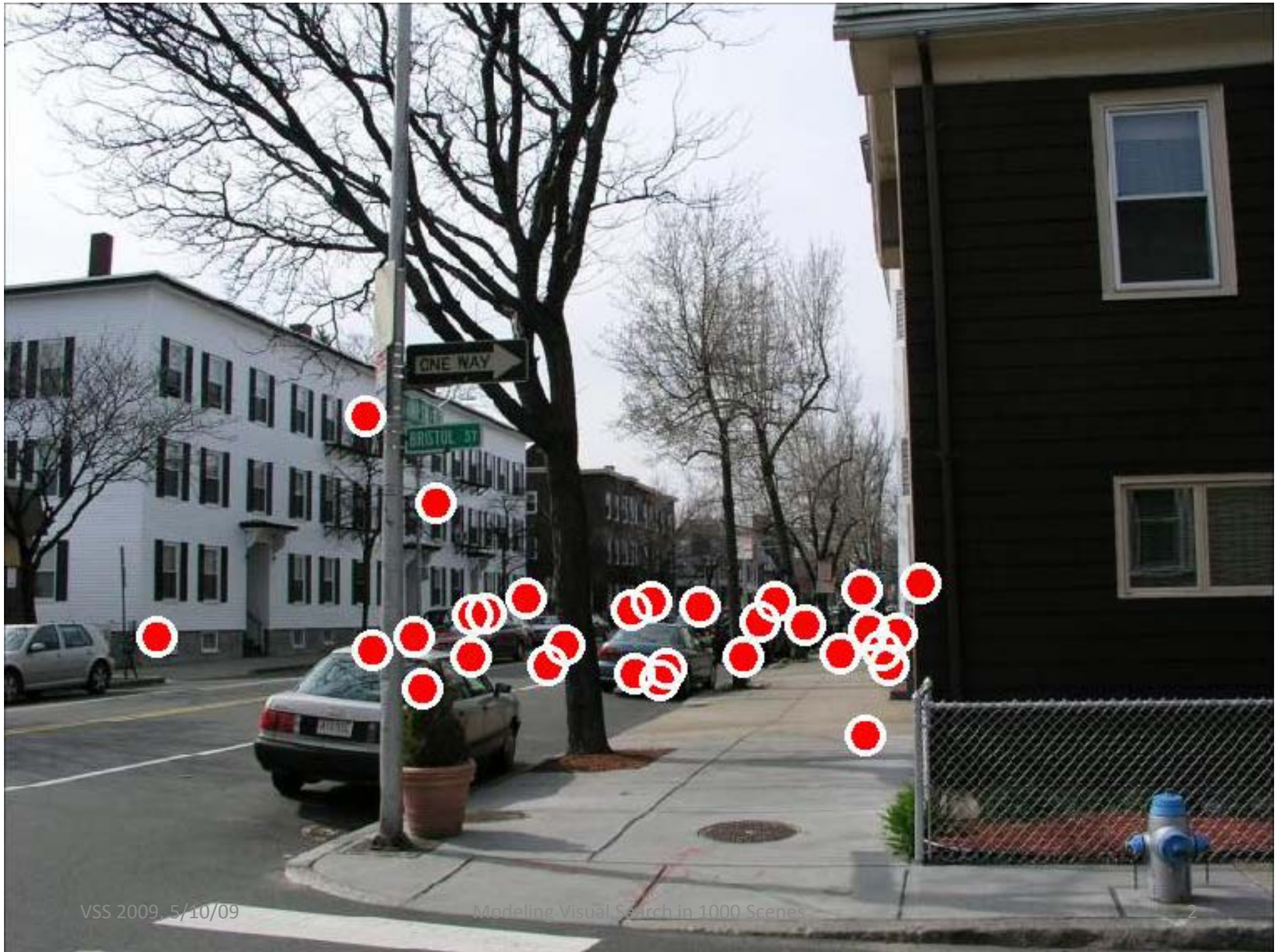


Modeling visual search in a thousand scenes: The roles of saliency, target features, and scene context

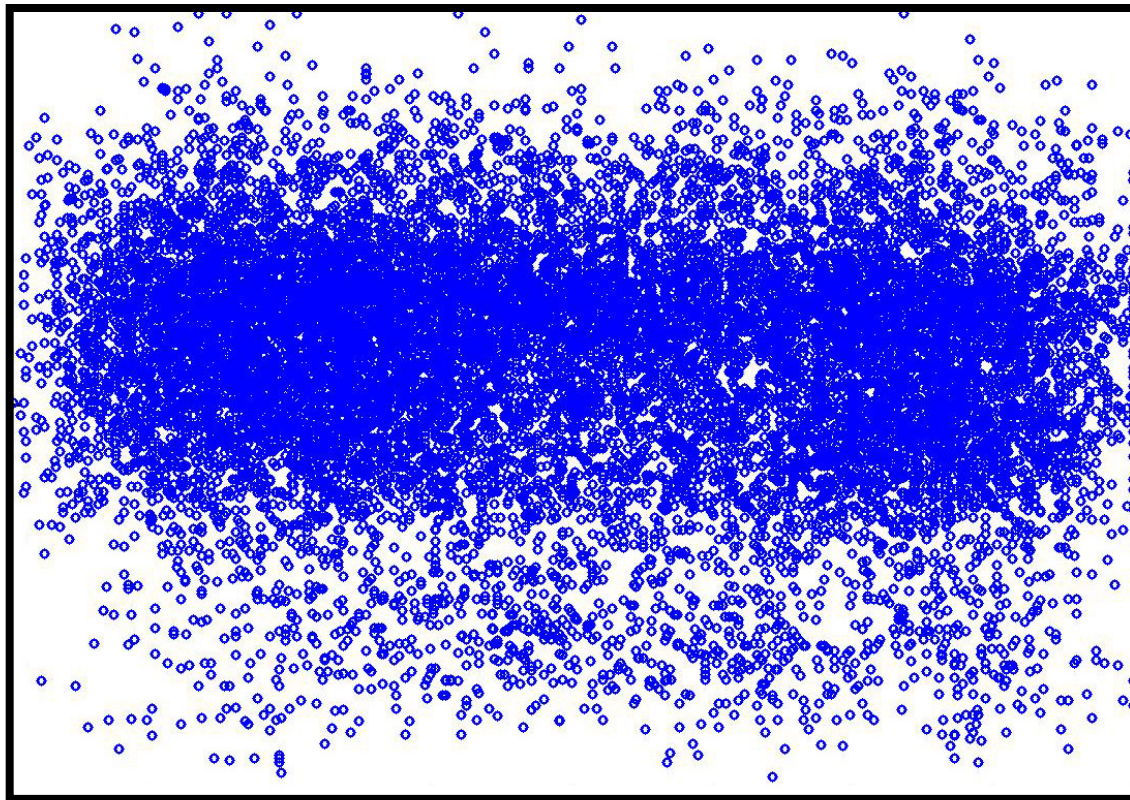


Krista Ehinger, Barbara Hidalgo-Sotelo, Antonio Torralba, & Aude Oliva



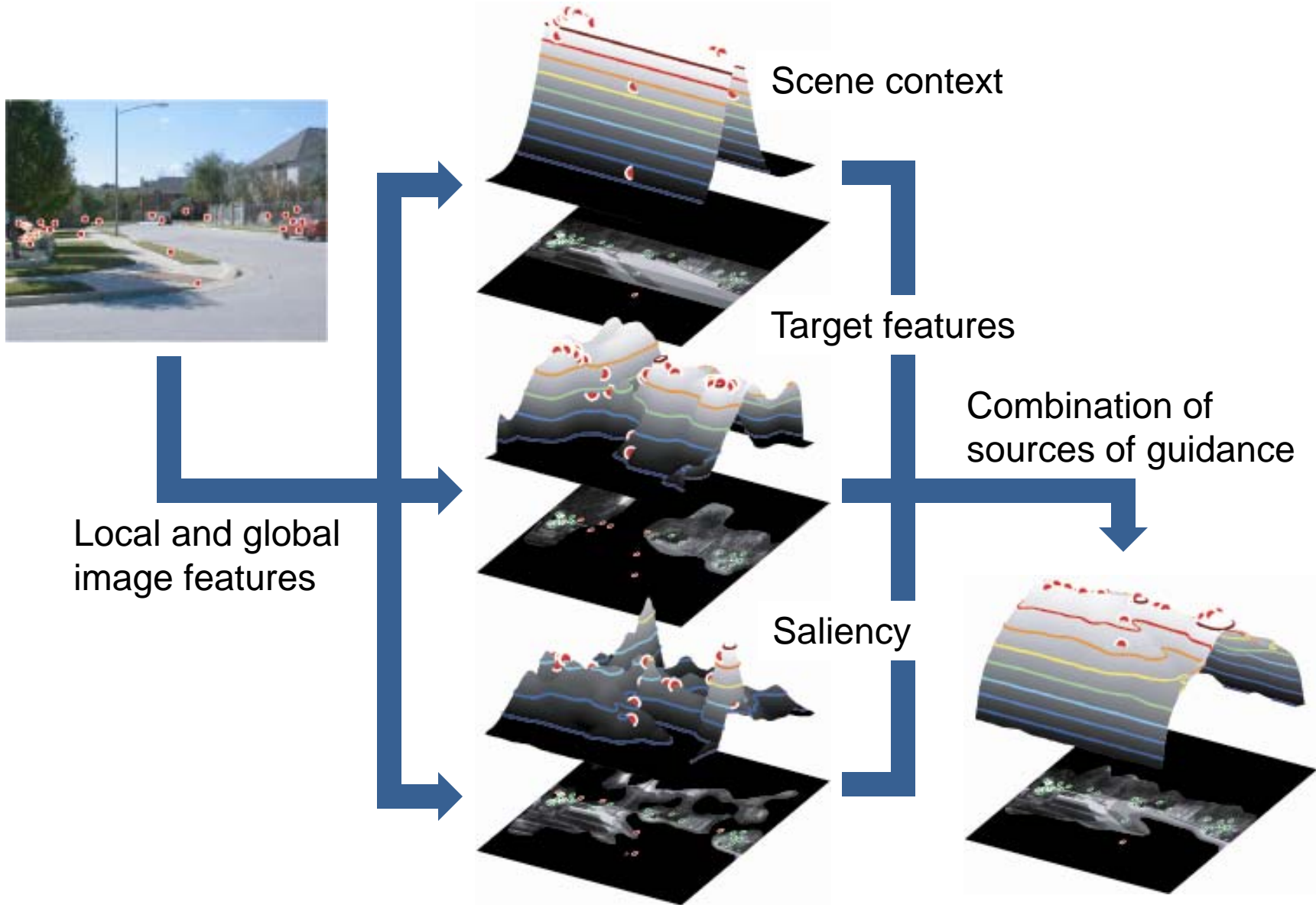
Overview of experiment

- 14 participants, 912 images = 45,144 fixations
- Person present/absent?



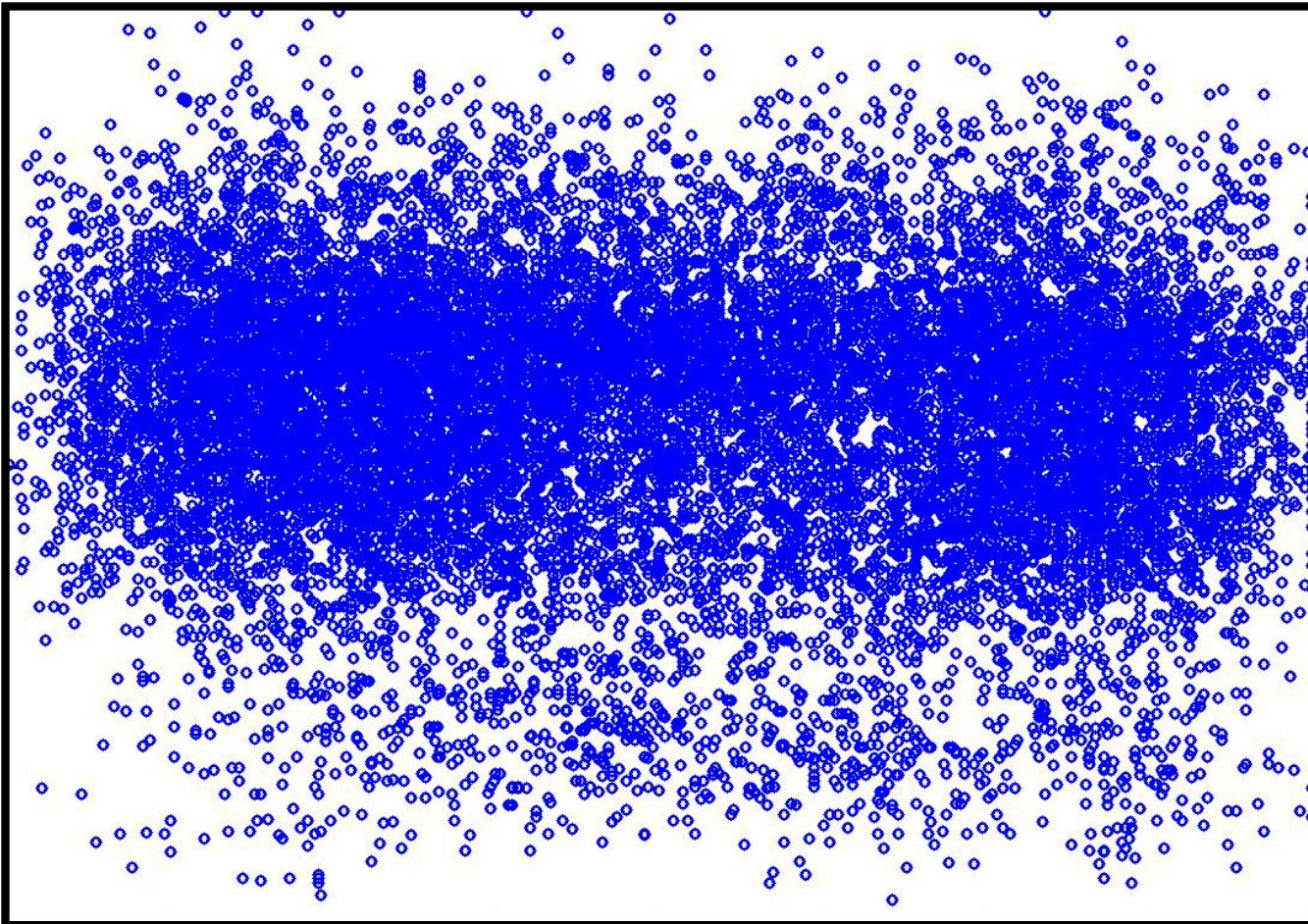
Fixations 1-3 on target-absent scenes

Overview of Model



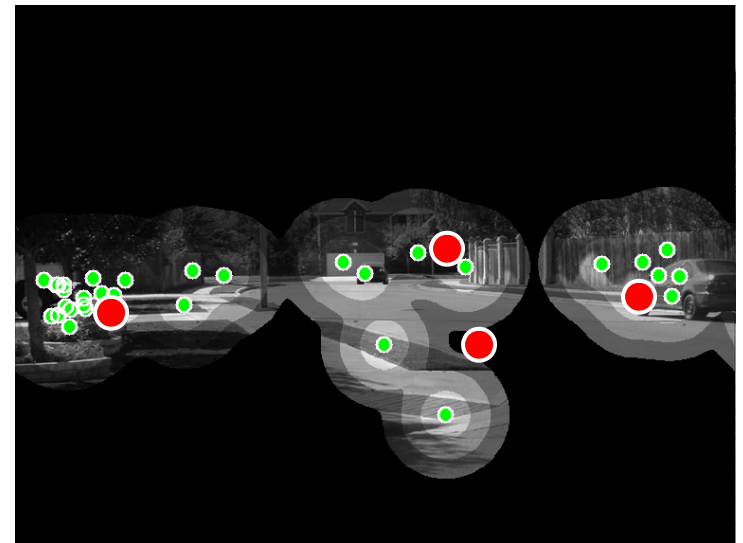
Overview of experiment

- 14 participants, 912 images = 45,144 fixations



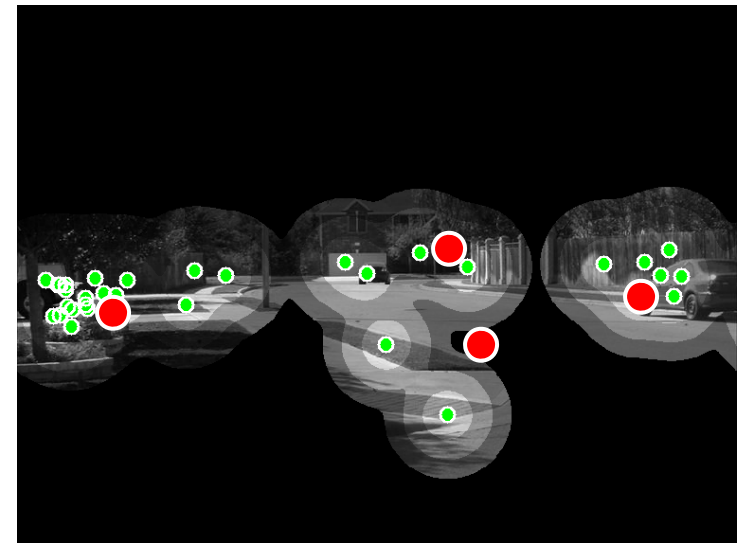
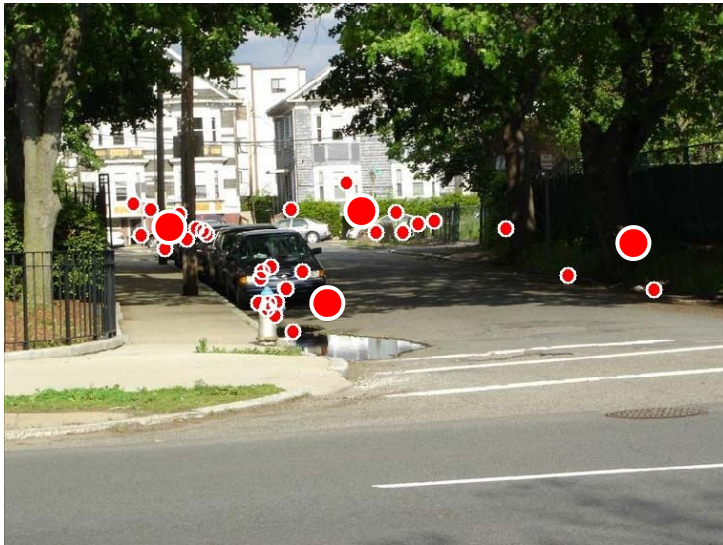
Human Agreement

