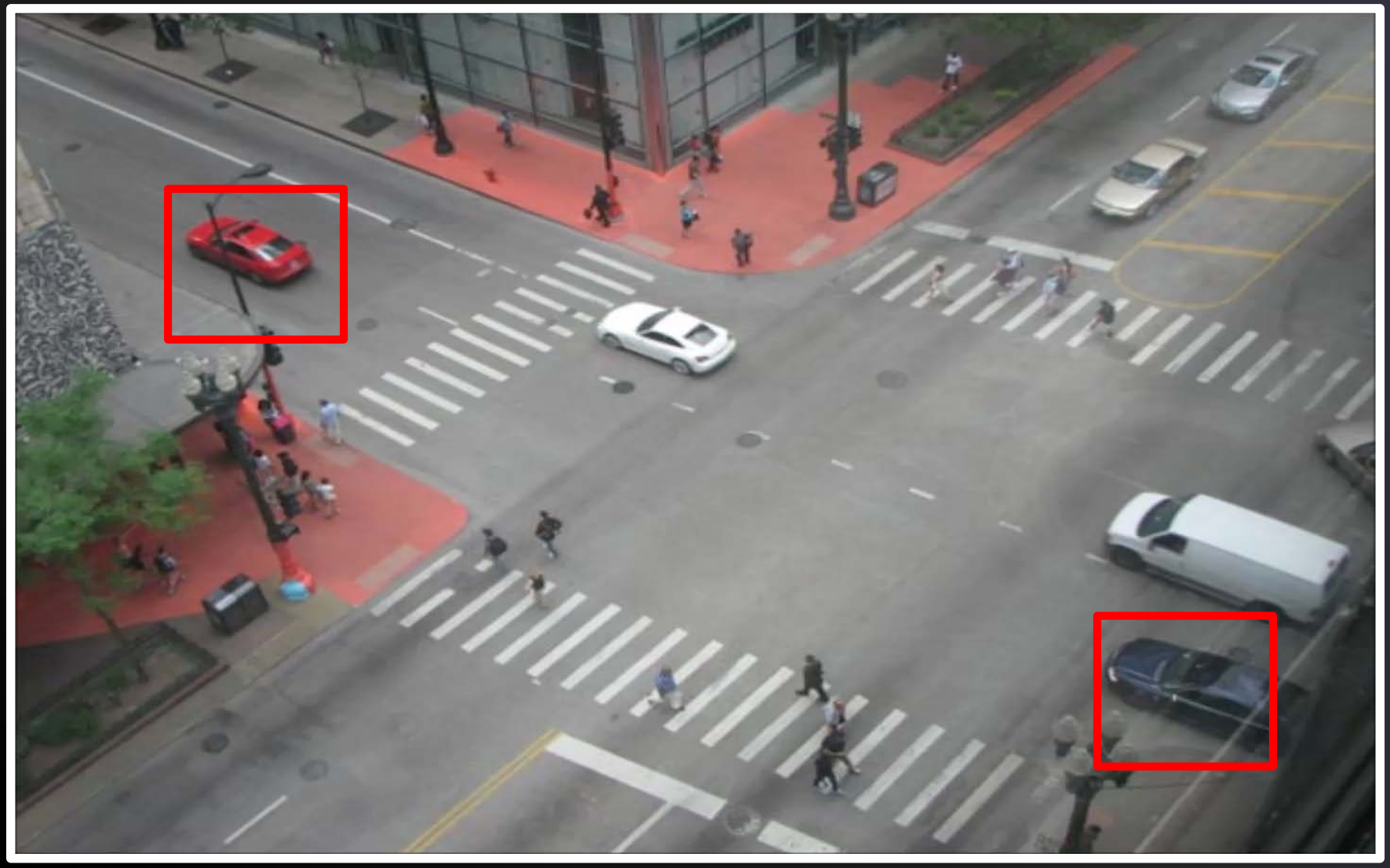
Patch to the Future: Unsupervised Visual Prediction

Jacob Walker, Abhinav Gupta, Martial Hebert
The Robotics Institute
Carnegie Mellon University

Visual Prediction



Goal

Both the *what* and the *how*

Goal

Both the *what* and the *how*

Goal

Both the *what* and the *how*

Goal

Both the *what* and the *how*

Goal

Both the *what* and the *how*

Goal

Both the *what* and the *how*

Background

Data-Driven

Yuen et al. 2010

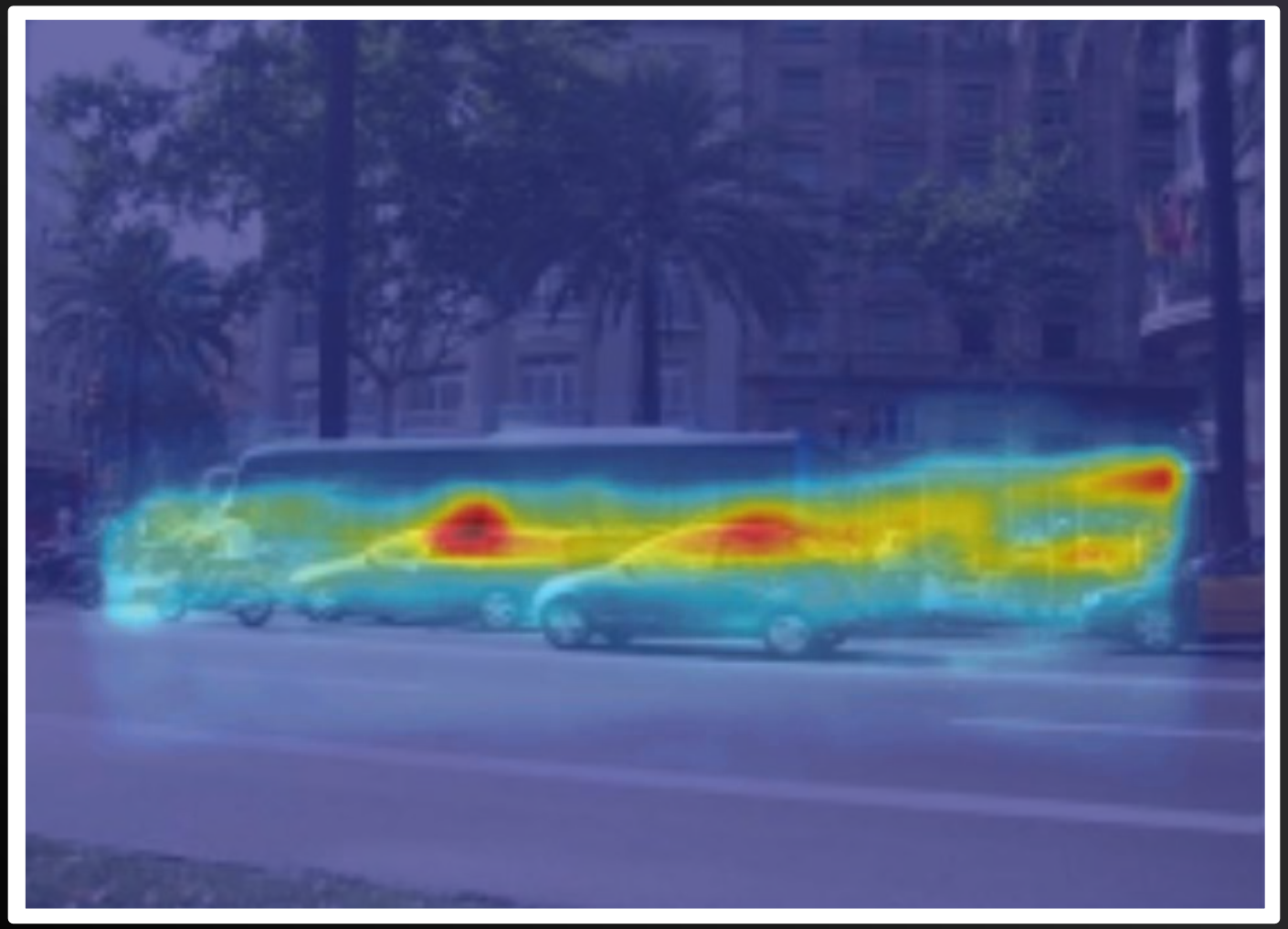
Background

Data-Driven

Yuen et al. 2010

Background

Data-Driven



Yuen et al. 2010

Background

Agent-Centric

Kitani et al. 2012, Koppula et al. 2013, etc.