- **Inter-observer agreement** = upper bound for model performance



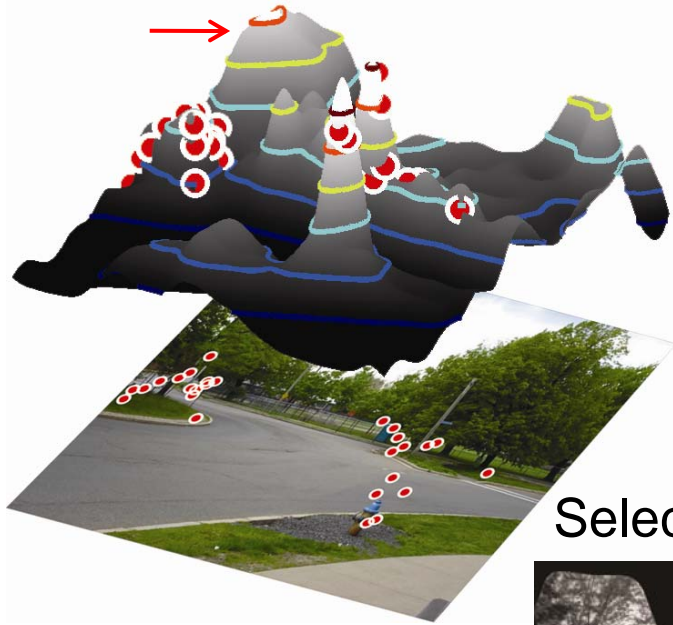
Human Agreement

- **Inter-observer agreement** = upper bound for model performance
- **Cross-image control** = lower bound for model performance

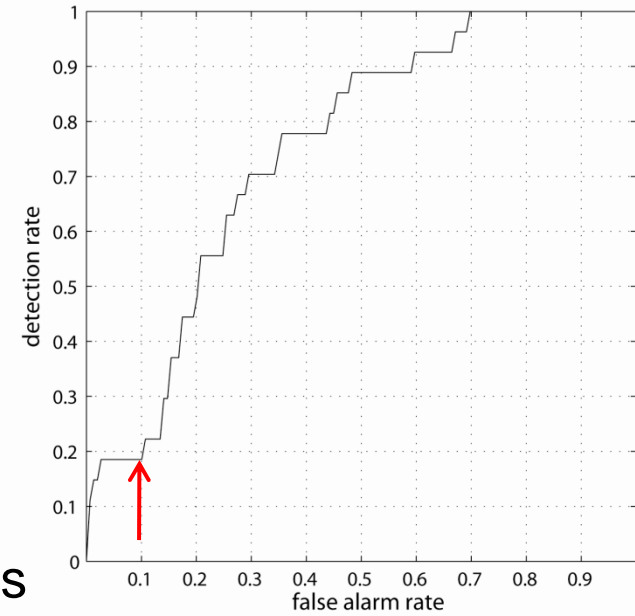
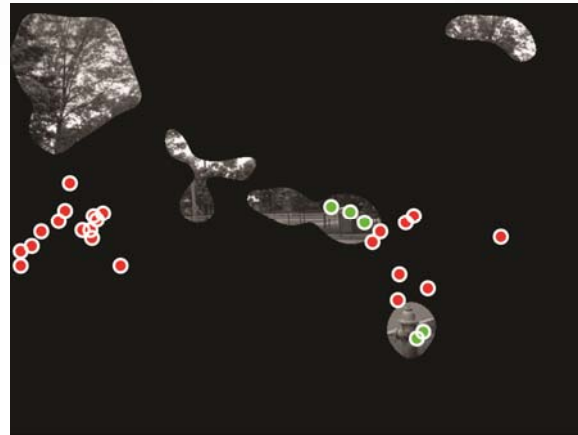


The ROC curve

Model



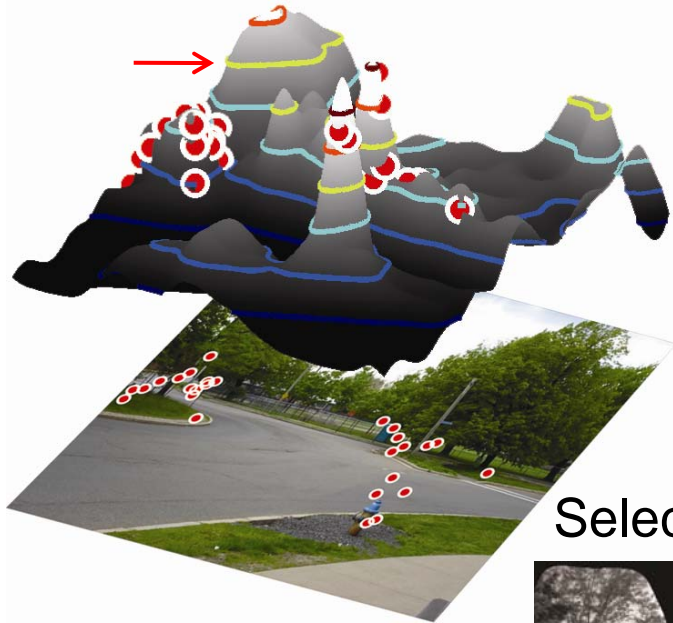
Selected image regions



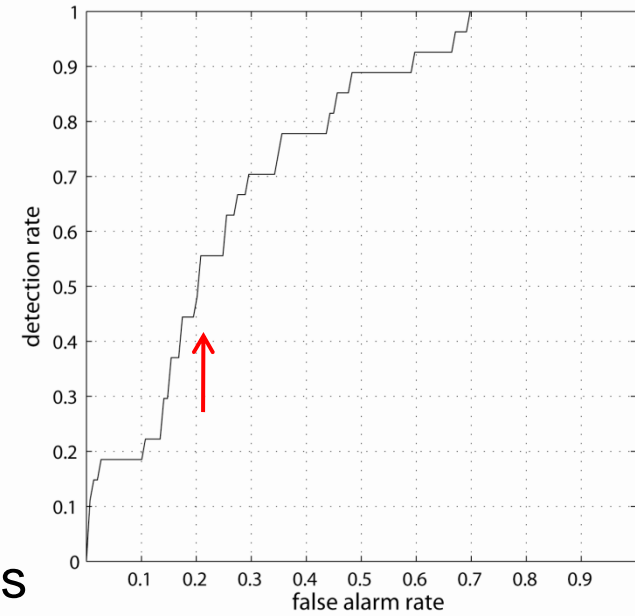
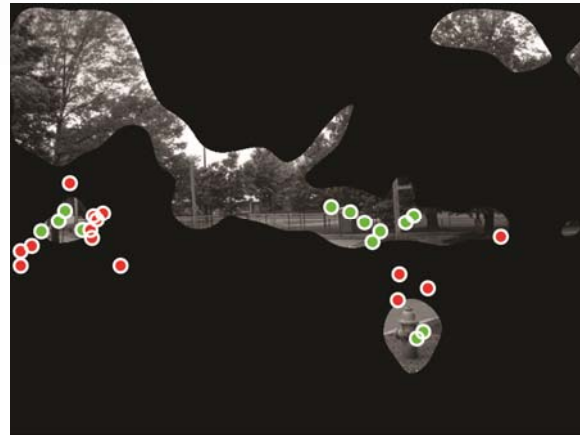
ROC curve

The ROC curve

Model



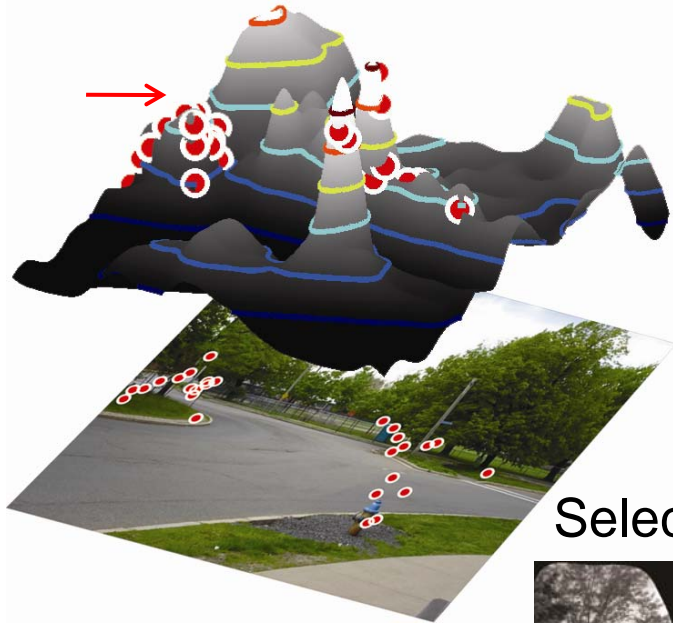
Selected image regions



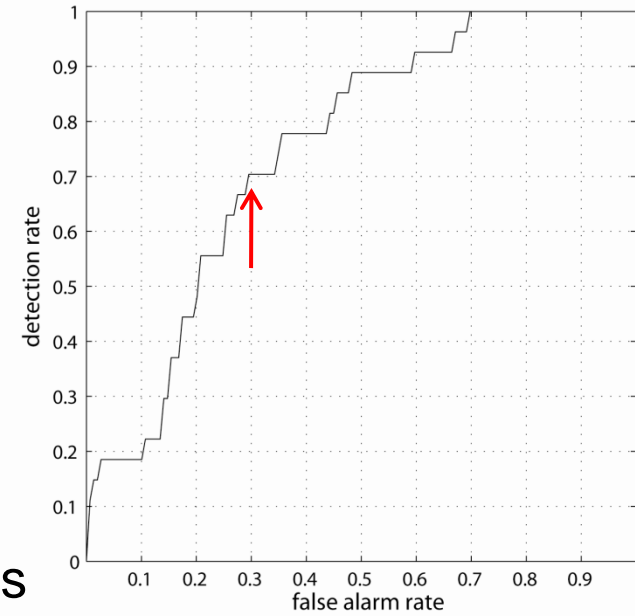
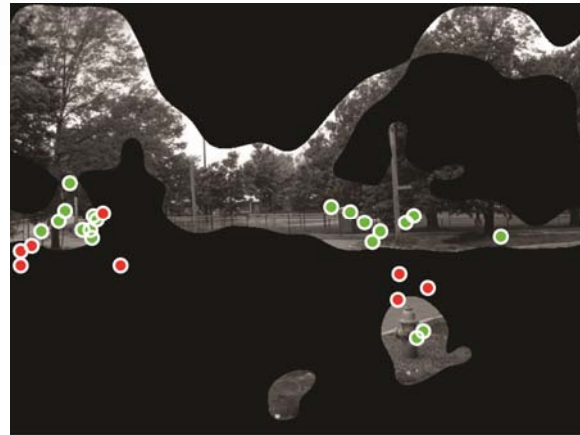
ROC curve

The ROC curve

Model



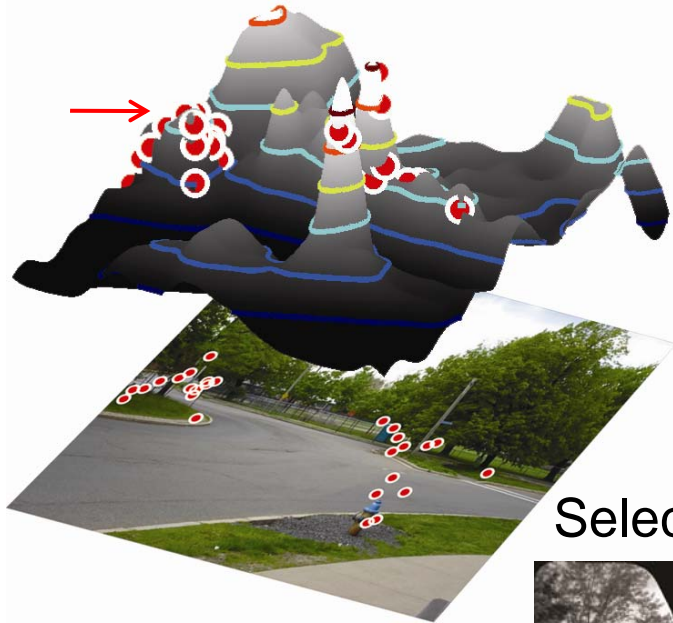
Selected image regions



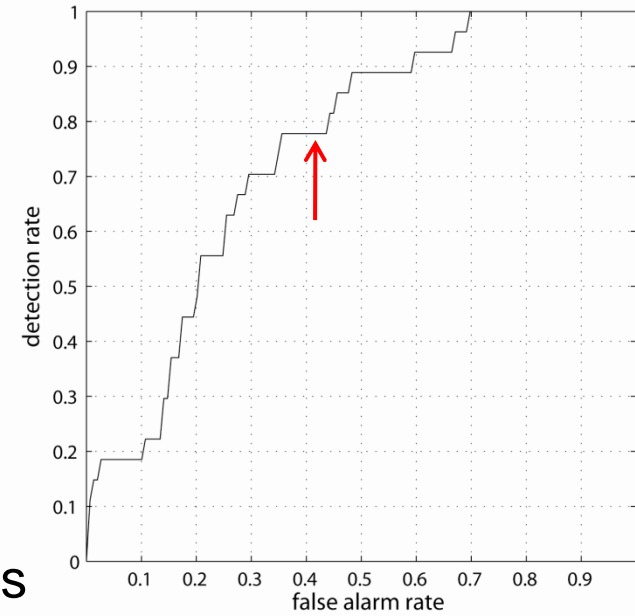
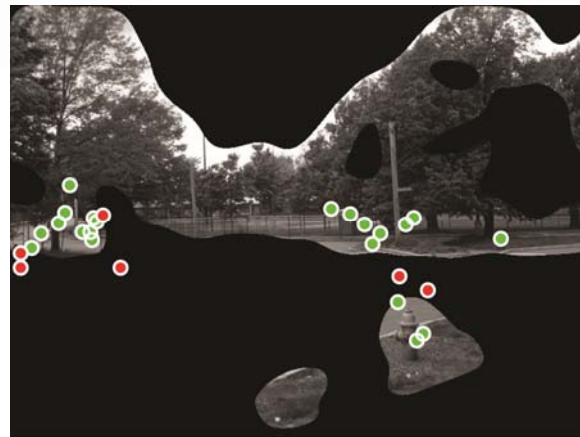
ROC curve

The ROC curve

Model



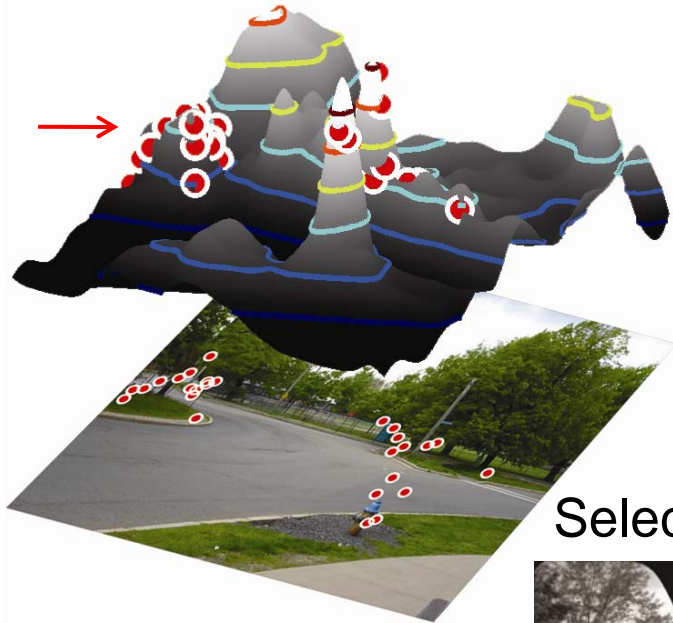
Selected image regions



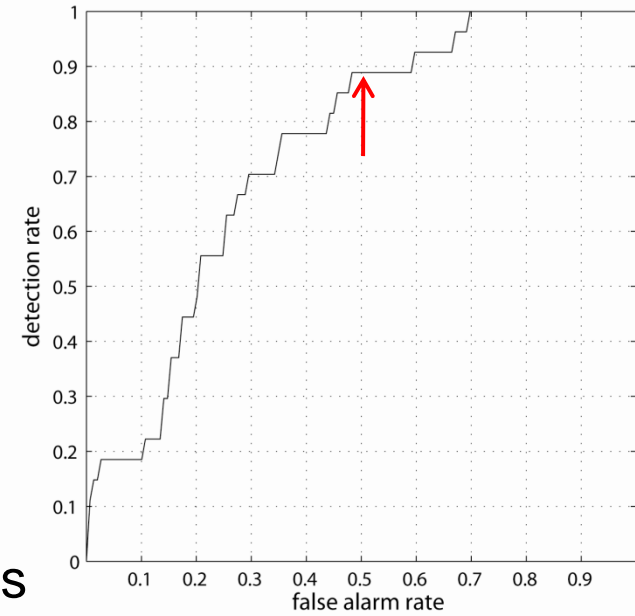
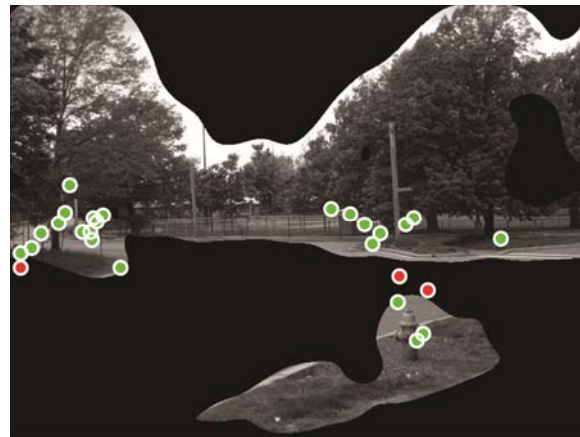
ROC curve

The ROC curve

Model



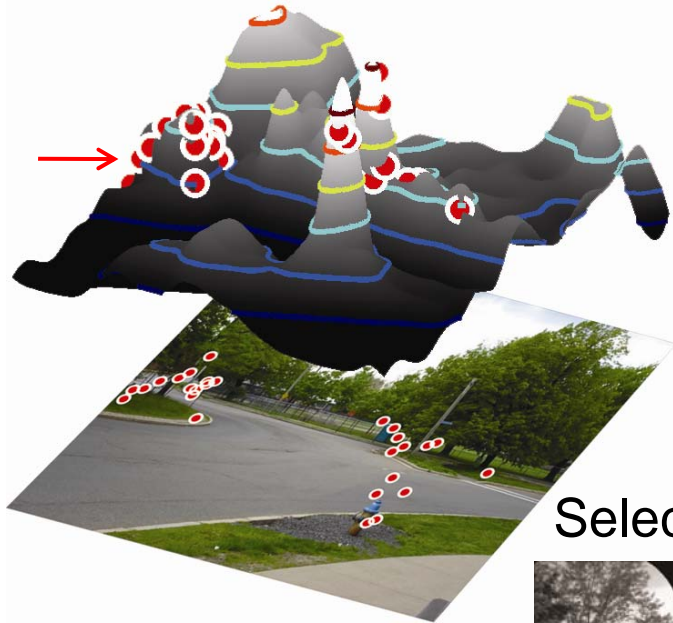
Selected image regions



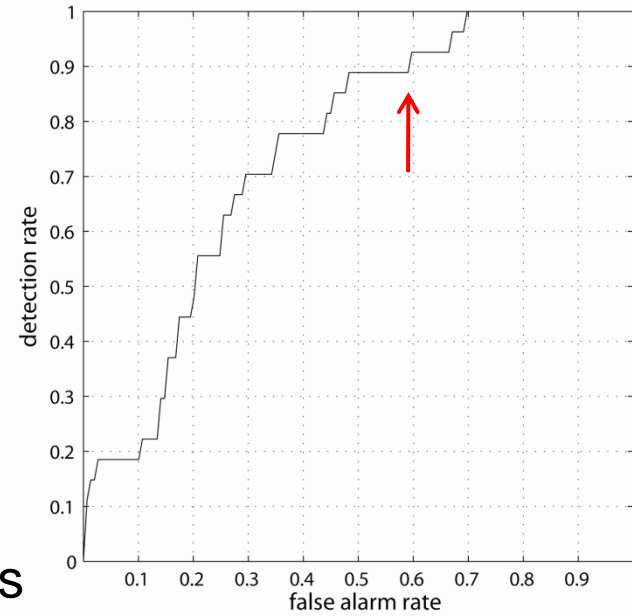
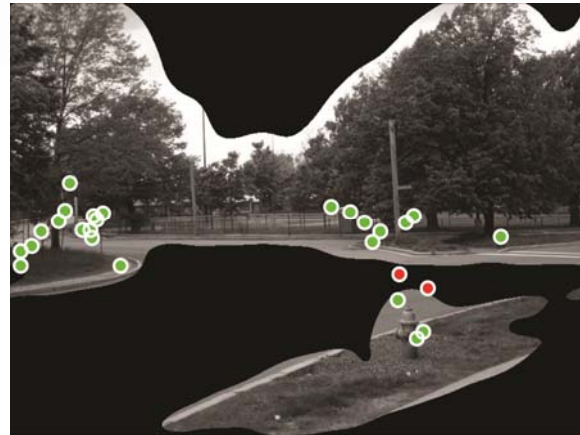
ROC curve

The ROC curve

Model



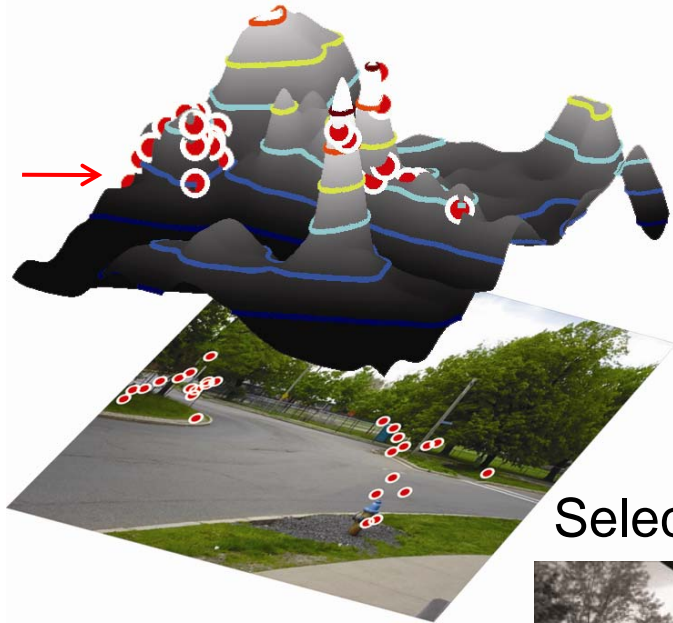
Selected image regions



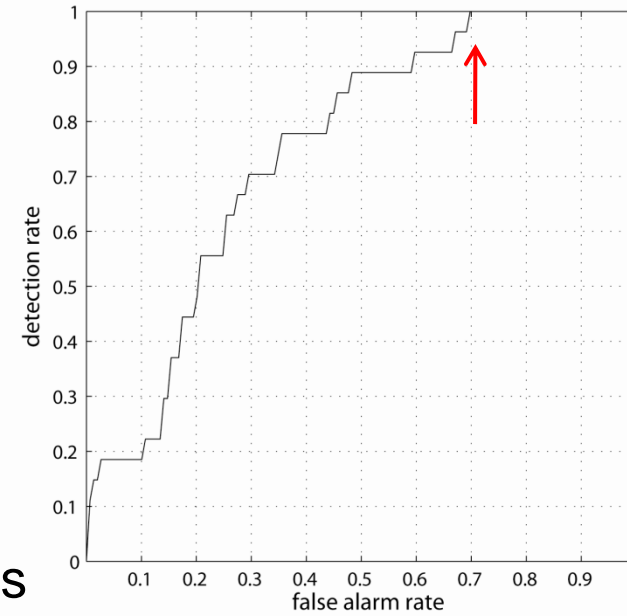
ROC curve

The ROC curve

Model



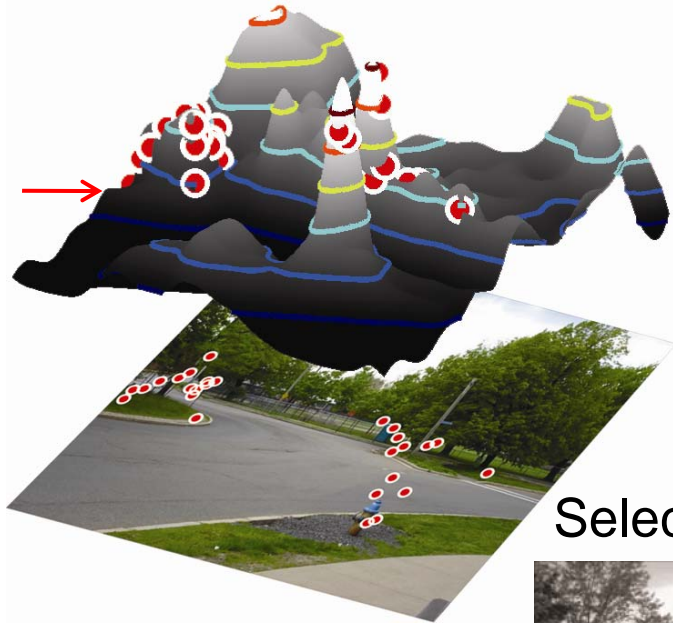
Selected image regions



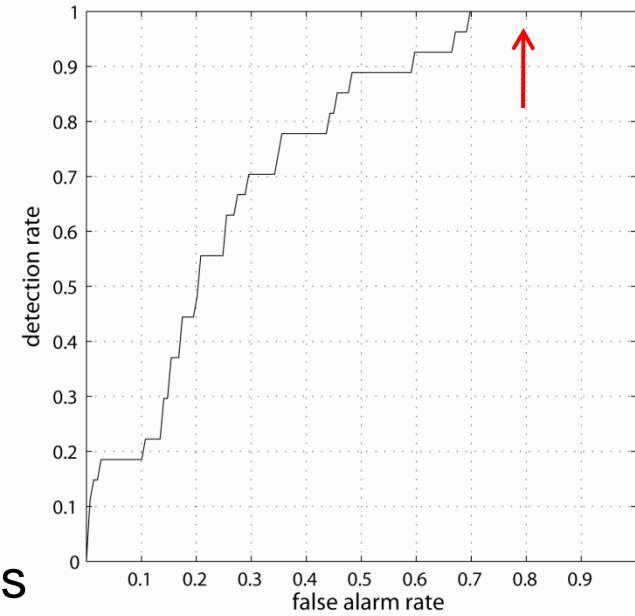
ROC curve

The ROC curve

Model



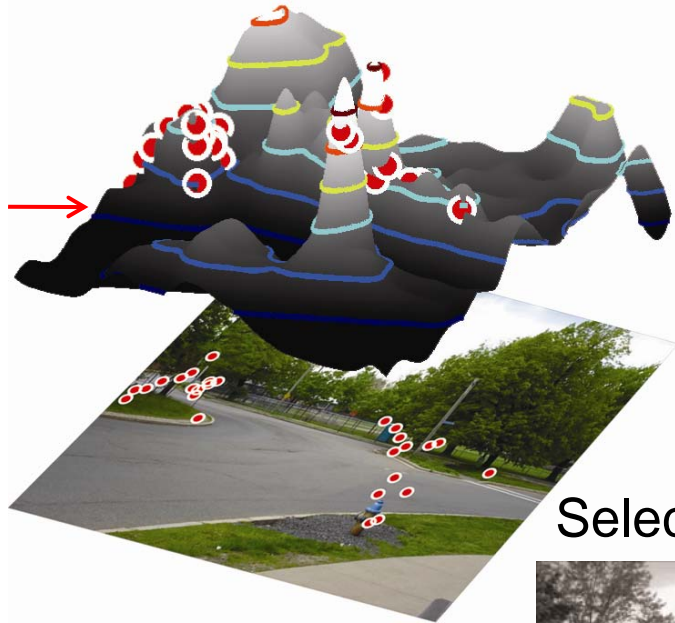
Selected image regions



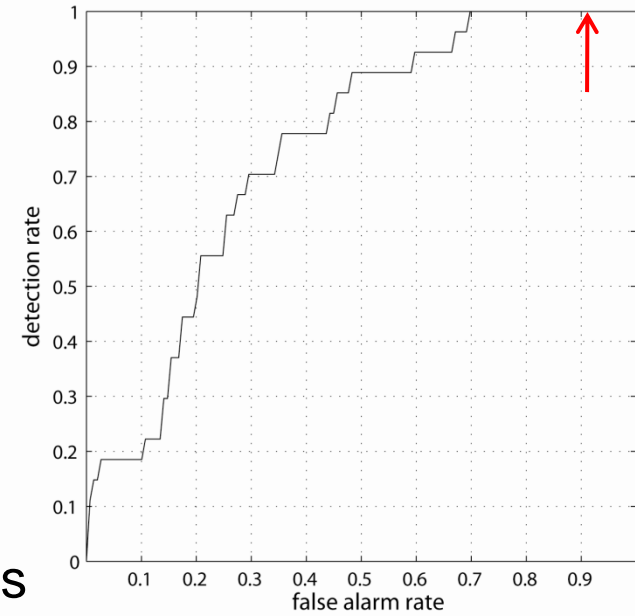
ROC curve

The ROC curve

Model



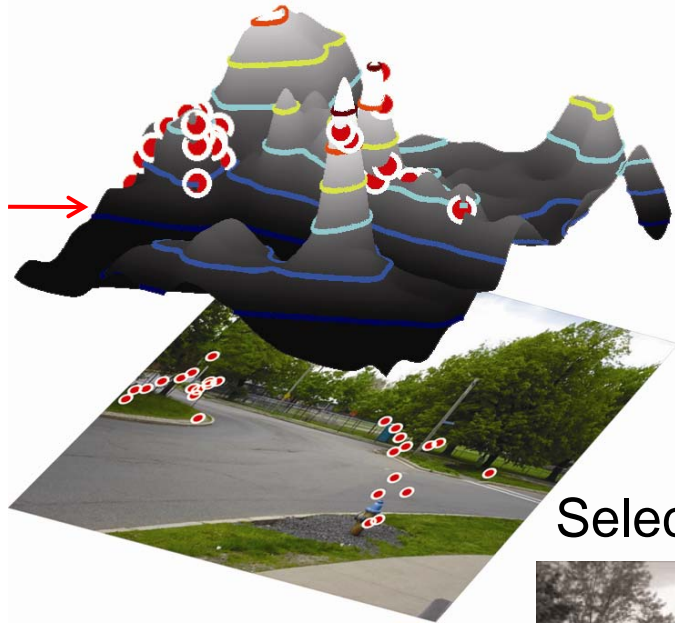
Selected image regions



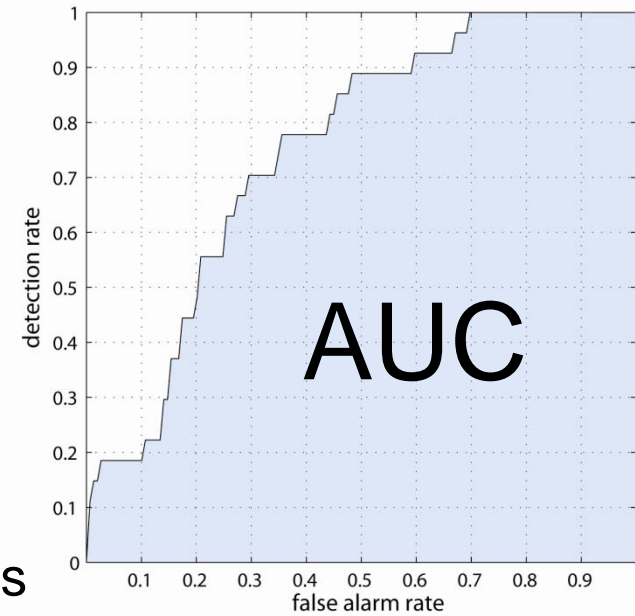
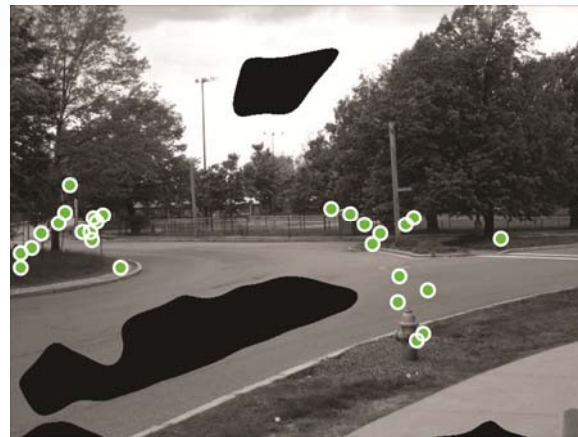
ROC curve