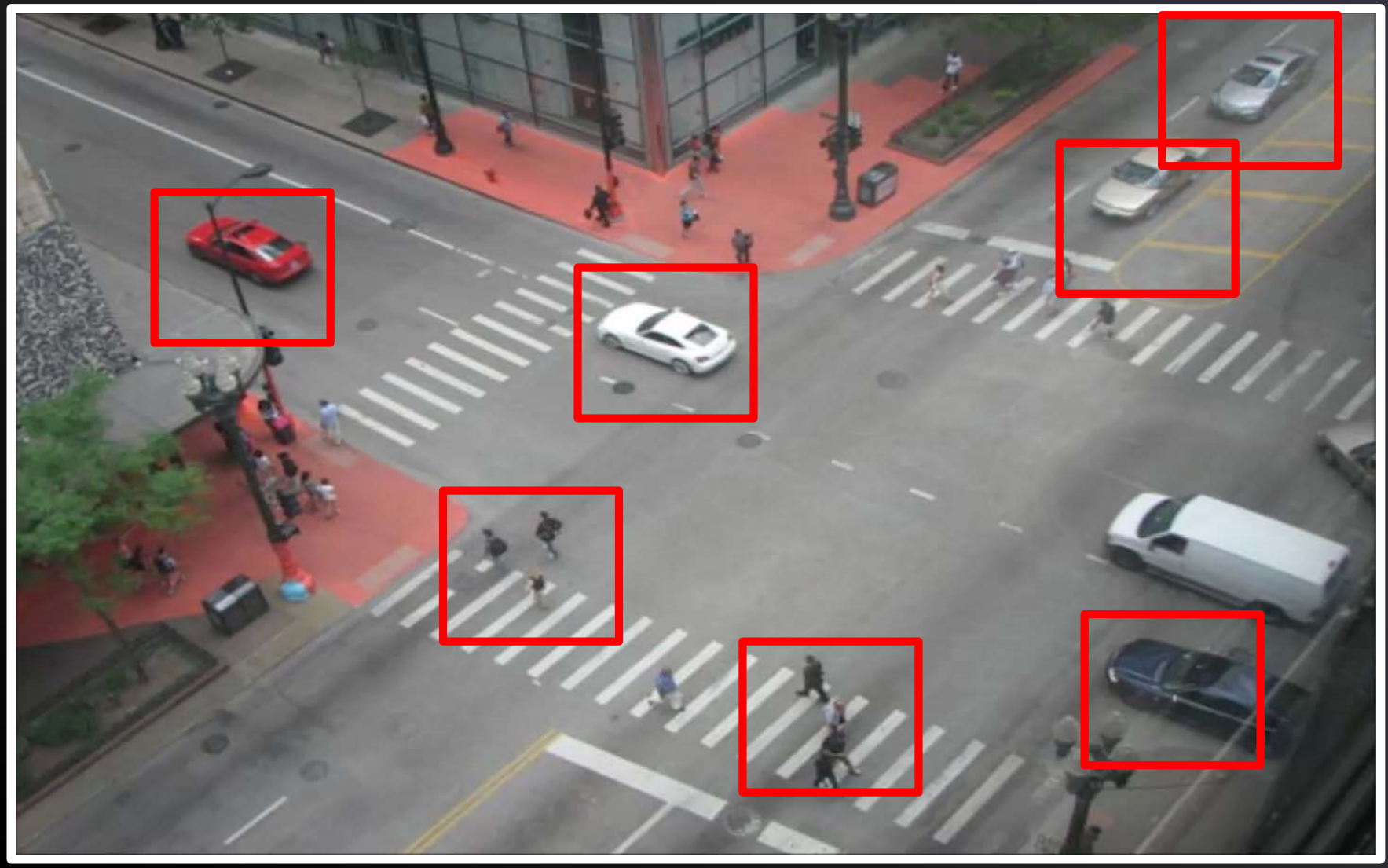
Background

Agent-Centric

Kitani et al. 2012, Koppula et al. 2013, etc.

Our Approach

Data-Driven



Our Approach

Data-Driven + Agent-Centric

Our Approach

Unsupervised

Limitations

Domain-Dependent

Train

Test

Limitations

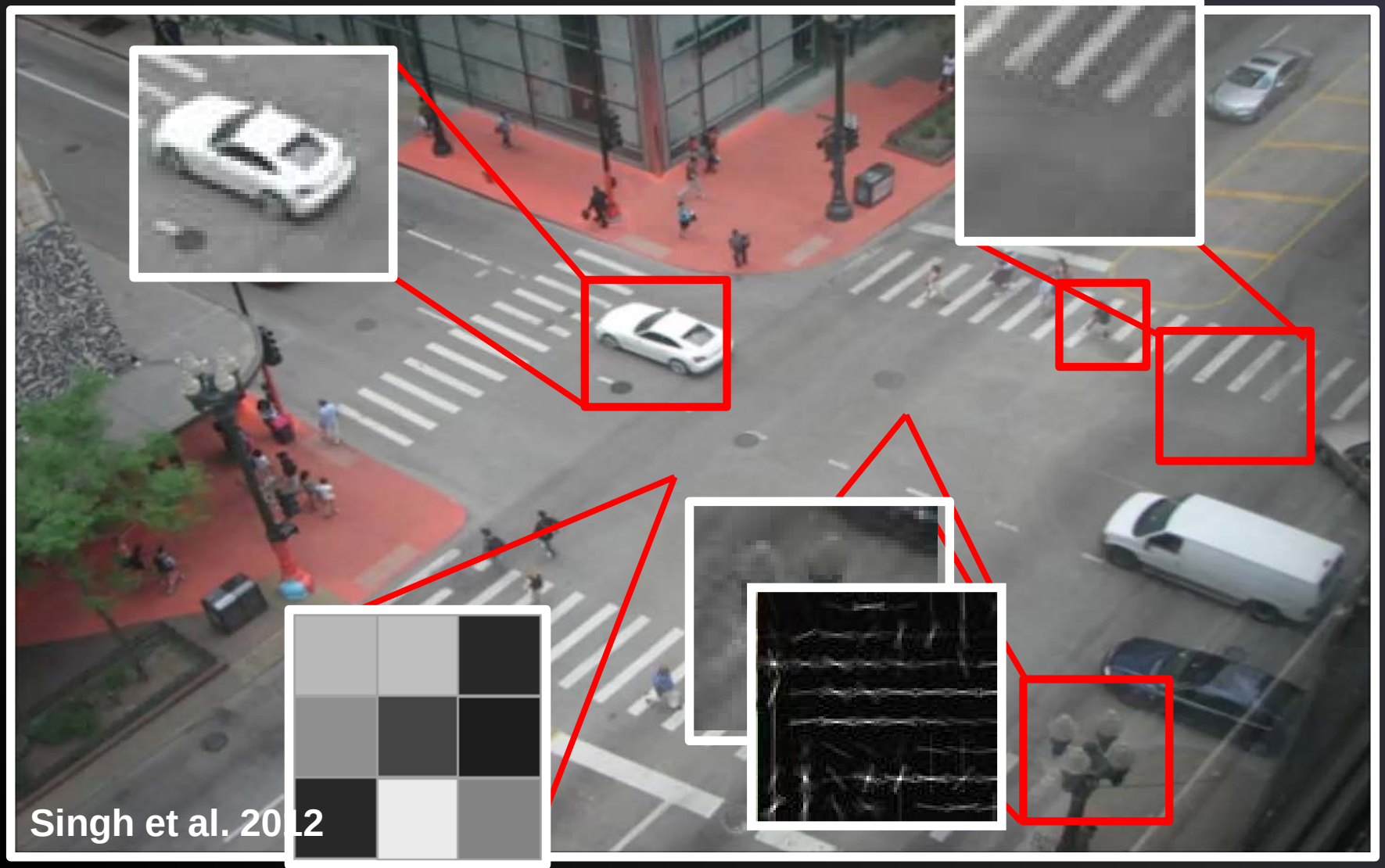
Goal-Driven

Limitations

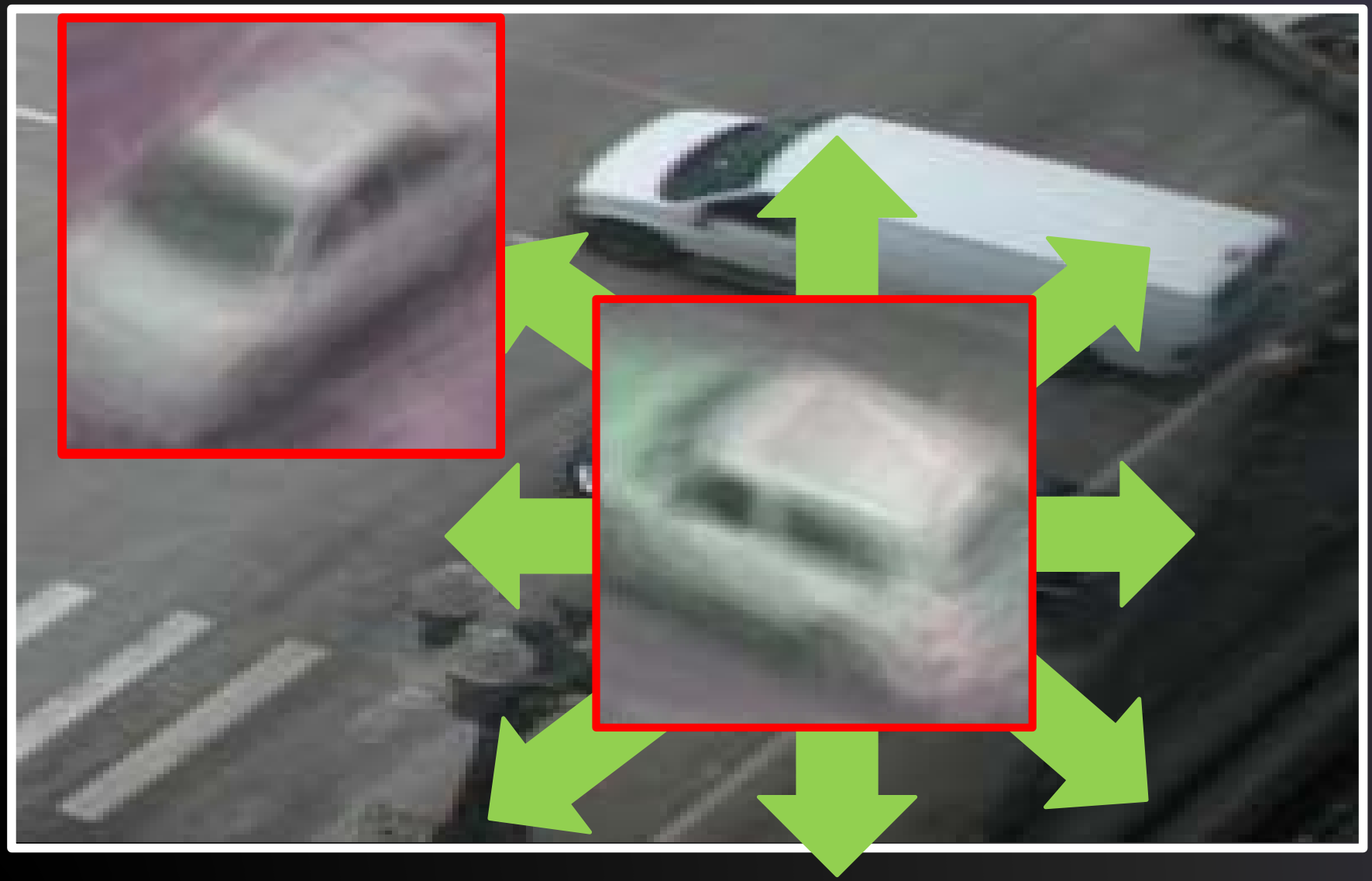
No Inter-Element Prediction

Overview