The ROC curve

Model

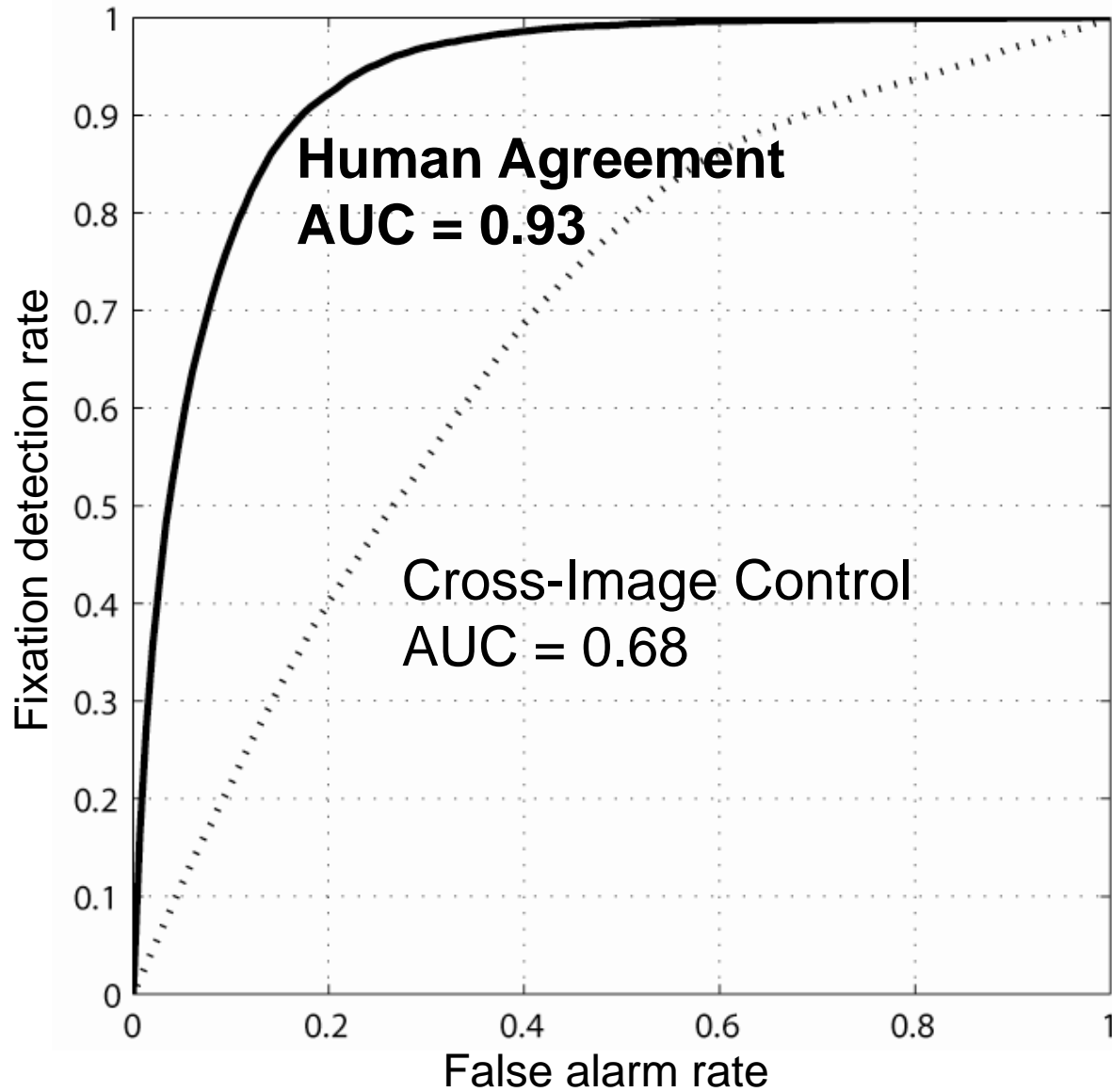


Selected image regions



ROC curve

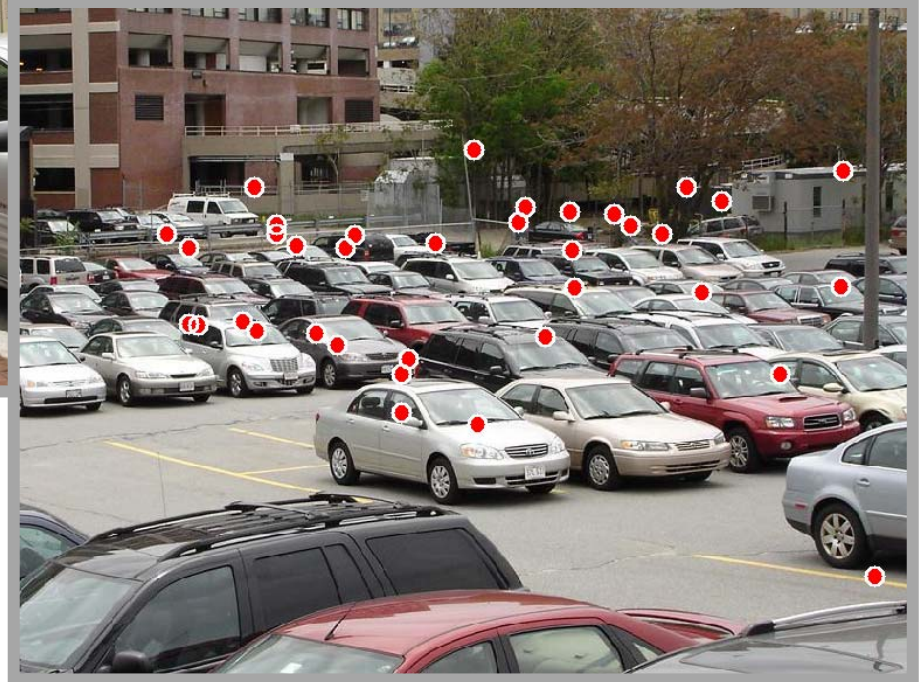
Human Agreement



Human agreement examples

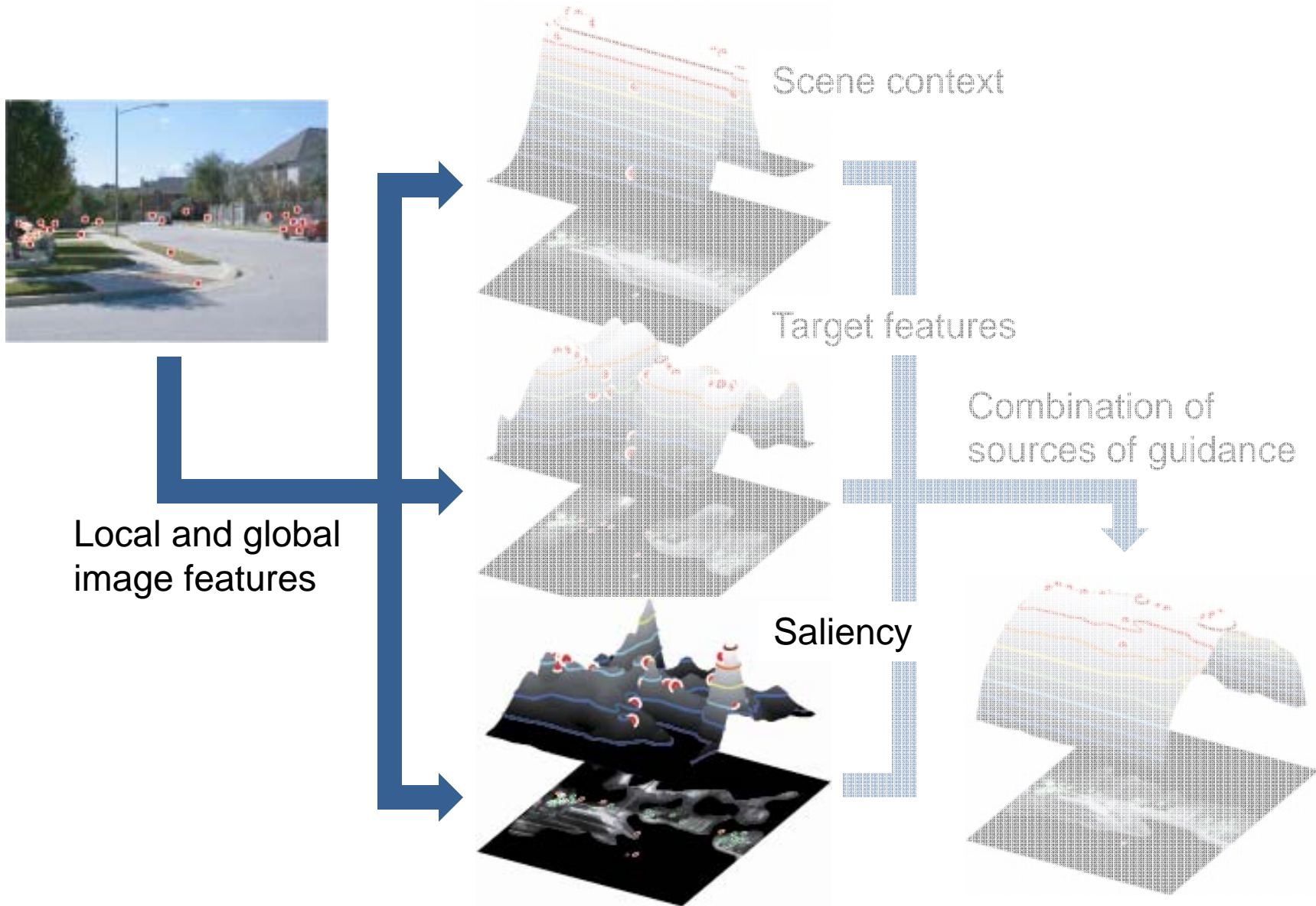


High inter-observer agreement

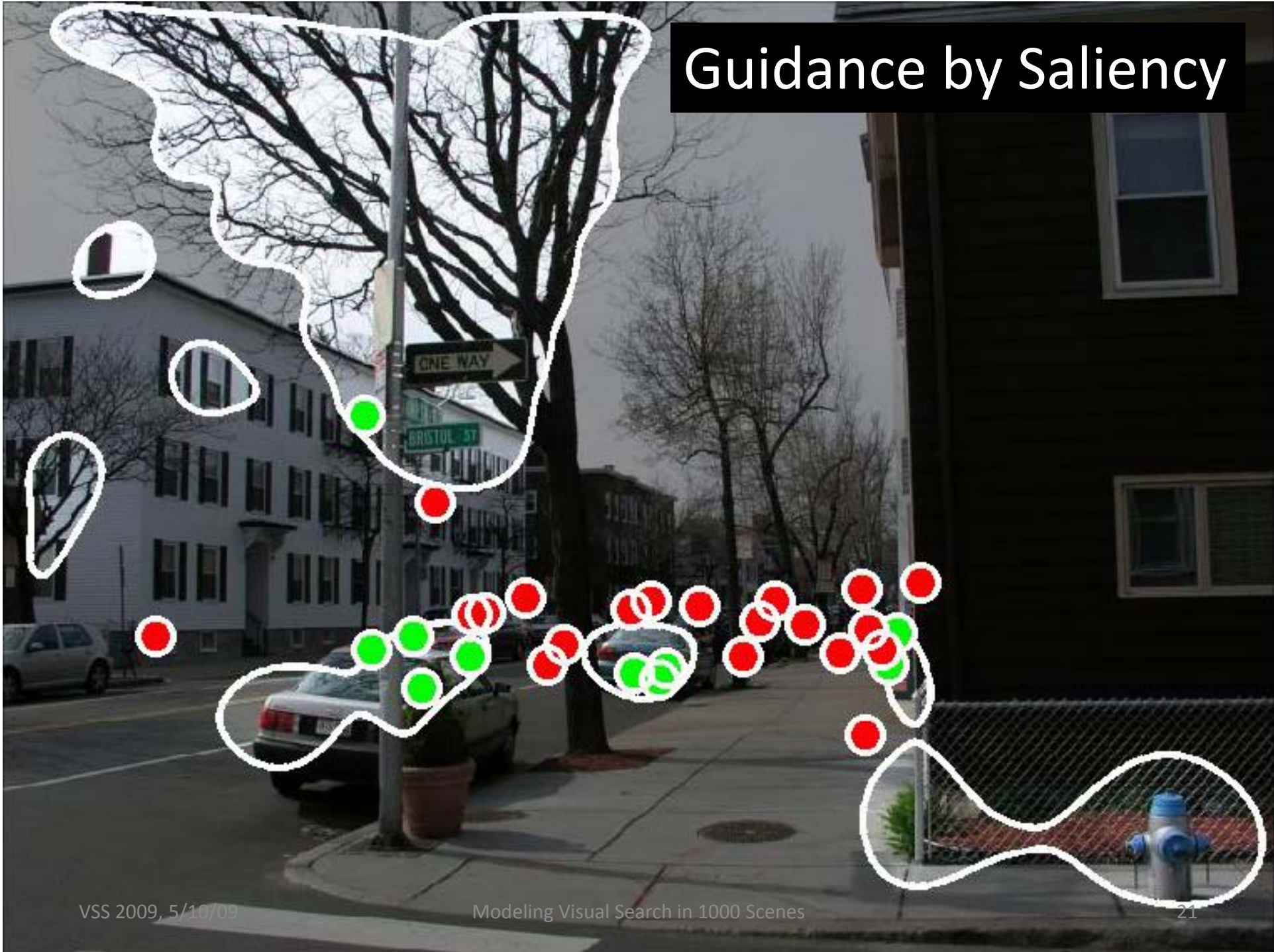


Low inter-observer agreement

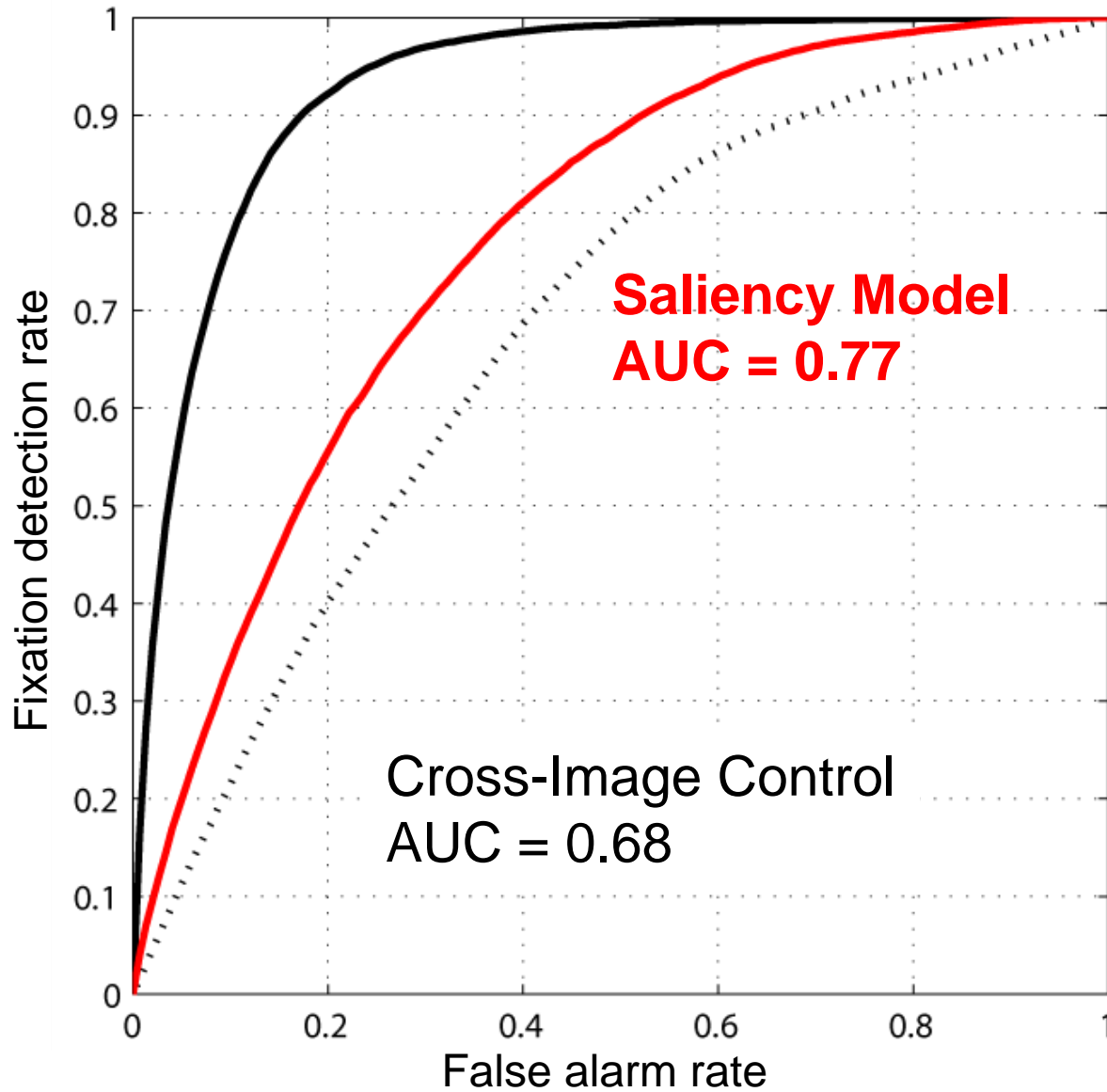
Overview of Model



Guidance by Saliency



Saliency Model

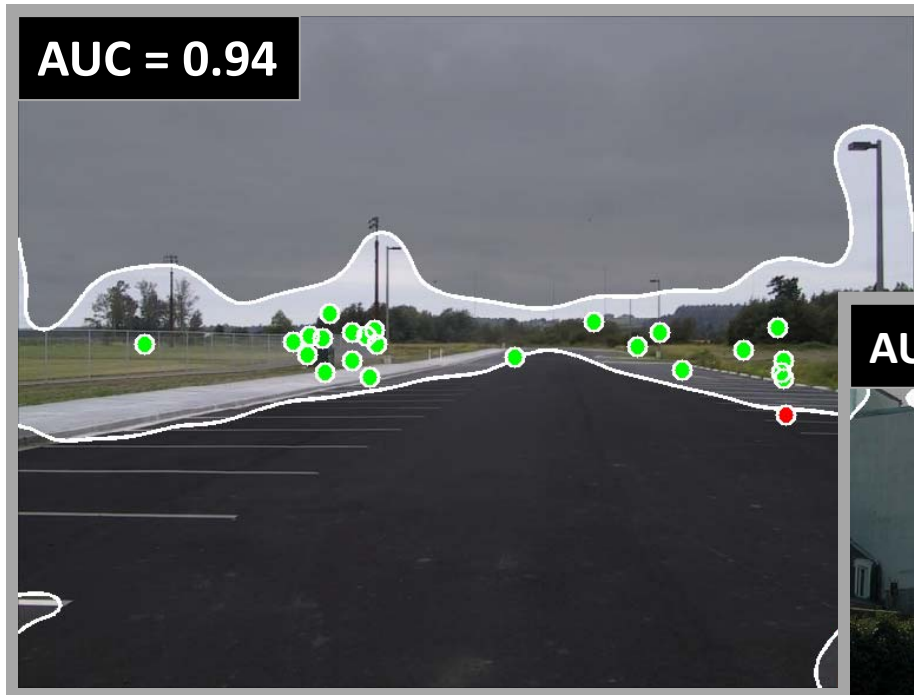


**Human Agreement
AUC = 0.93**

**Saliency Model
AUC = 0.77**

**Cross-Image Control
AUC = 0.68**

Saliency Model: Examples

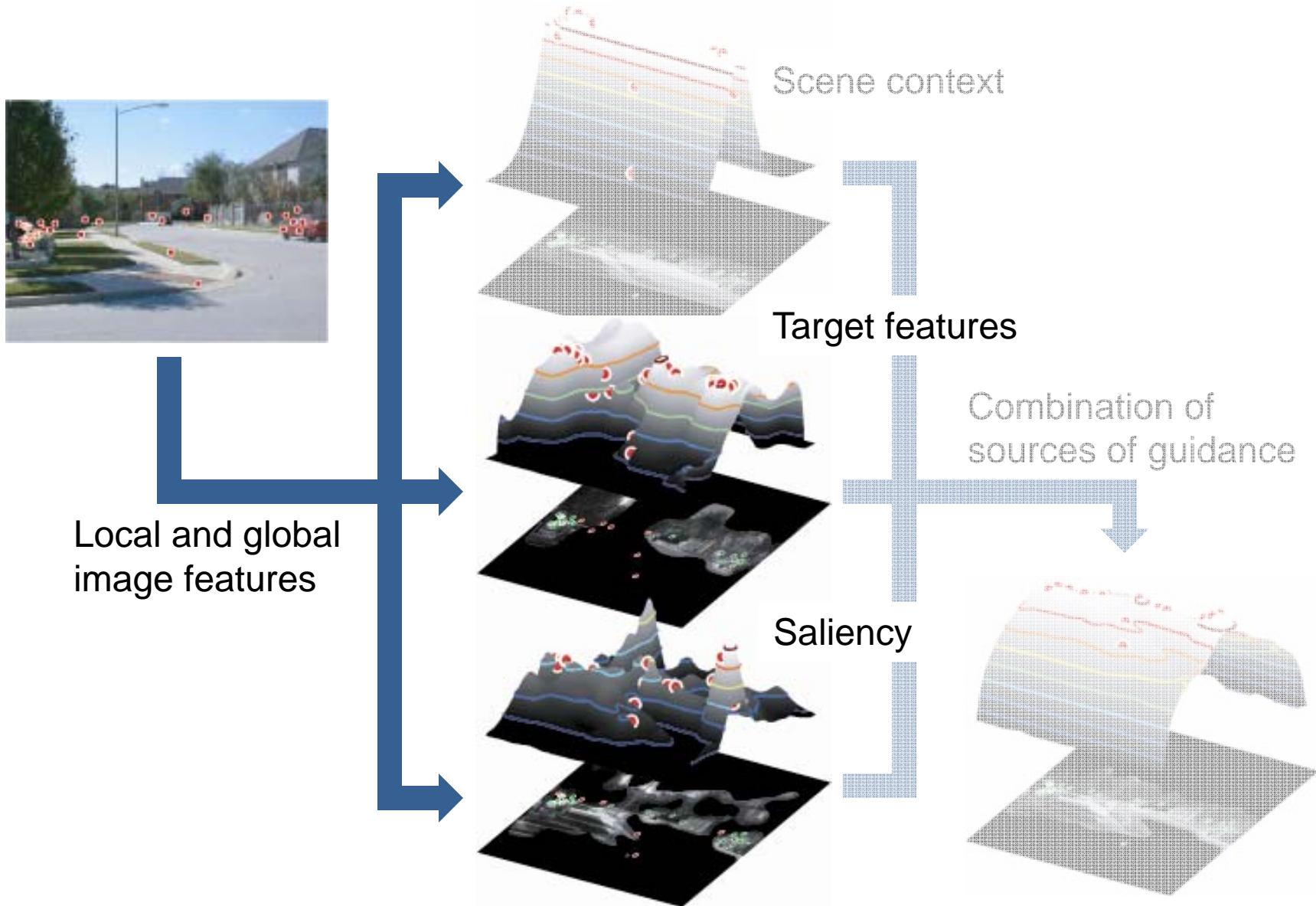


Best performance



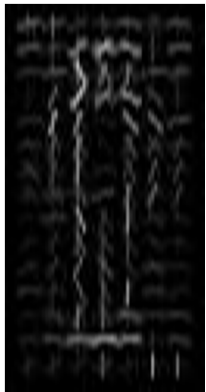
Worst performance

Overview of Model

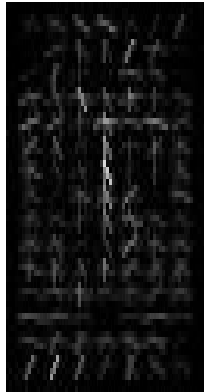