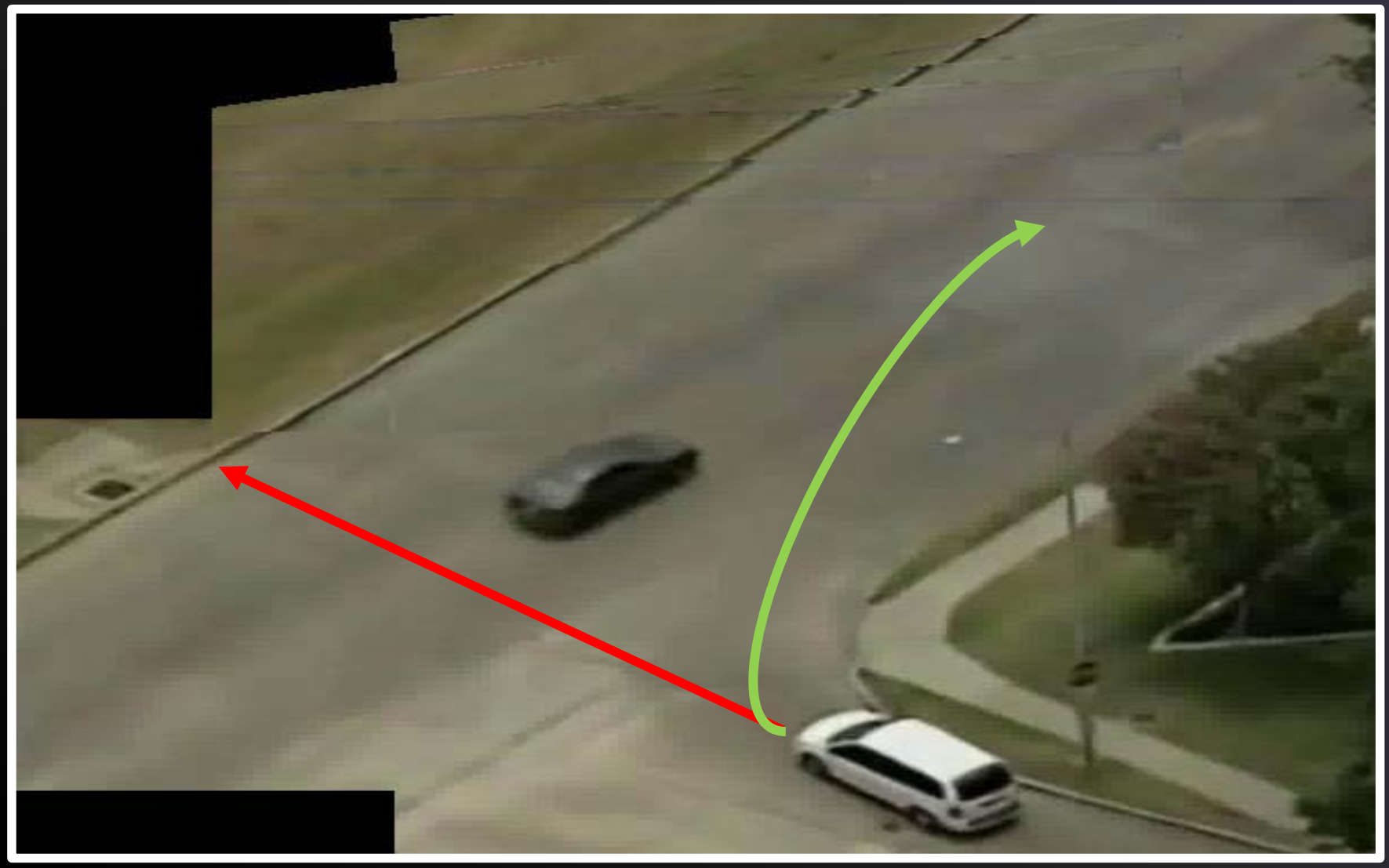
Representation



Action Space

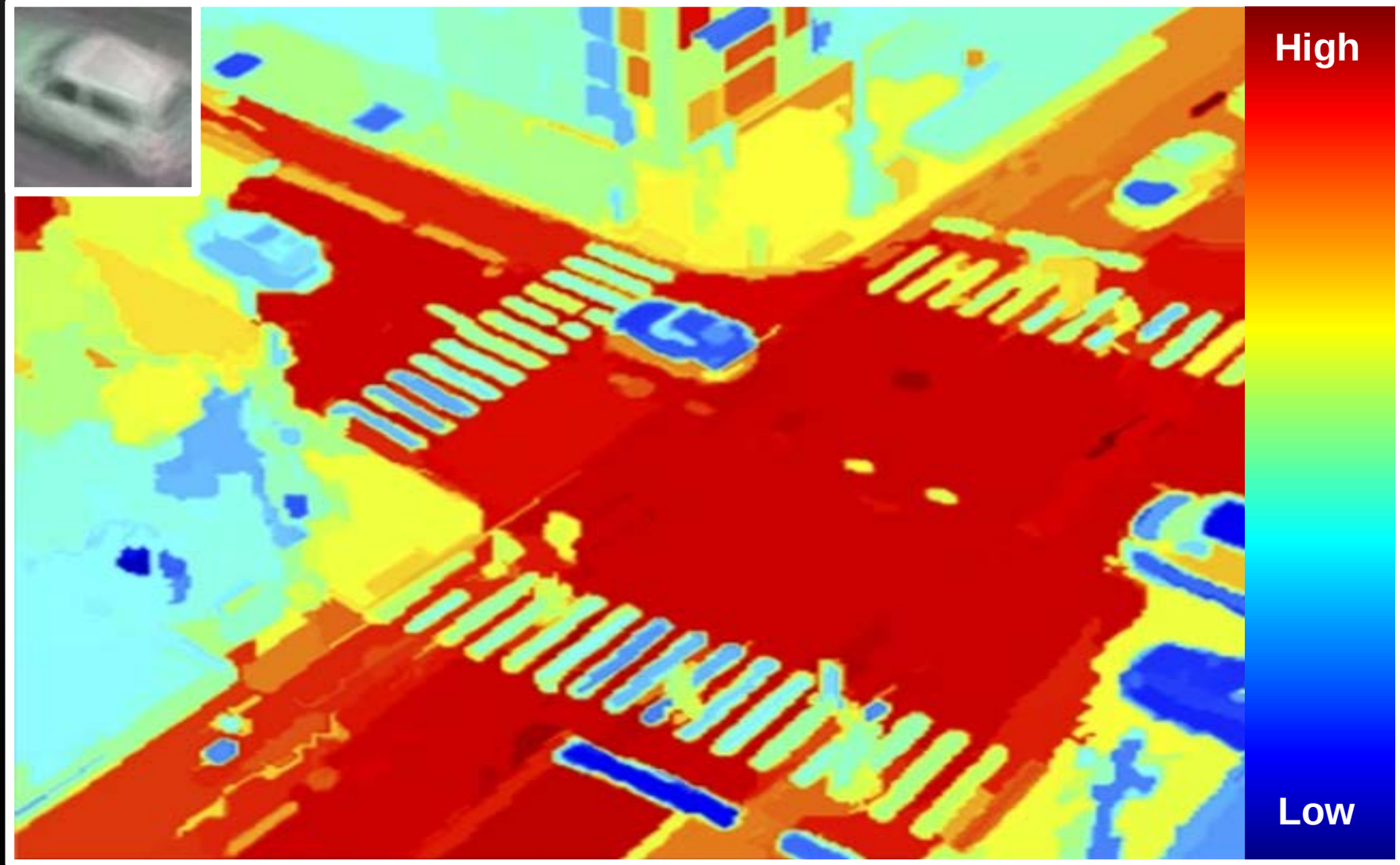


Scene Interaction



Scene Interaction

Scene Interaction



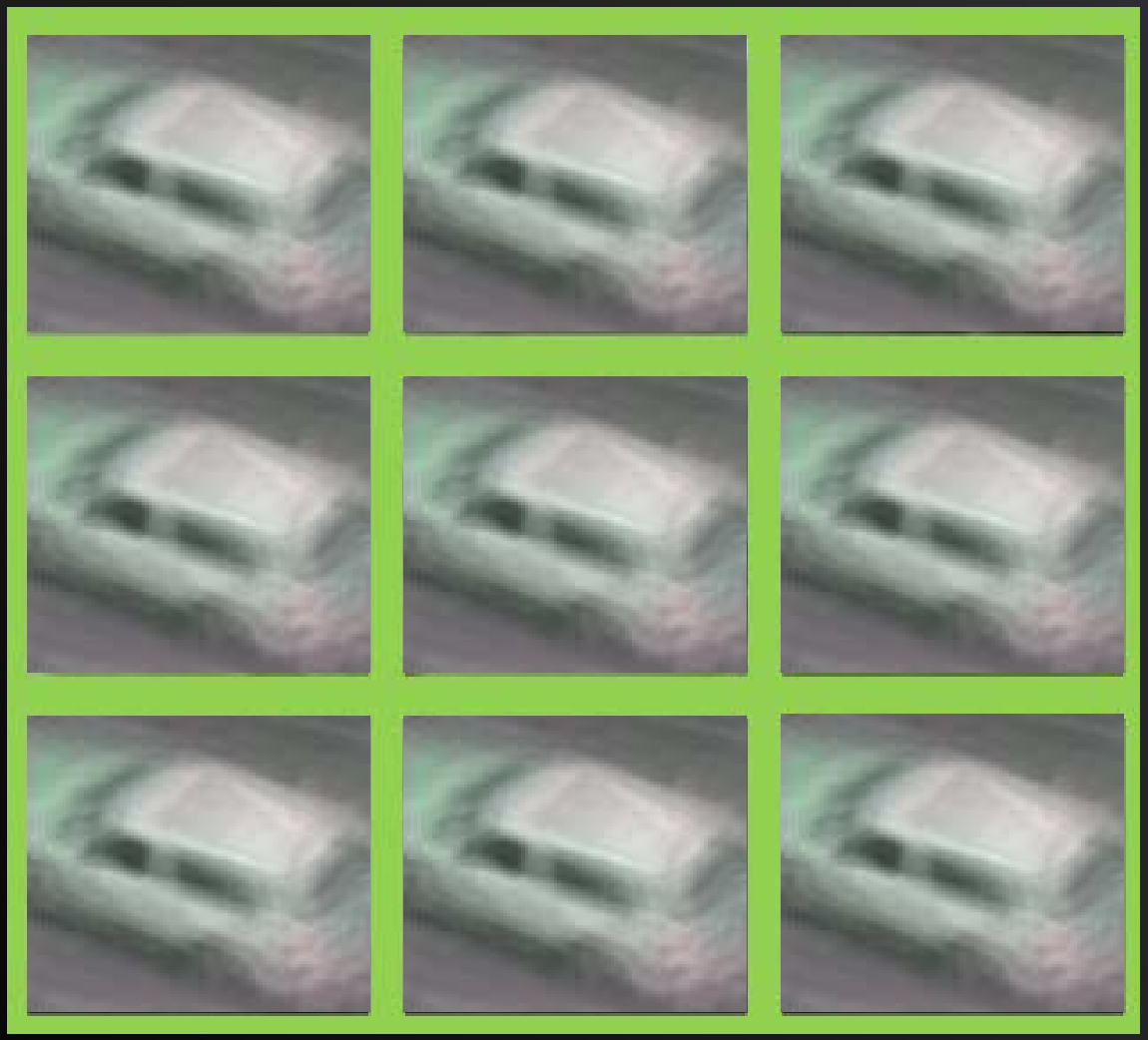
Expected Reward

P(Transition)

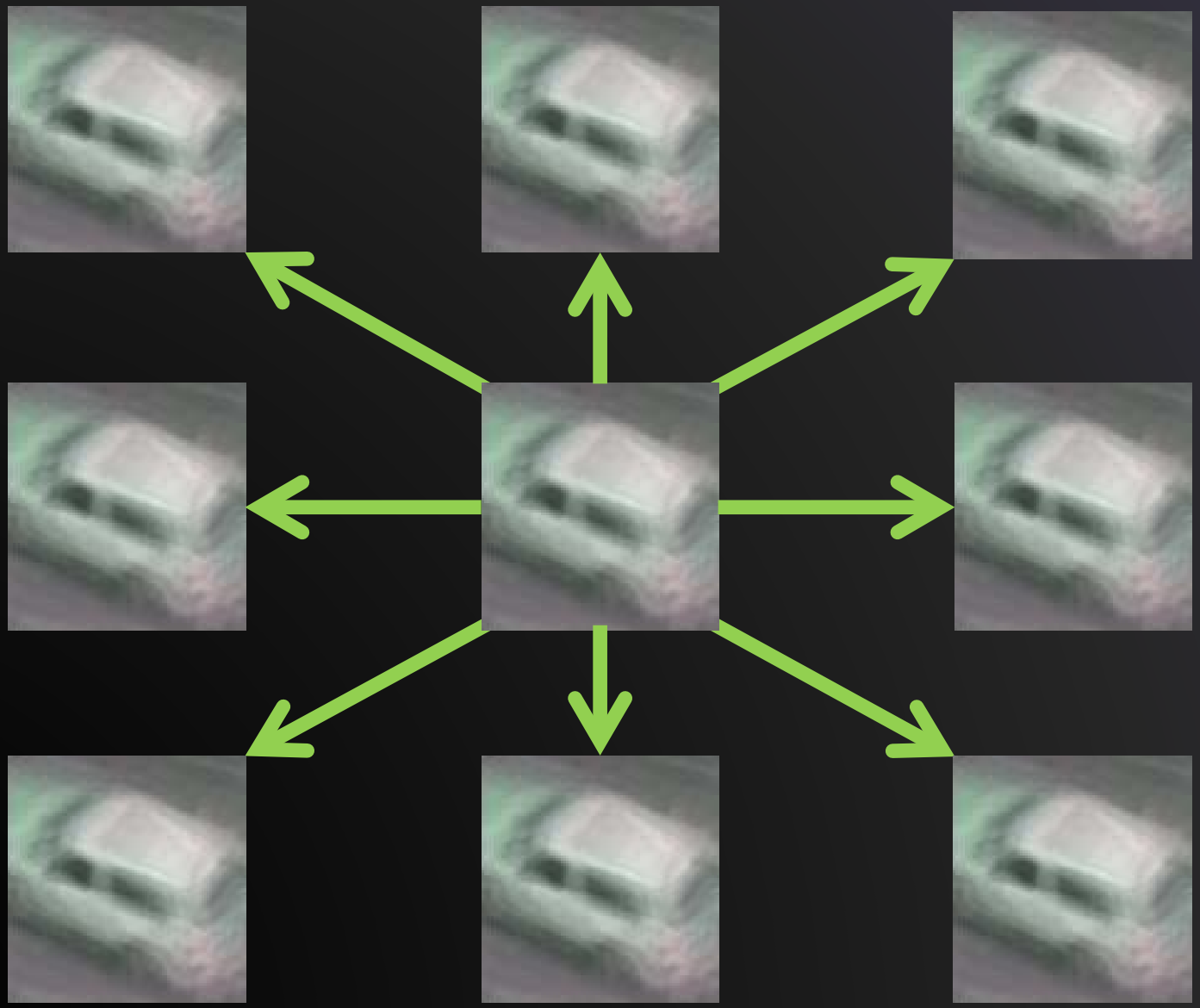
Reward(X,Y,C)

E(Reward) = P(T) * R

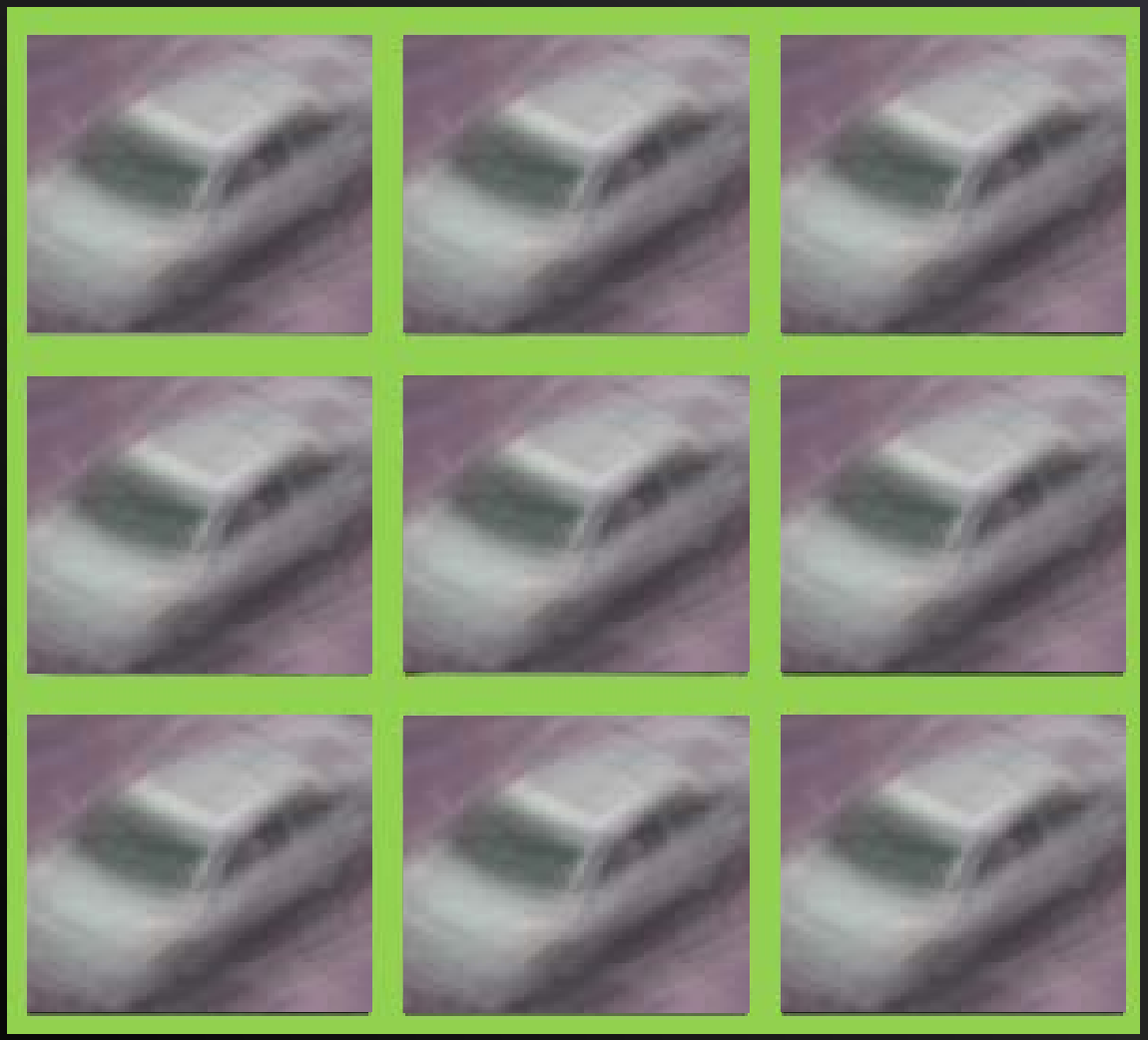
Planning



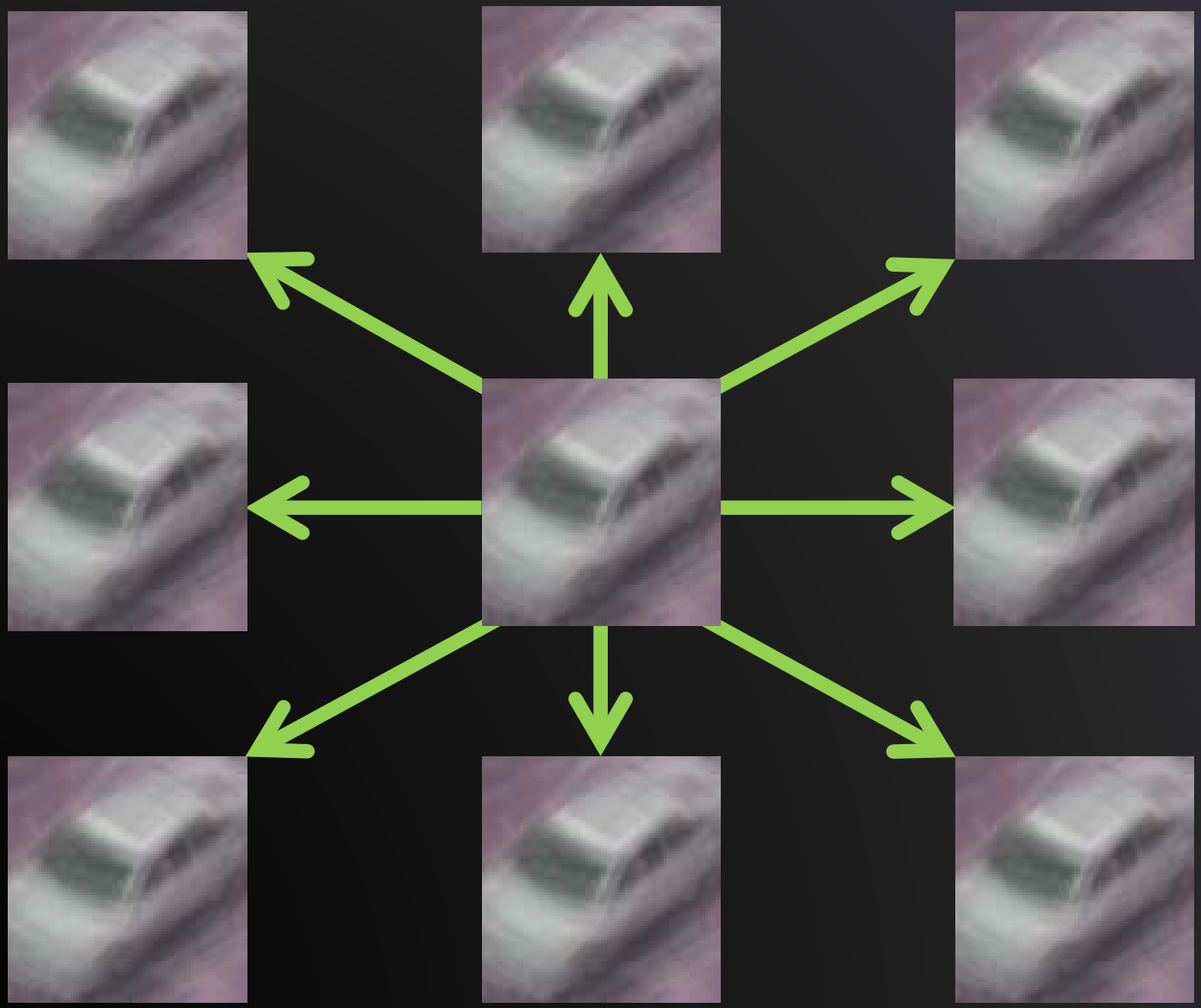
Planning



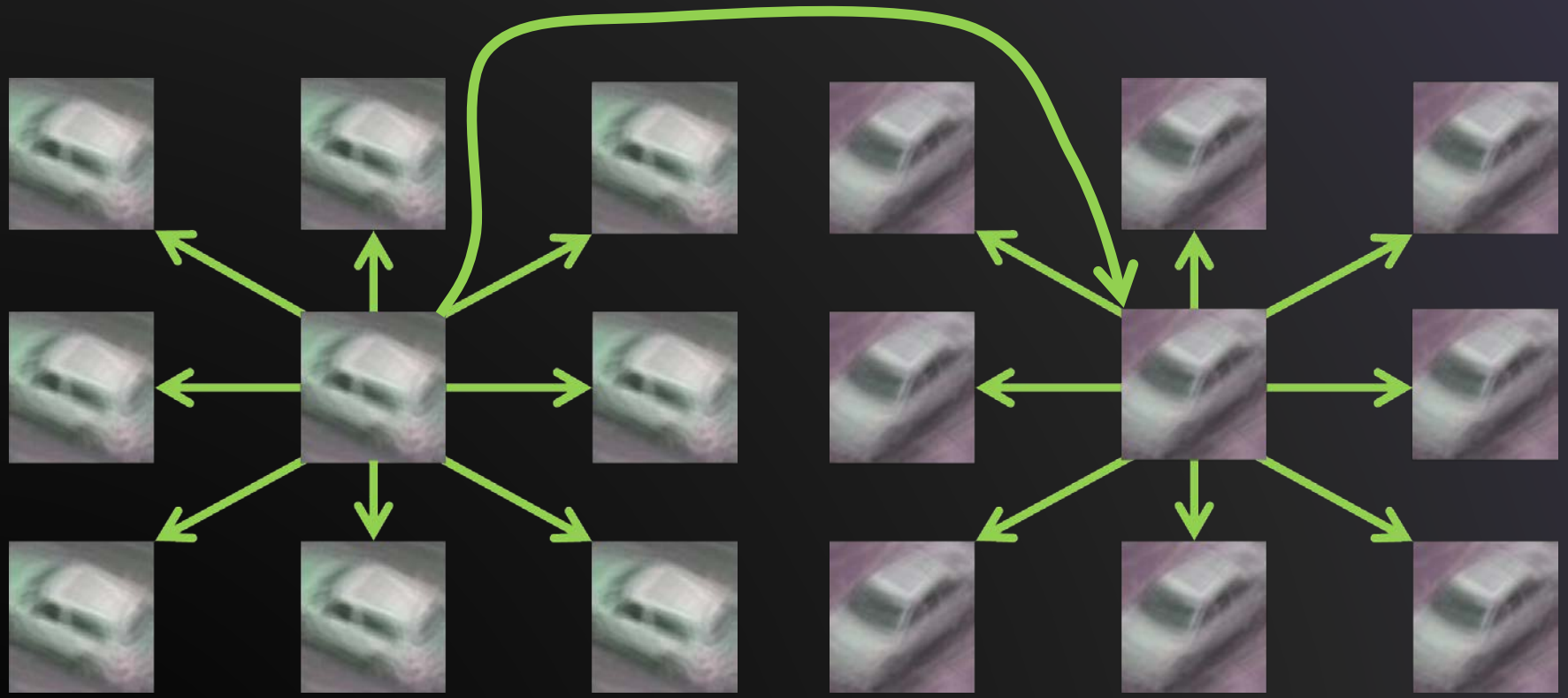