Pedestrian Detector

- Histograms of Oriented Gradients (HOG) detector by Dalal & Triggs



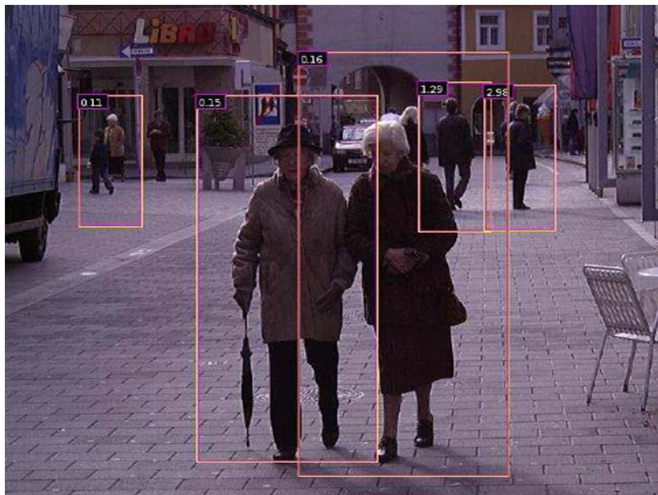
Positive features



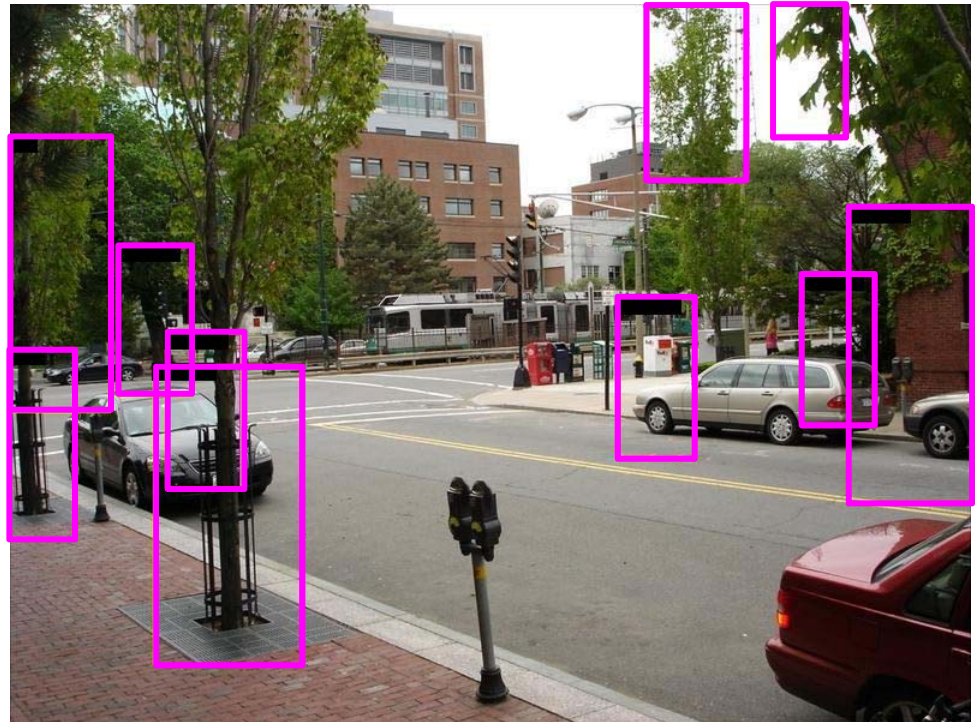
Negative features



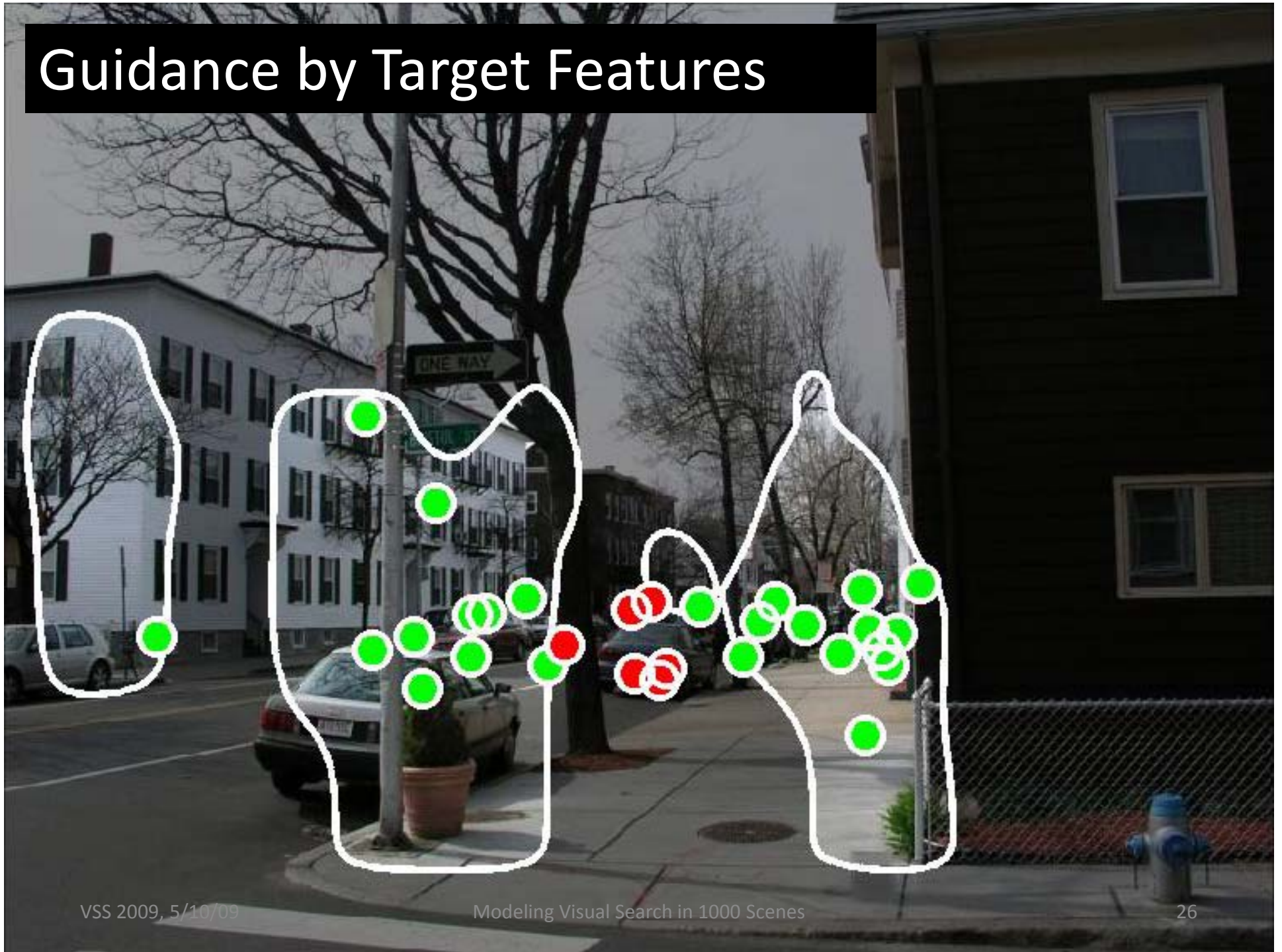
Average gradient



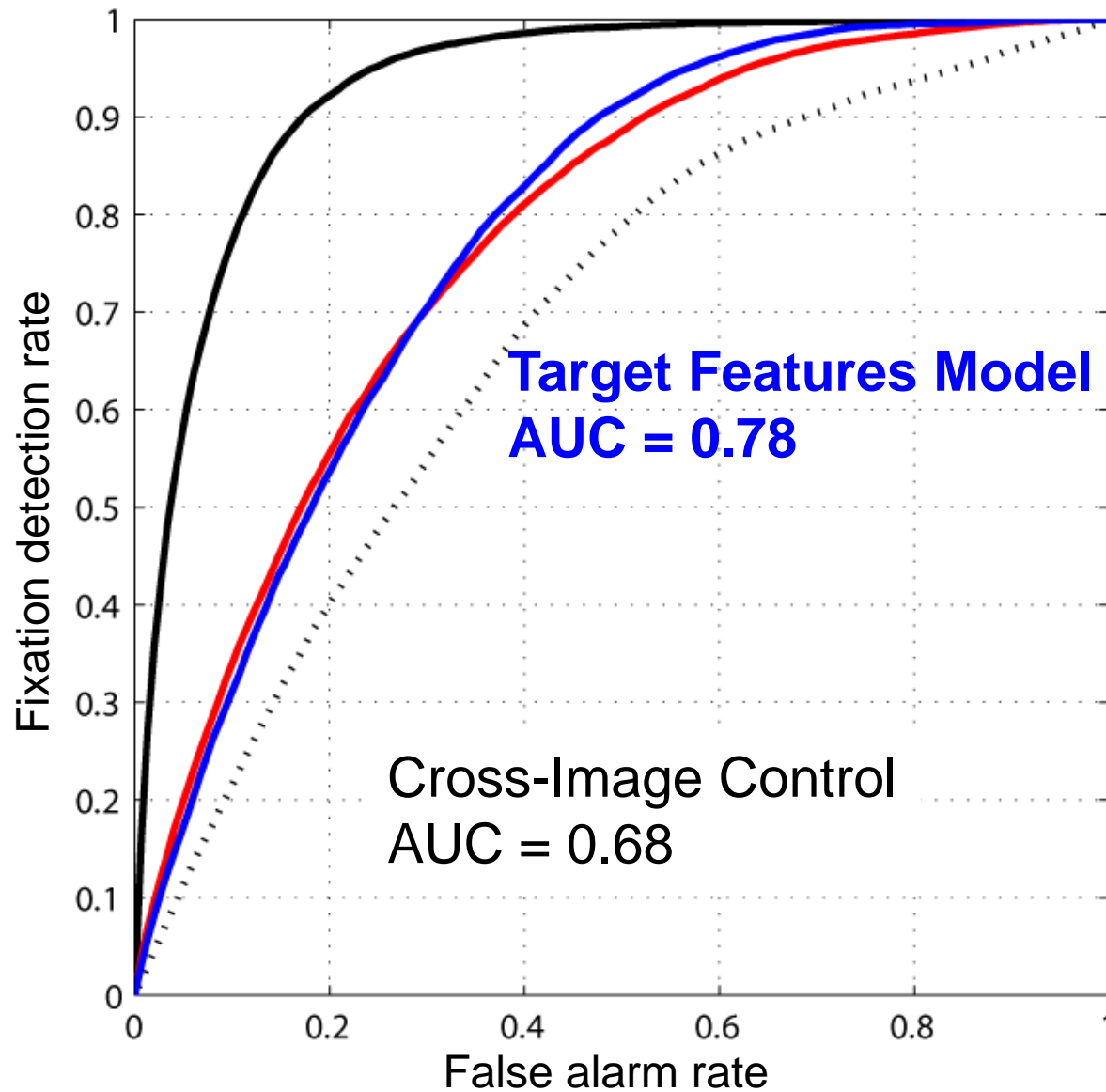
Dalal & Triggs, 2005 CVPR



Guidance by Target Features



Target Features Model



**Human Agreement
AUC = 0.93**

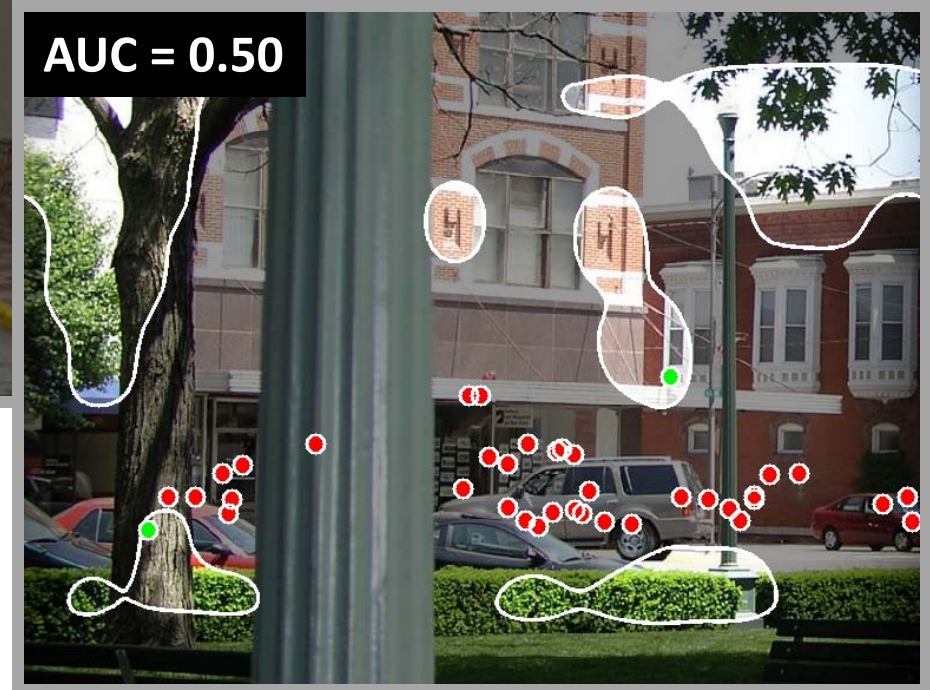
**Target Features Model
AUC = 0.78**

Cross-Image Control
AUC = 0.68

Target Features Model: Examples

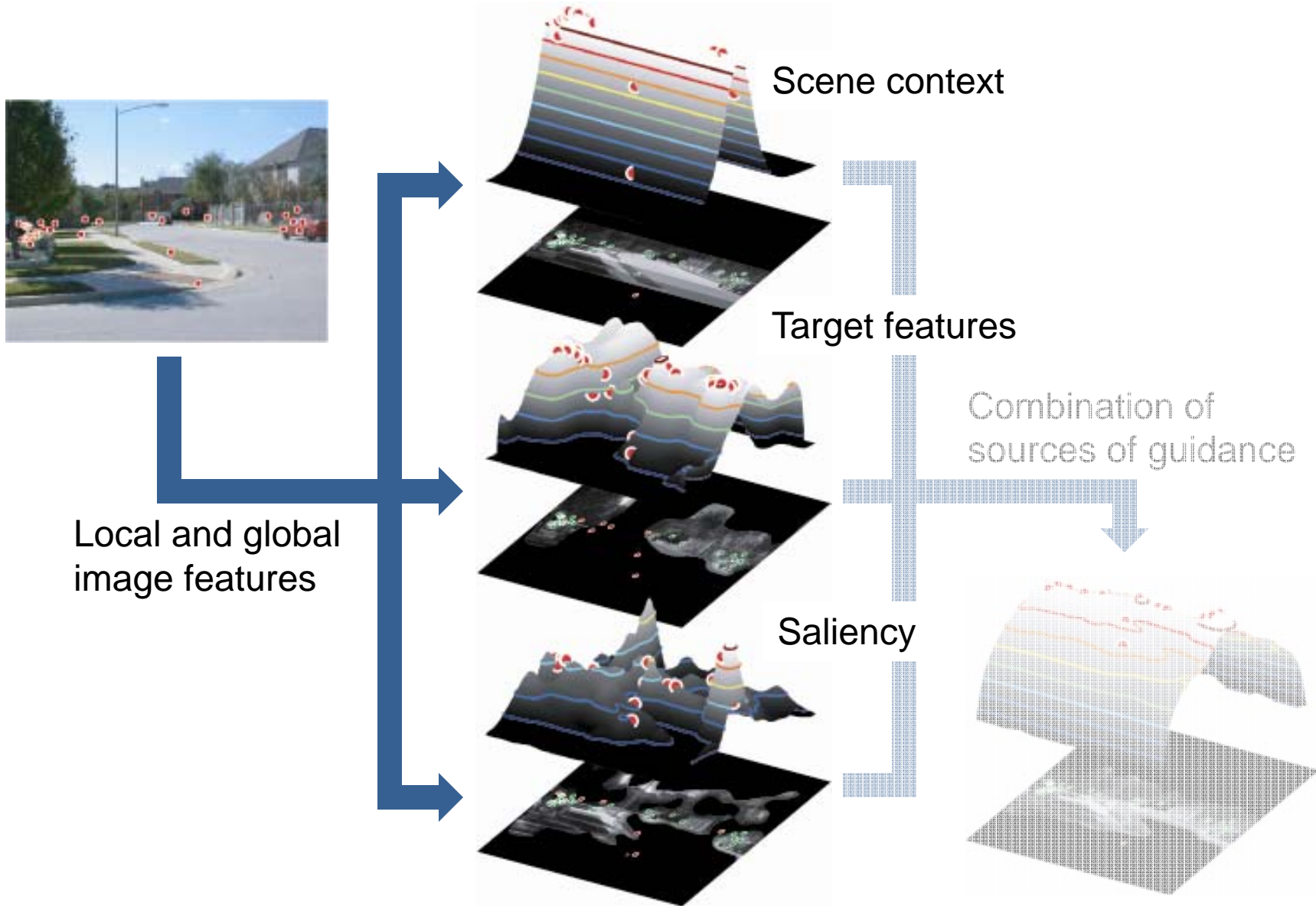


Best performance



Worst performance

Overview of Model



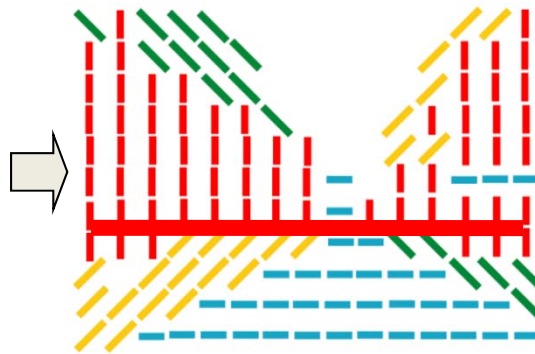
What is the context region for pedestrians?