Planning



Planning

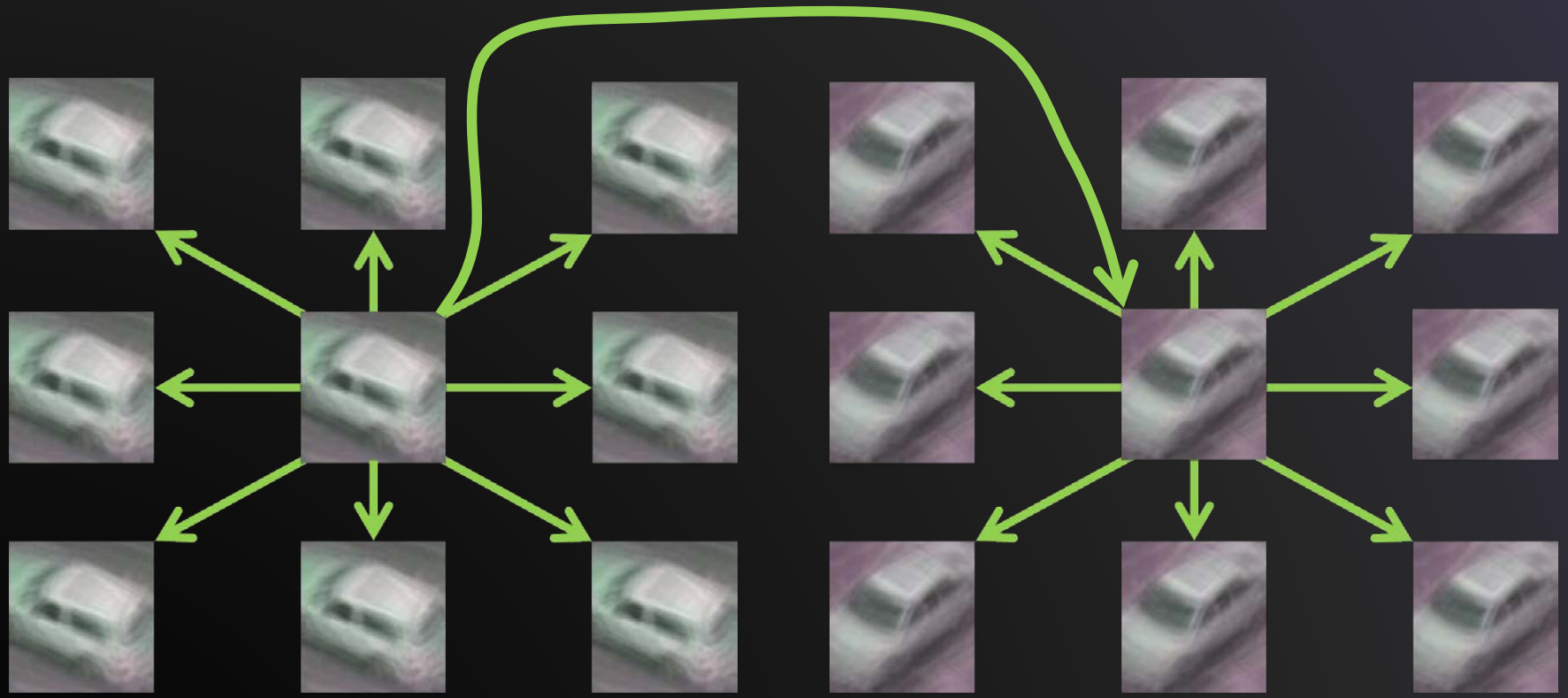


Planning

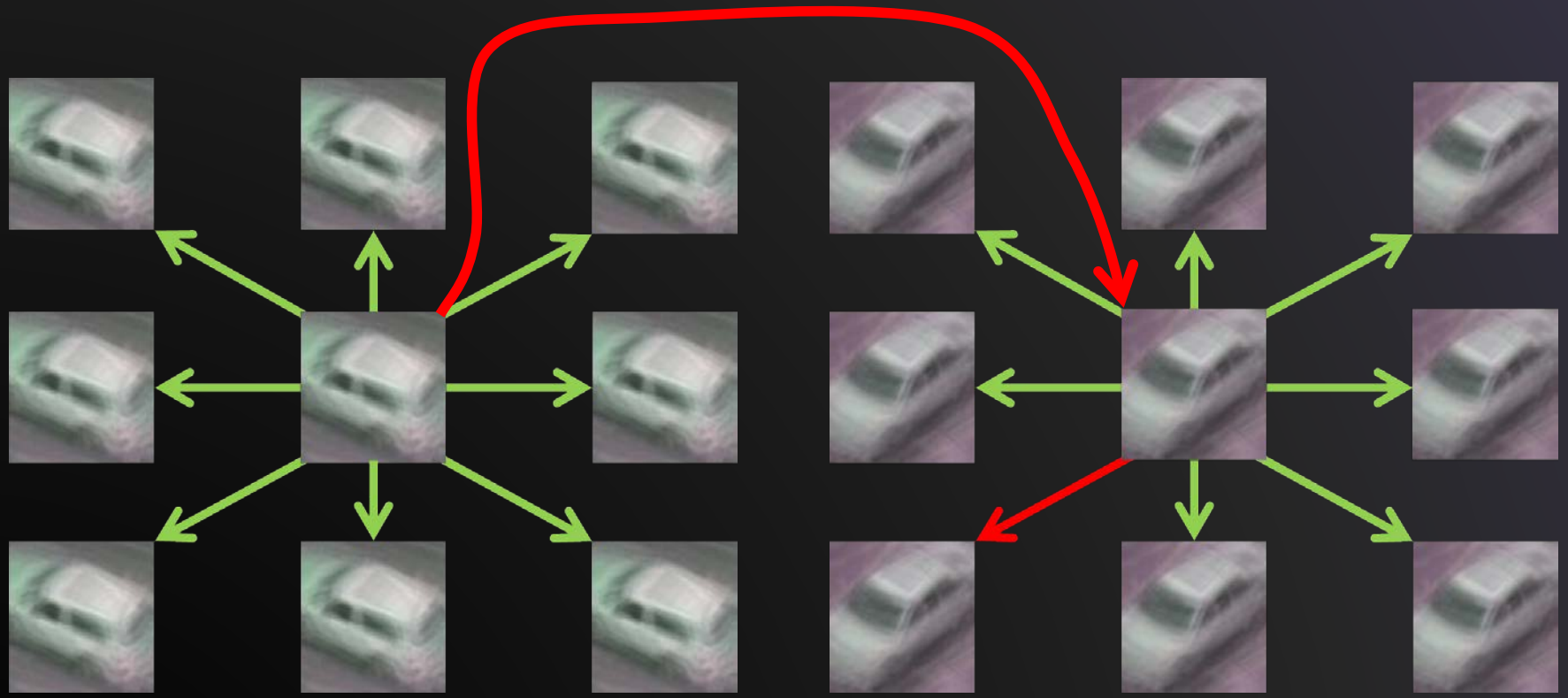


Planning

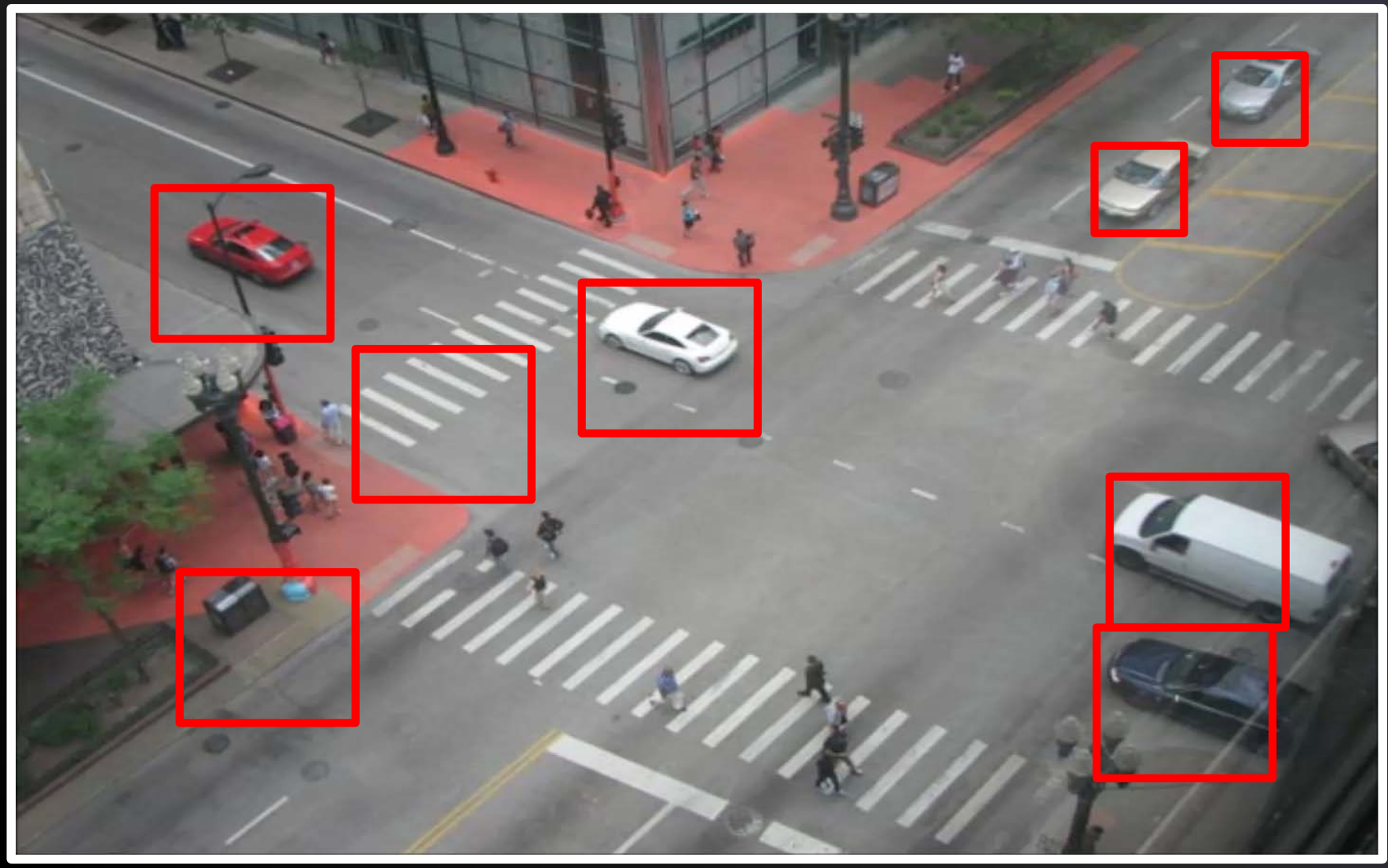
Planning



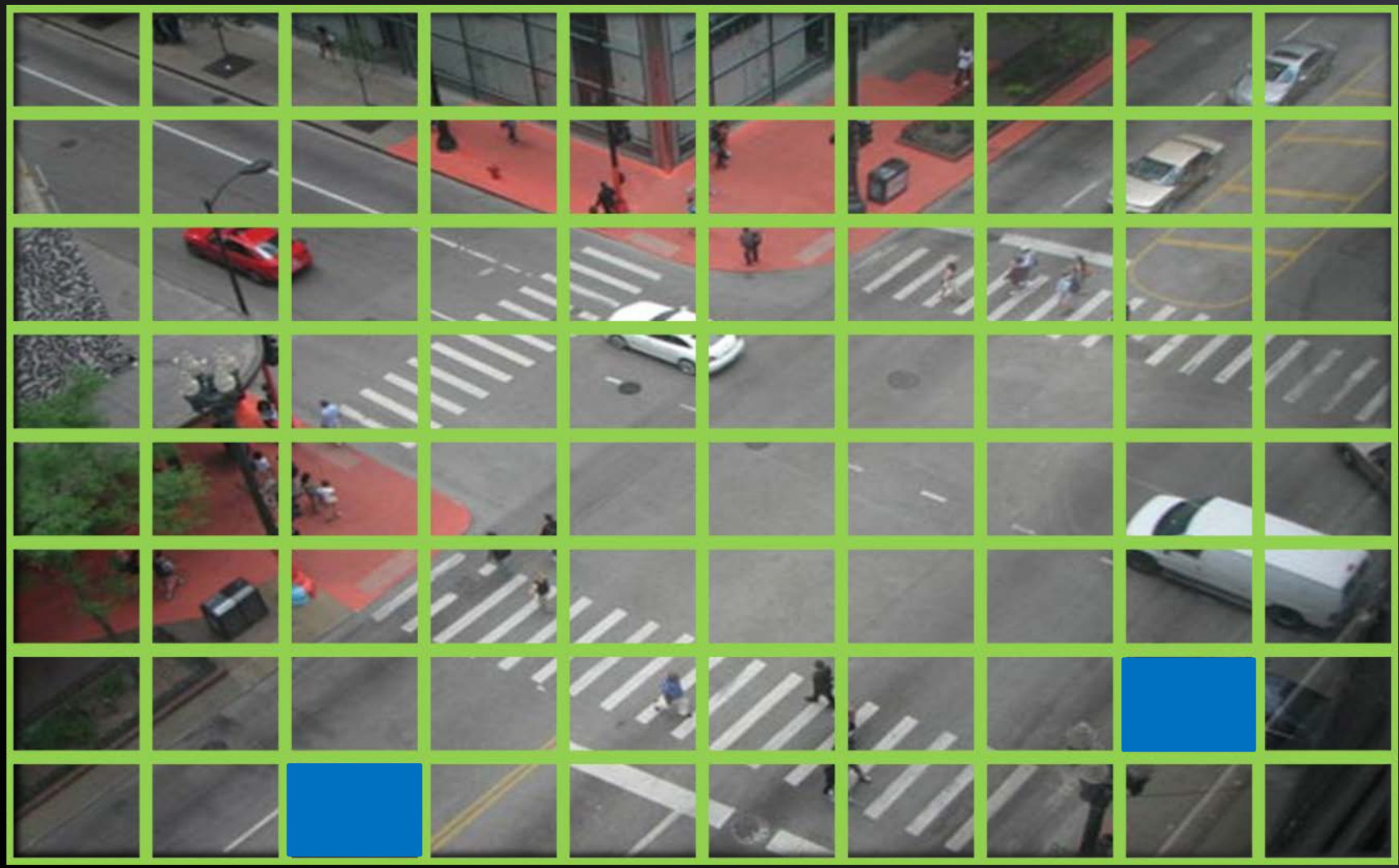
Planning



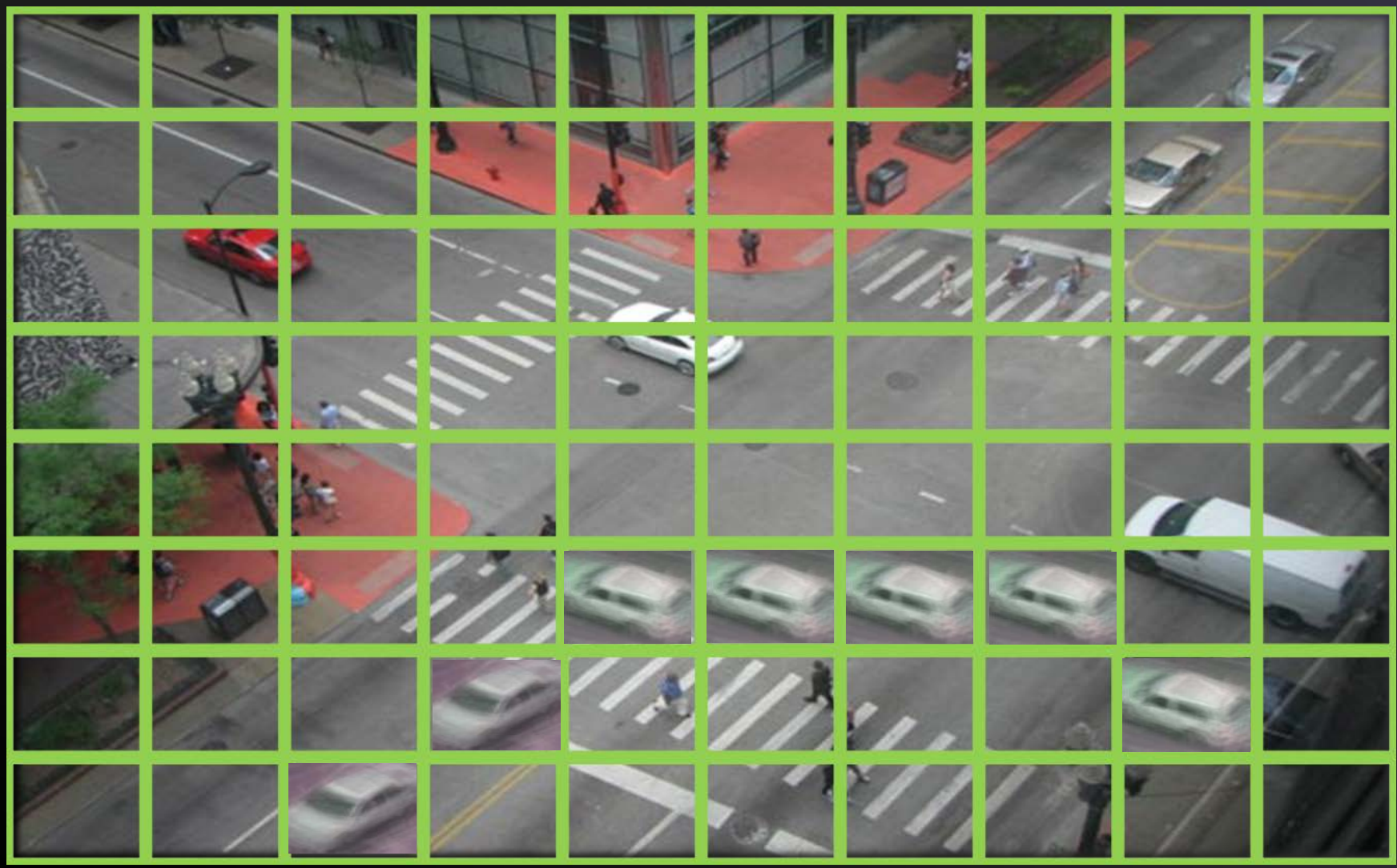
Planning



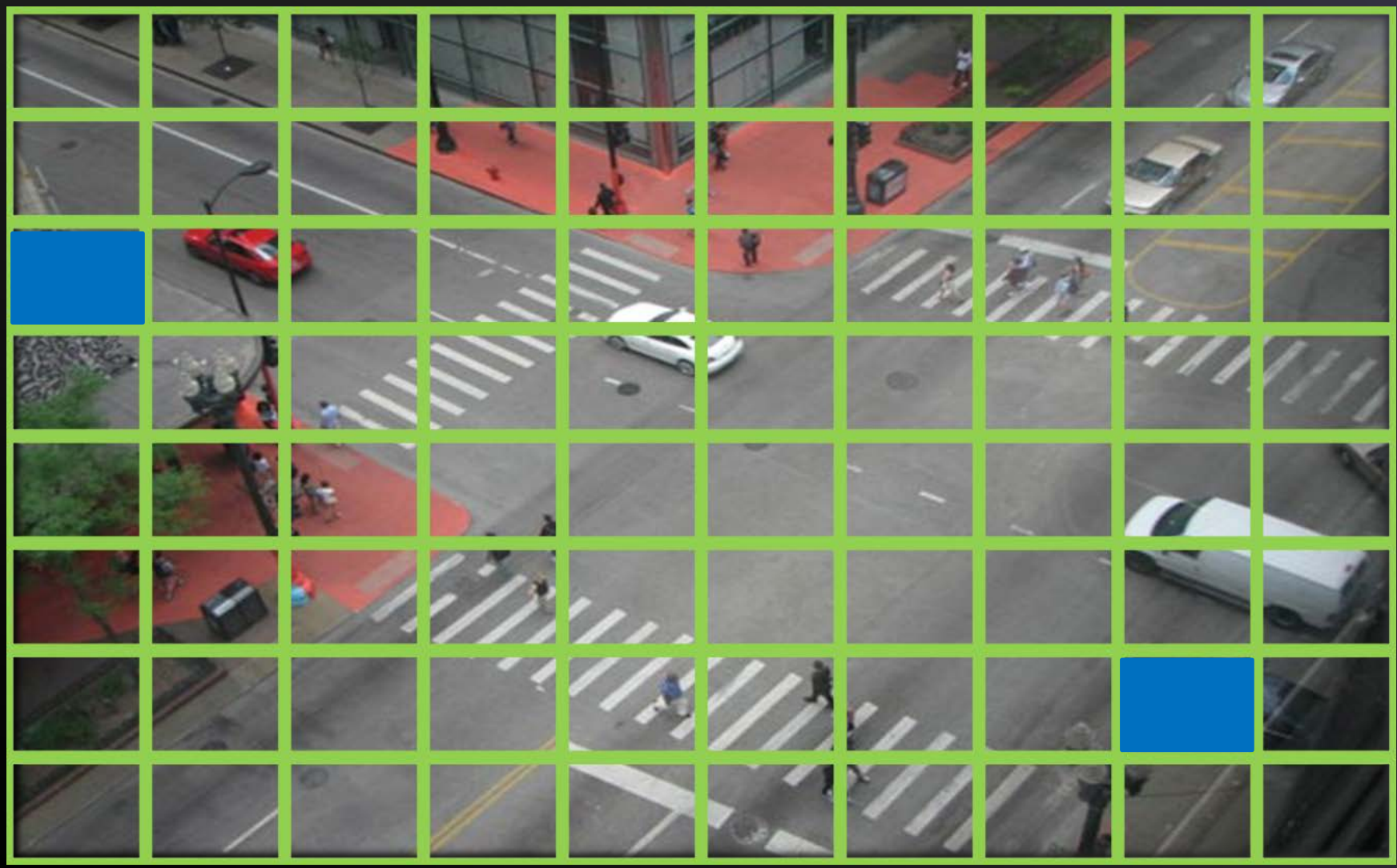