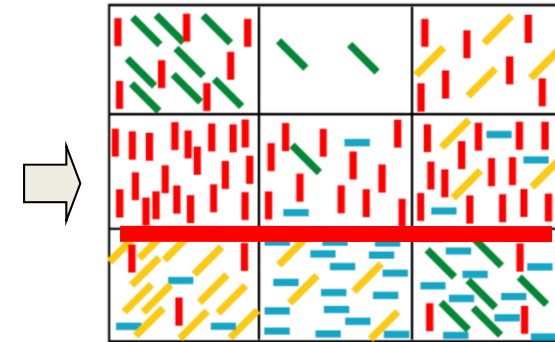
Scene Context Model



Training image
(contains a pedestrian)



Orientations at various
spatial scales

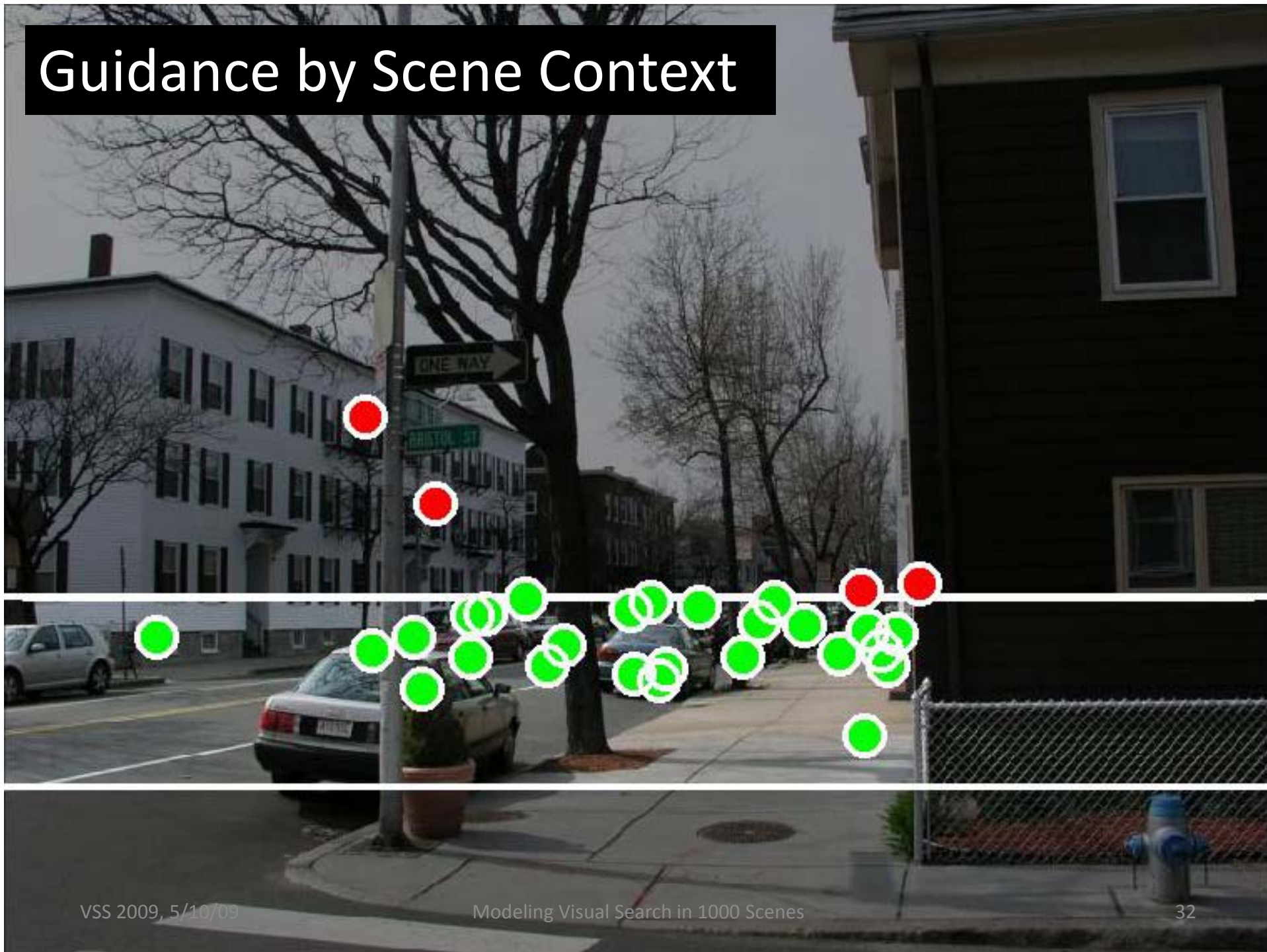


Scene "gist" + position
of pedestrian

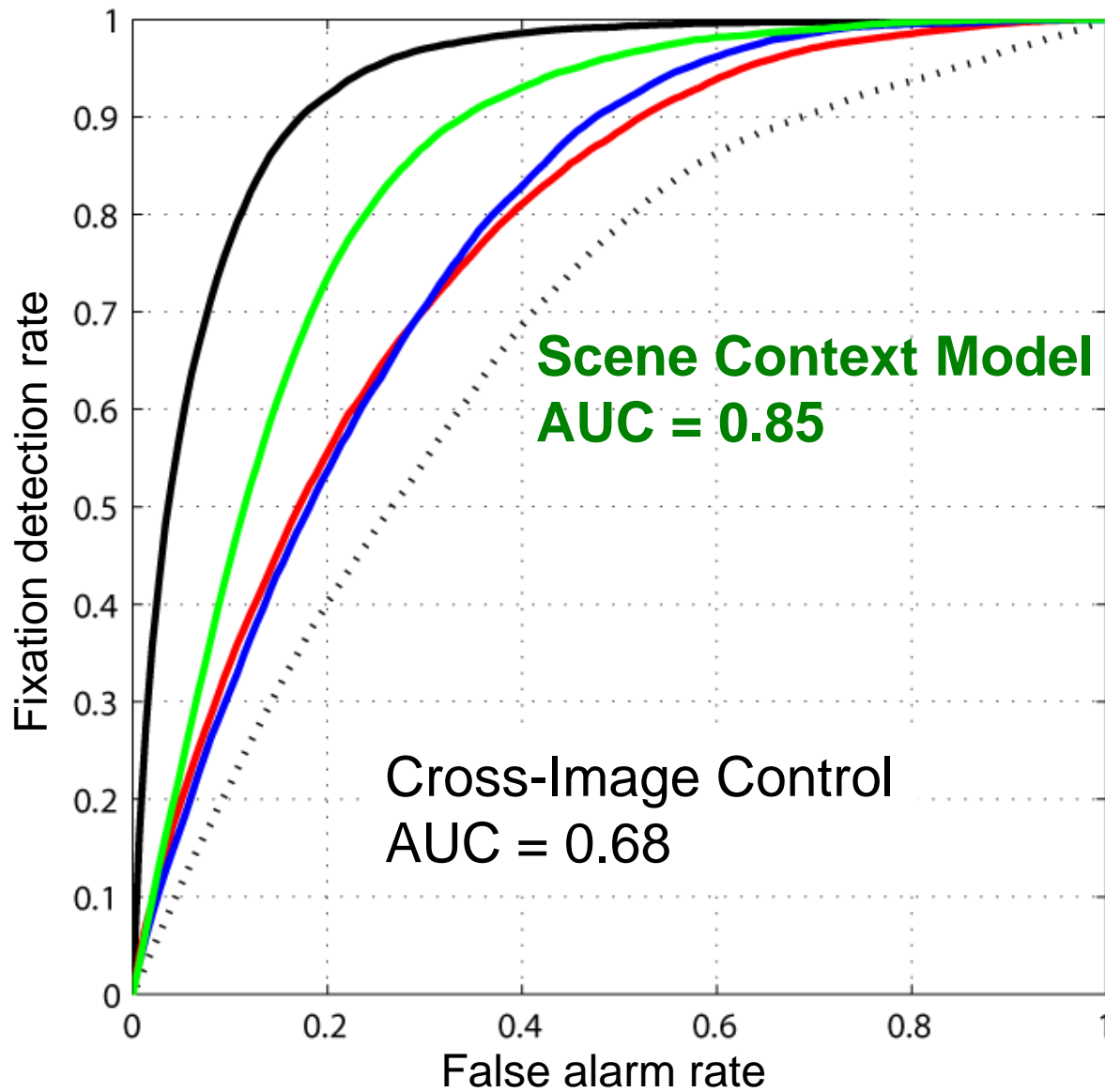


Oliva & Torralba, 2001

Guidance by Scene Context



Scene Context Model



Human Agreement
AUC = 0.93

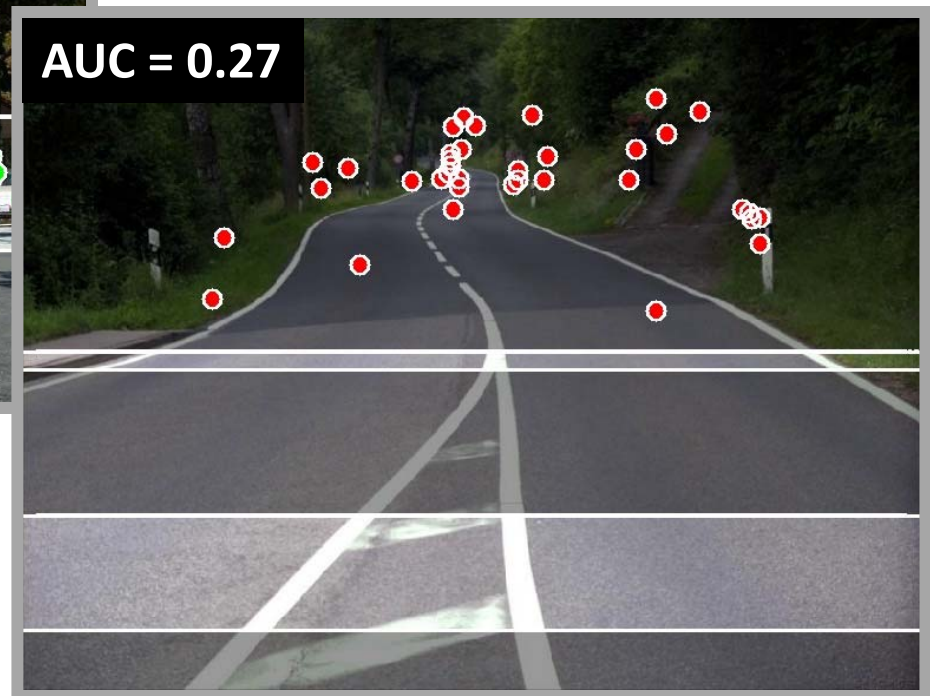
Scene Context Model
AUC = 0.85

Cross-Image Control
AUC = 0.68

Scene Context Model: Examples

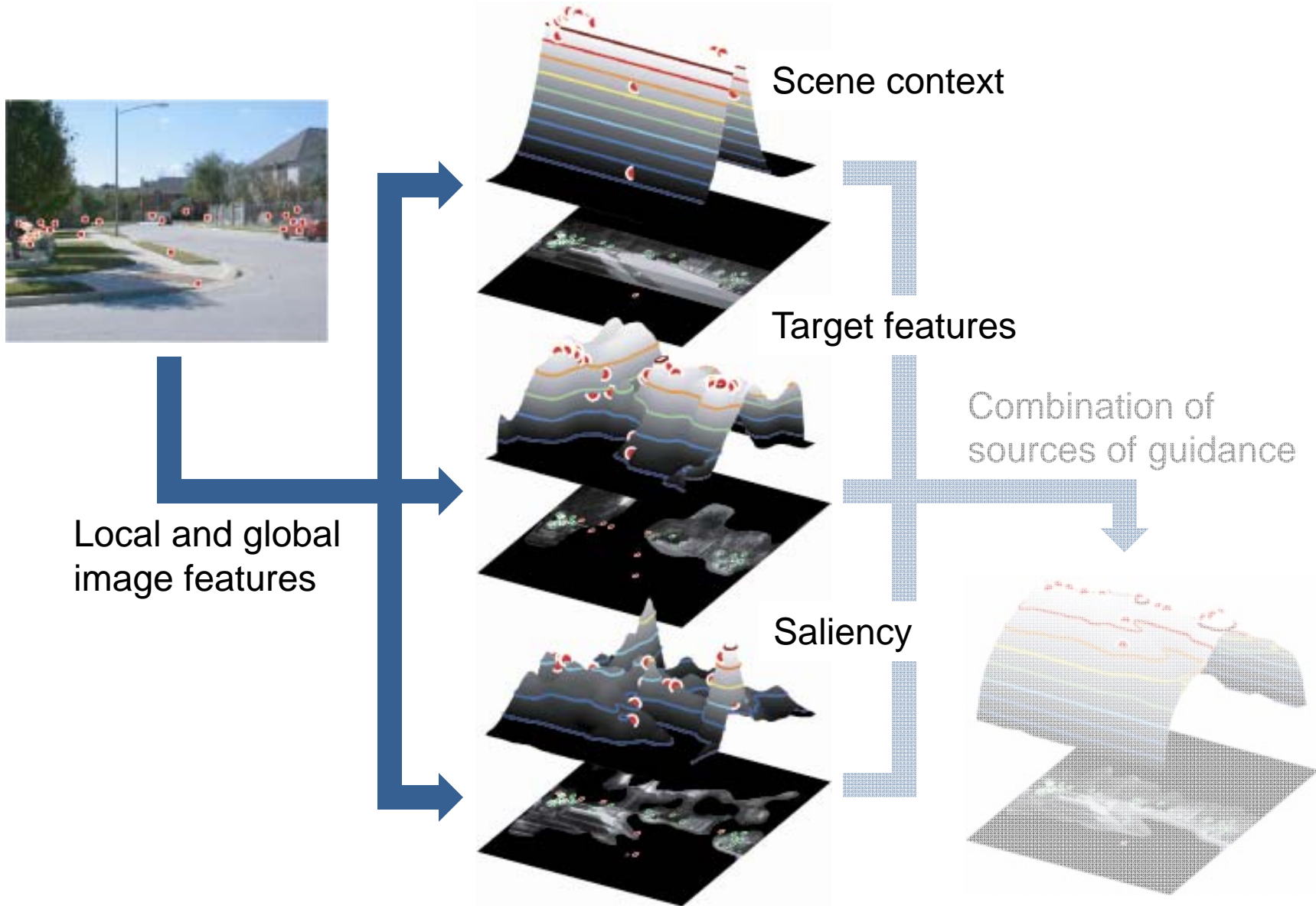


Best performance

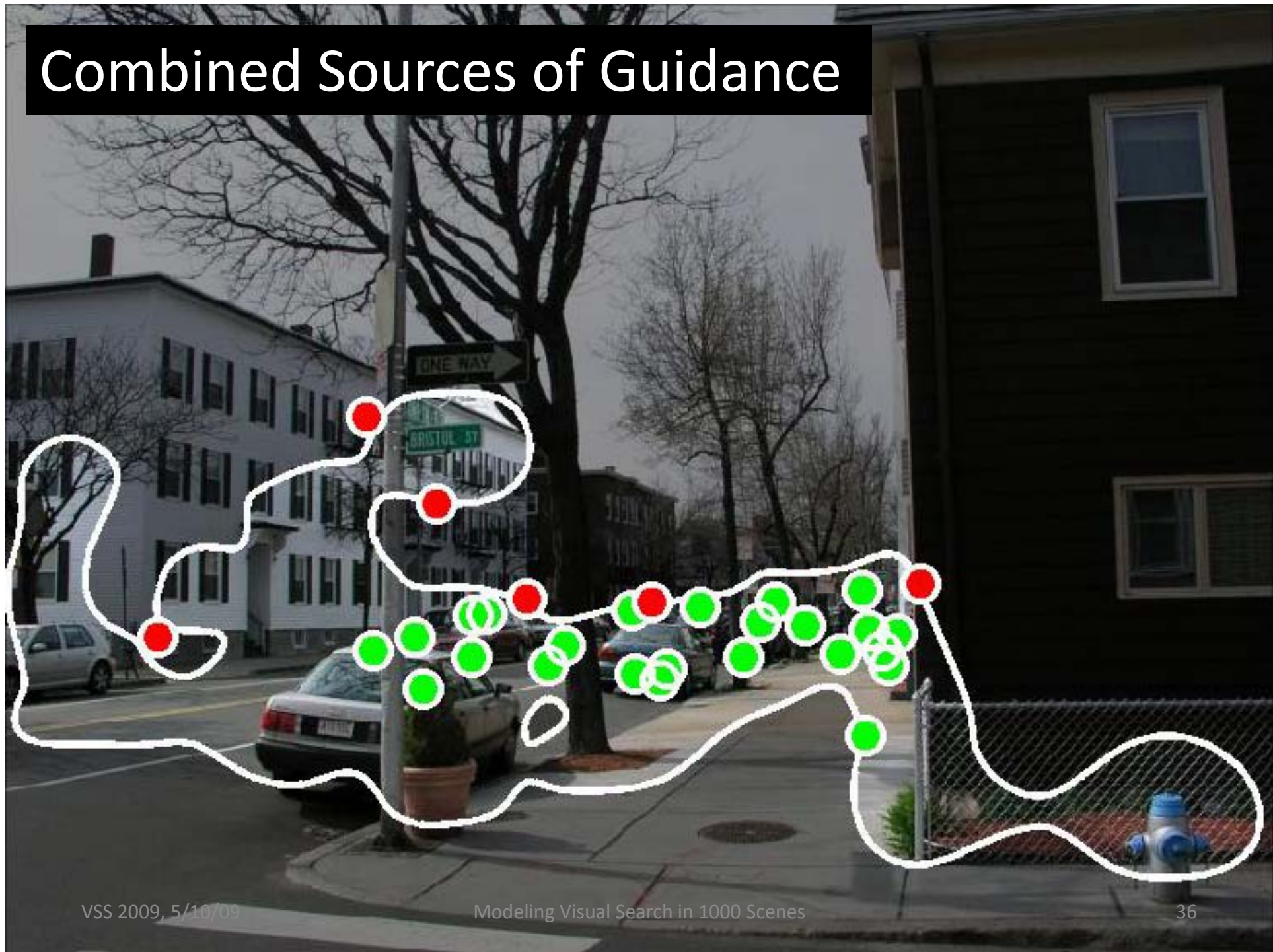


Worst performance

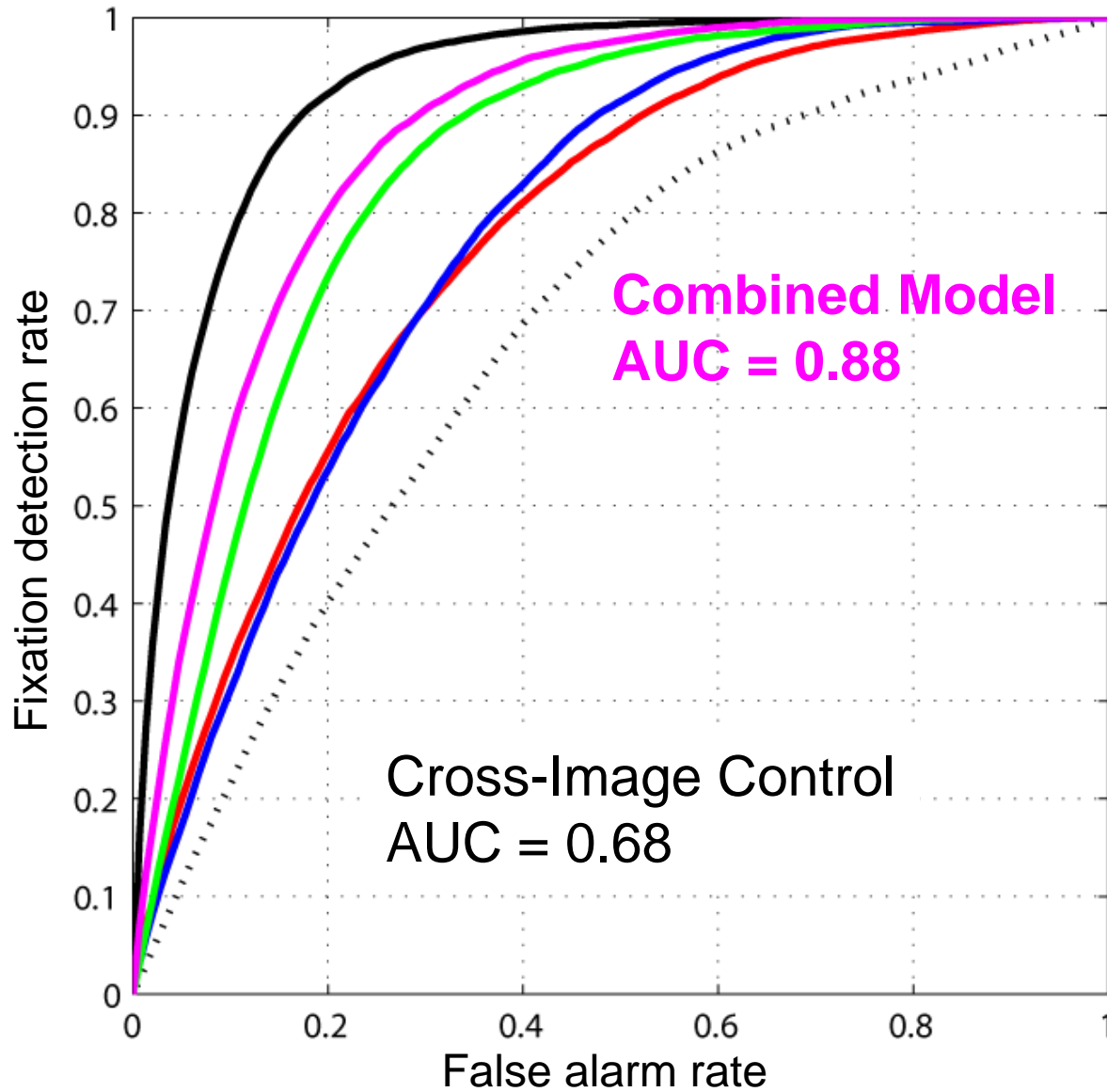
Overview of Model



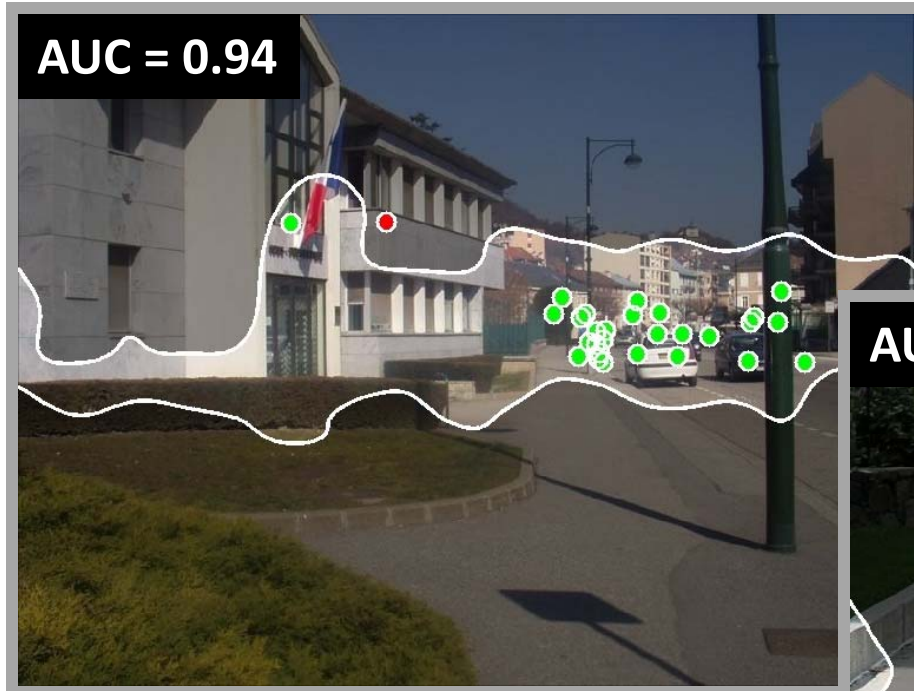
Combined Sources of Guidance



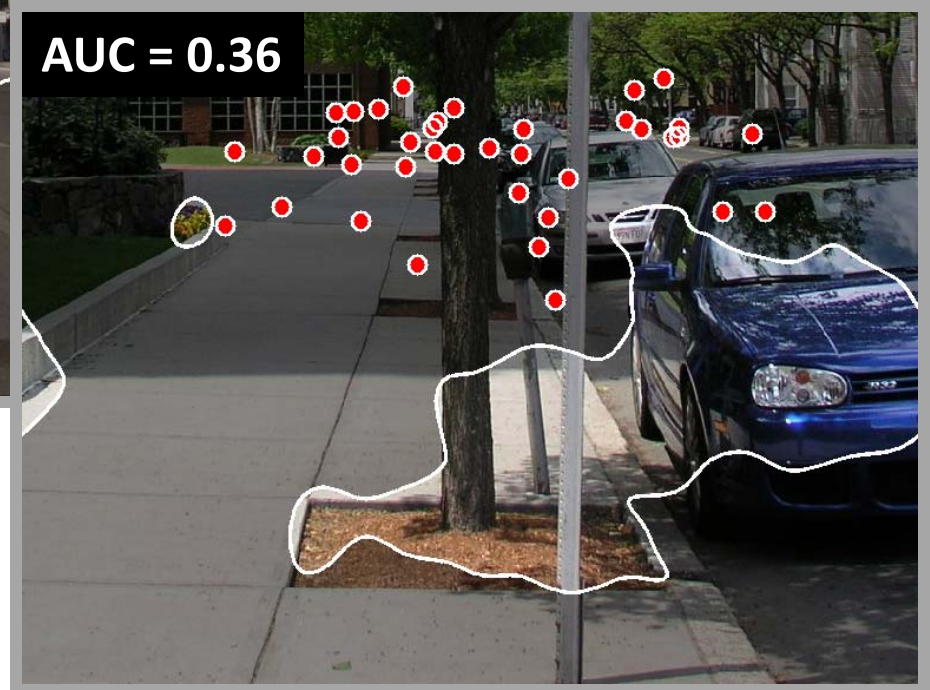
Combined Model



Combined Model: Examples

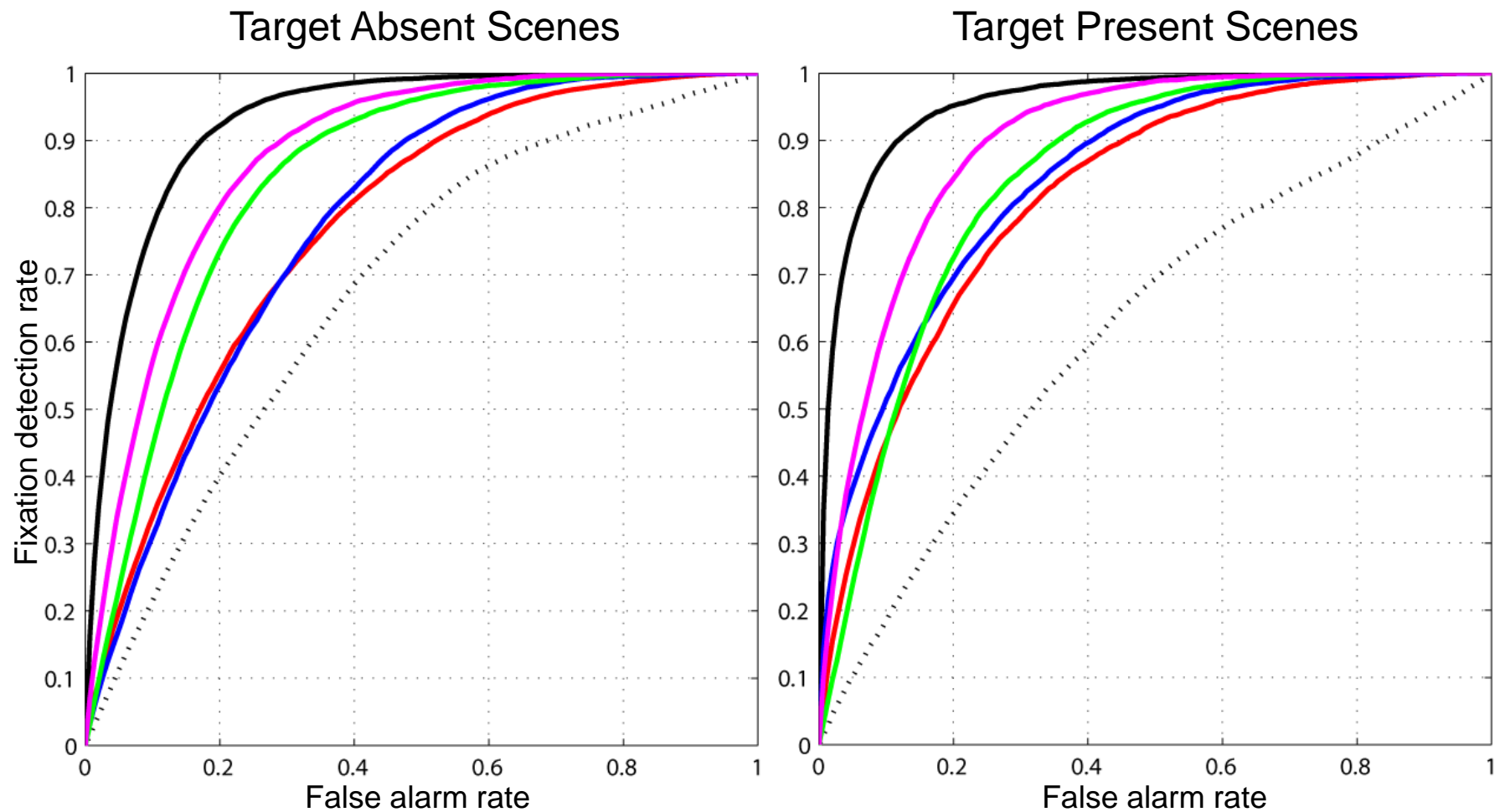


Best performance



Worst performance

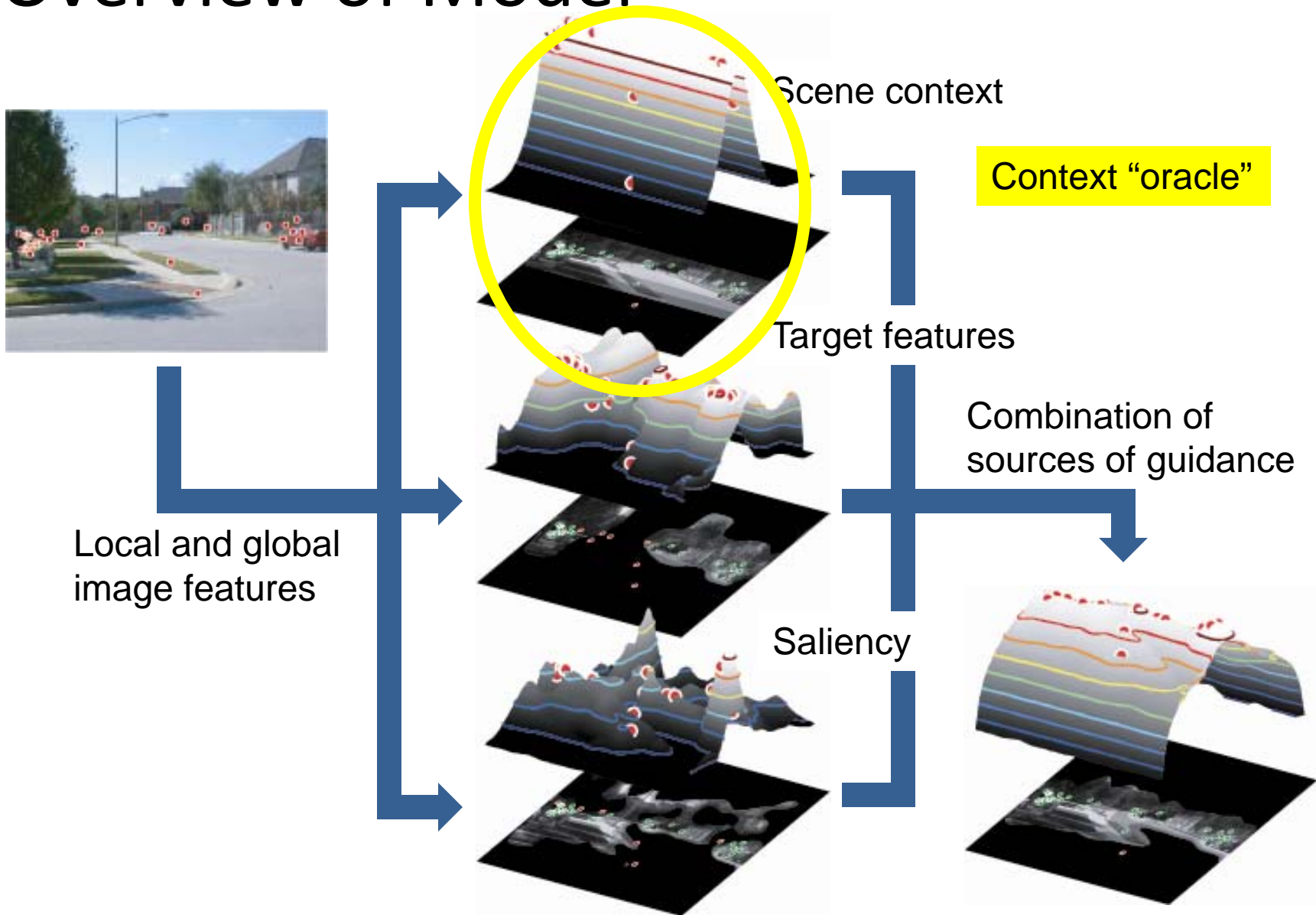
Target Absent vs. Target Present



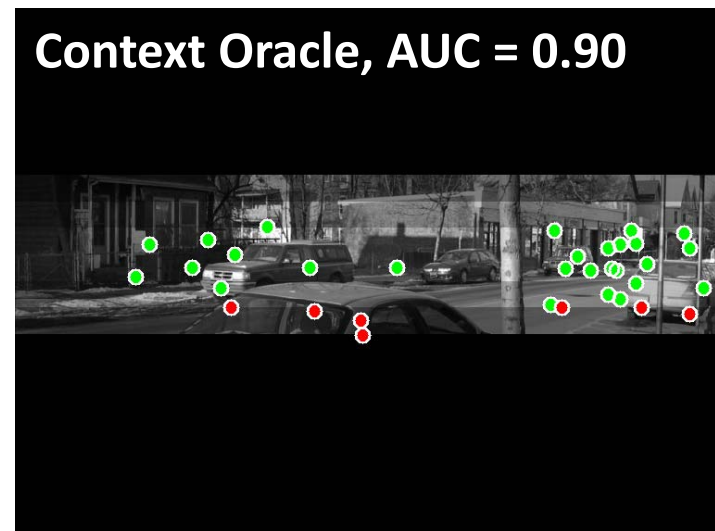
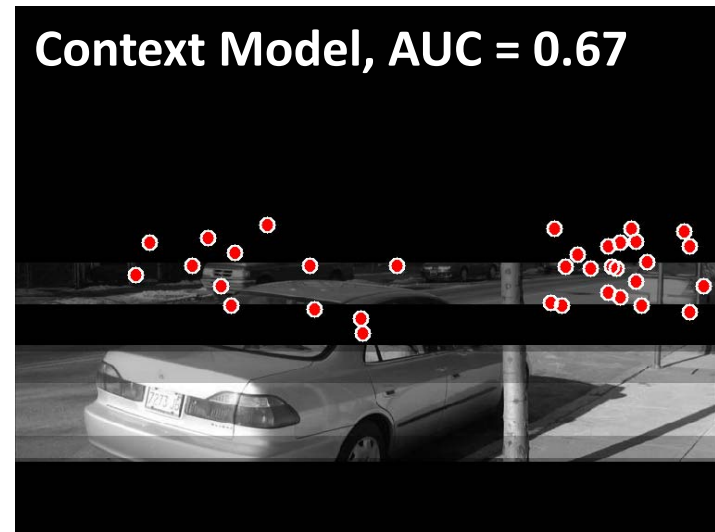
Summary of results

- Combined model accounts for 94% of human agreement in search fixations
- Scene context gives the best prediction of human search fixations in this task
- How to get that last 6%?

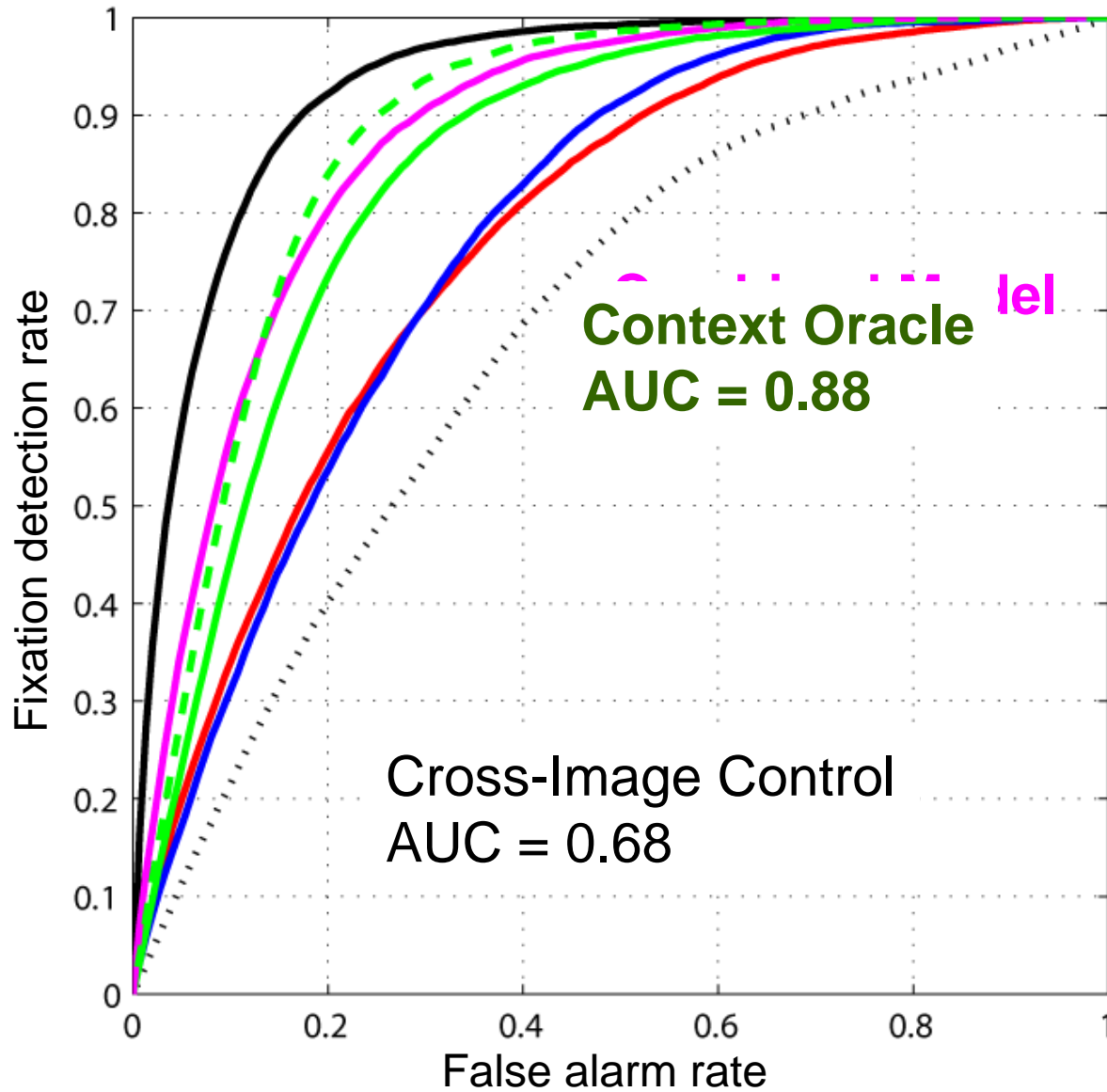
Overview of Model



“Context Oracle” Implementation



Context Oracle

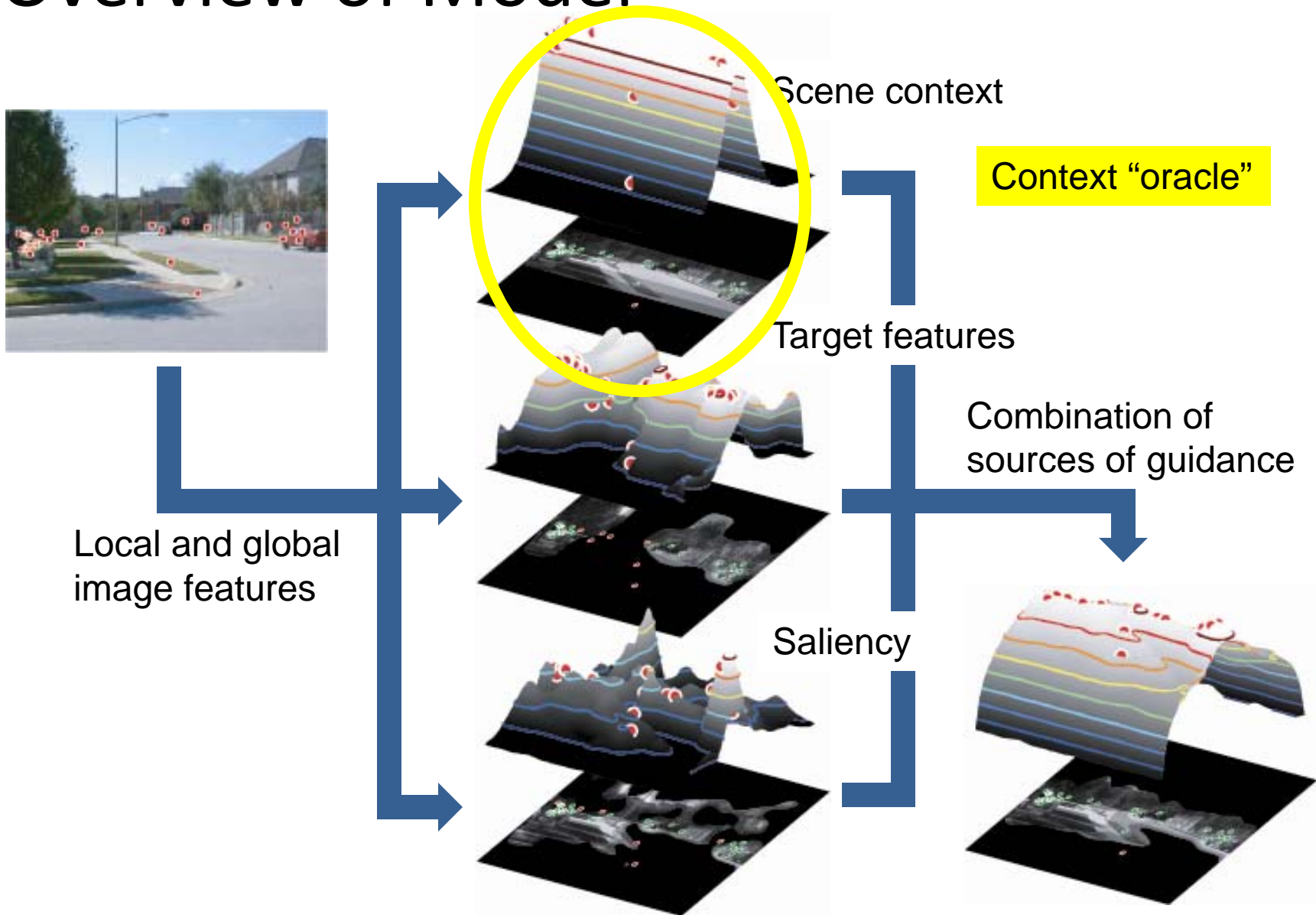


Human Agreement
AUC = 0.93

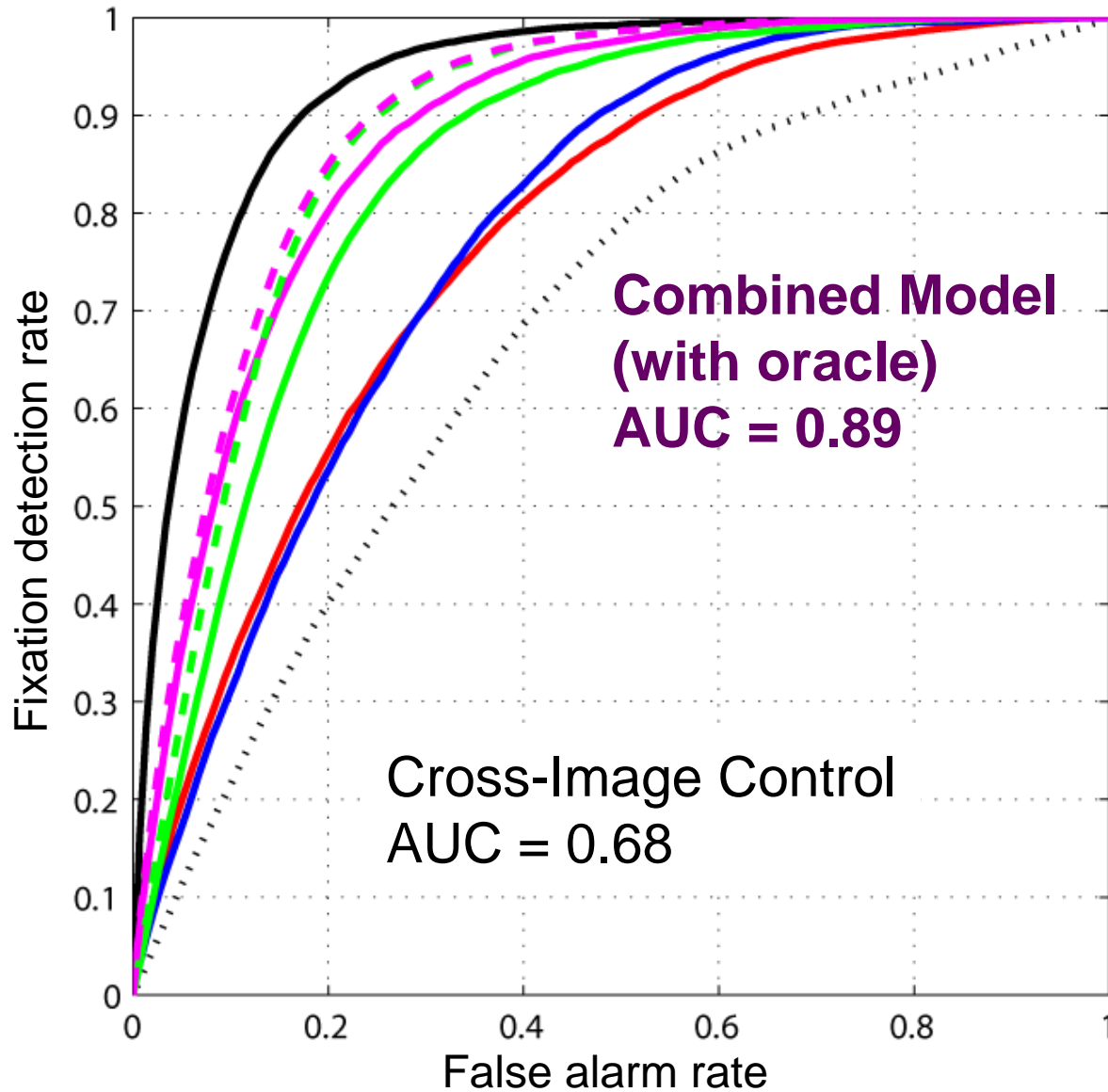
Context Oracle
AUC = 0.88

Cross-Image Control
AUC = 0.68

Overview of Model



Combined Model with Oracle



Human Agreement
AUC = 0.93

Combined Model
(computational)
AUC = 0.88

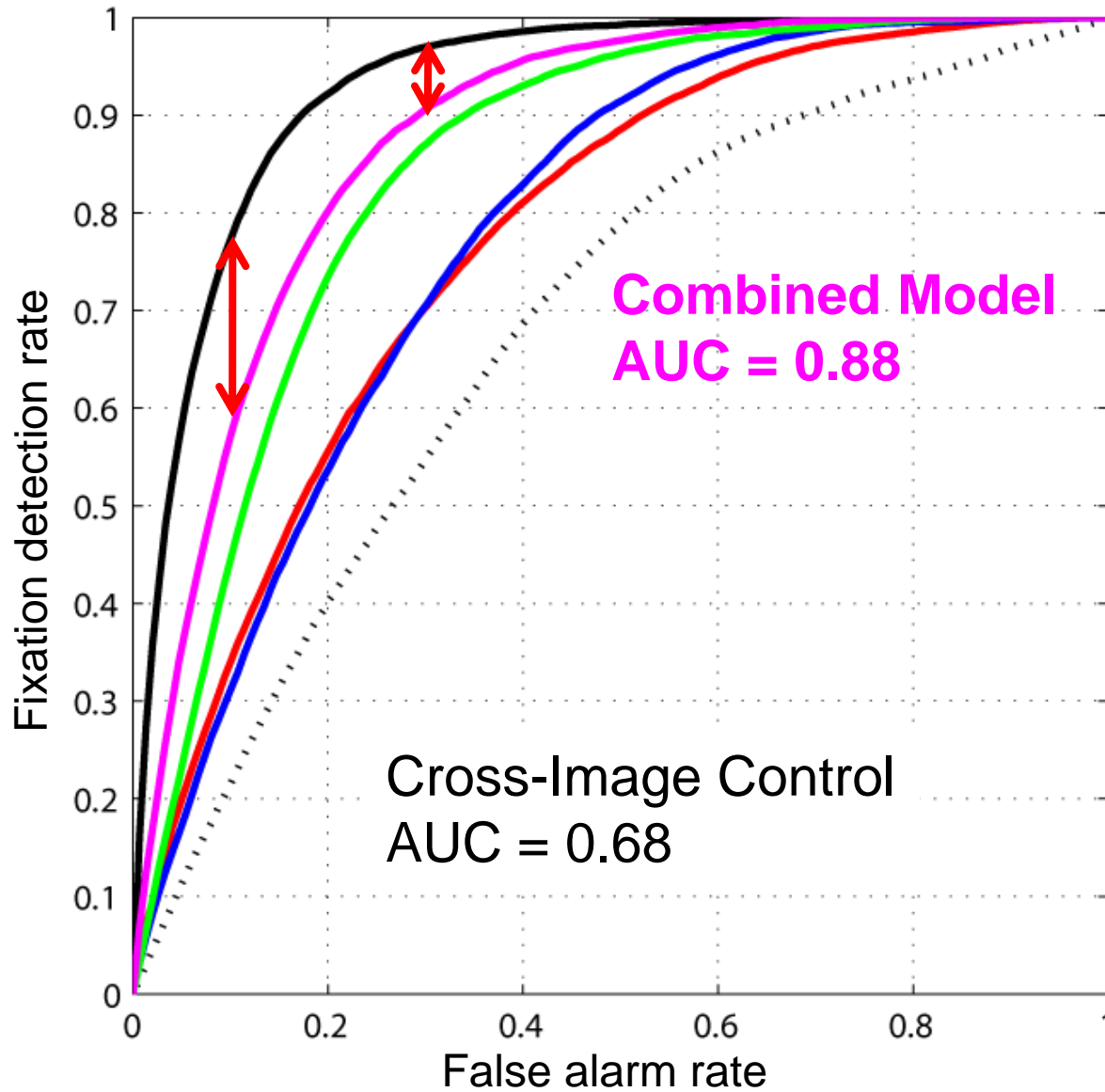
Combined Model
(with oracle)
AUC = 0.89

Cross-Image Control
AUC = 0.68

Summary of results

- Combined model accounts for 94% of human agreement in search fixations
- Context predicts human fixations better than saliency or target features in this search task
- How to get that last 6%?
 - Context “oracle”?
 - Improves performance to 95% of human agreement
 - Something else?

What's next?

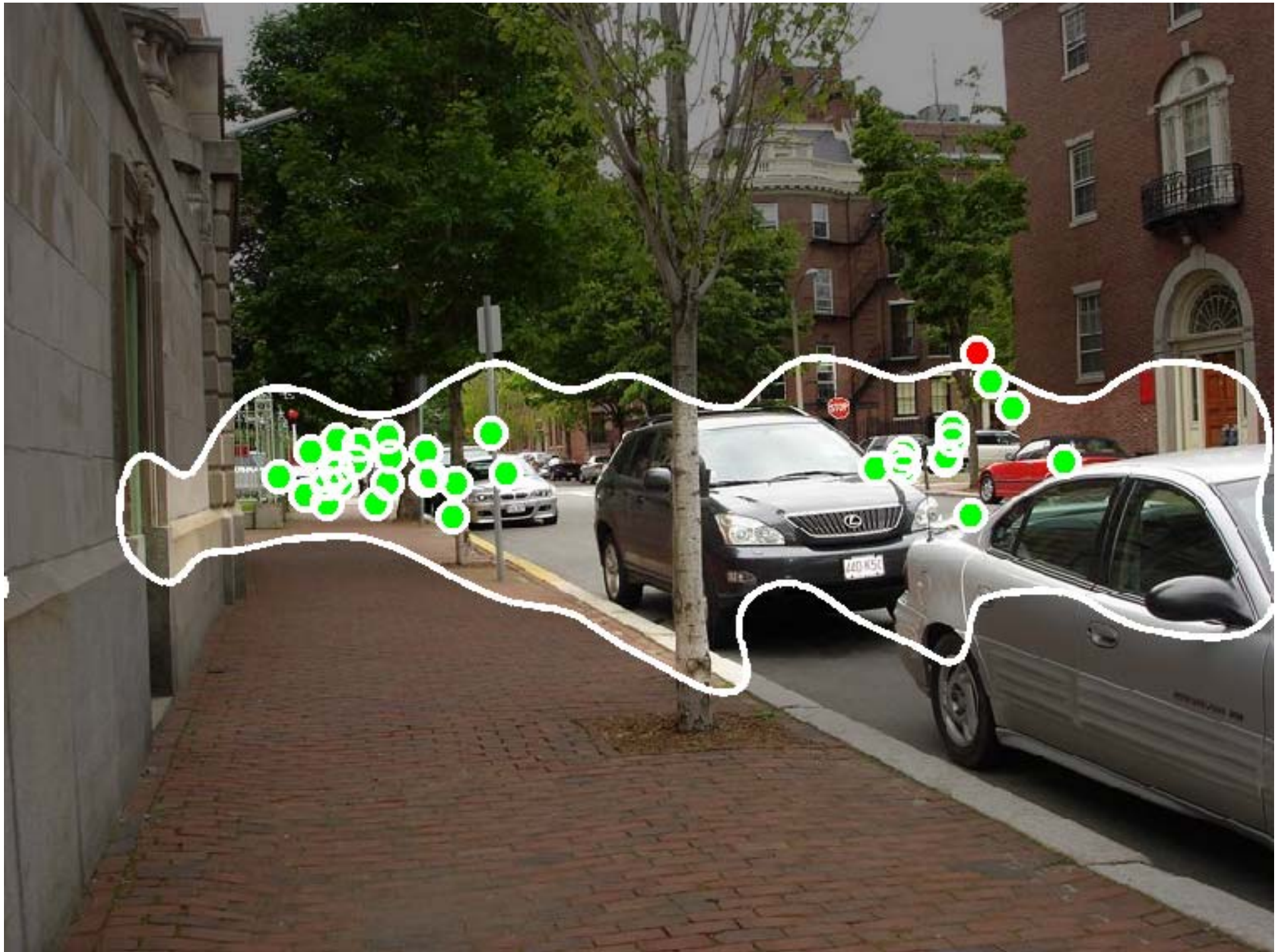


**Human Agreement
AUC = 0.93**

**Combined Model
AUC = 0.88**

**Cross-Image Control
AUC = 0.68**







Acknowledgements



Barbara Hidalgo-Sotelo



Antonio Torralba



Aude Oliva

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A. & Oliva, A. Modeling Search for People in 900 Scenes: A combined source model of eye guidance. *Visual Cognition*, in press.

Funded by a Singleton graduate research fellowship to KE, an NSF graduate research fellowship to BHS, and NSF CAREER awards to AT and AO.