Planning



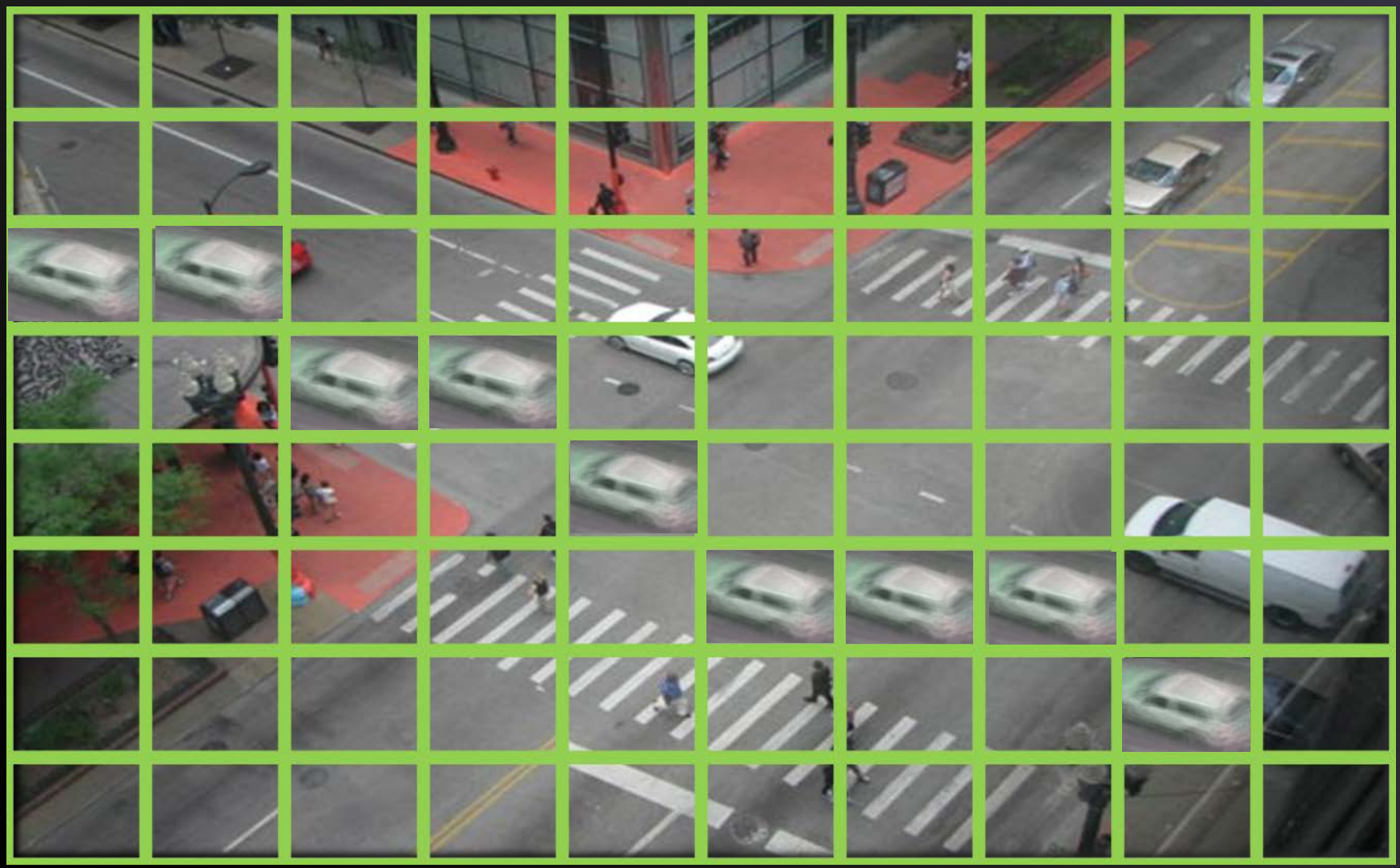
Planning



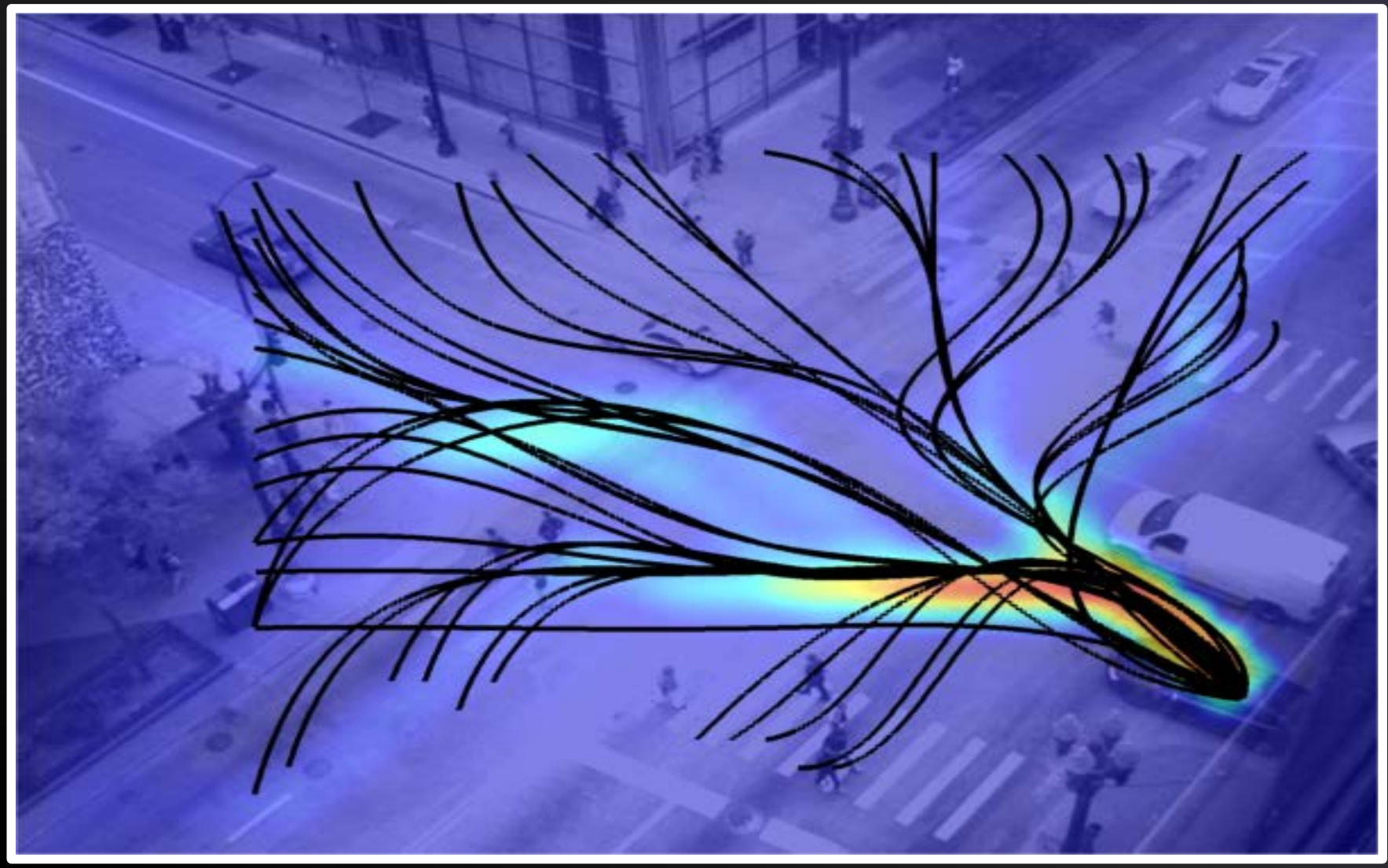
Planning



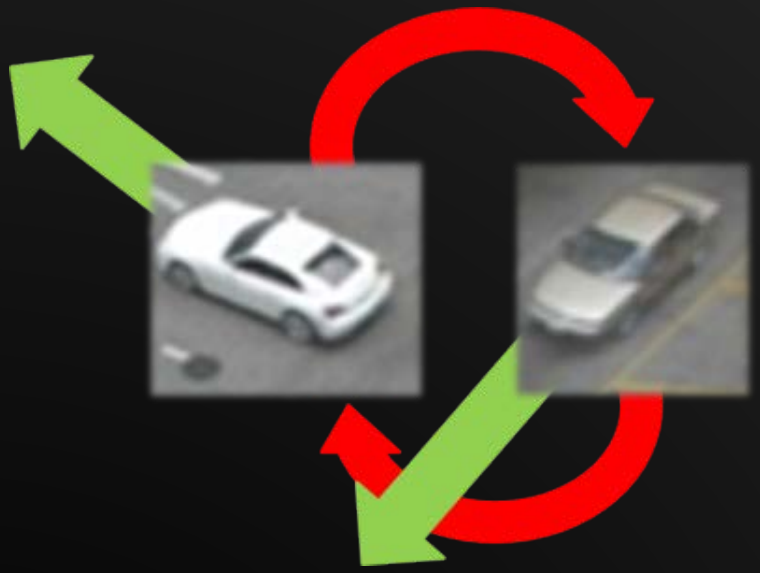
Planning



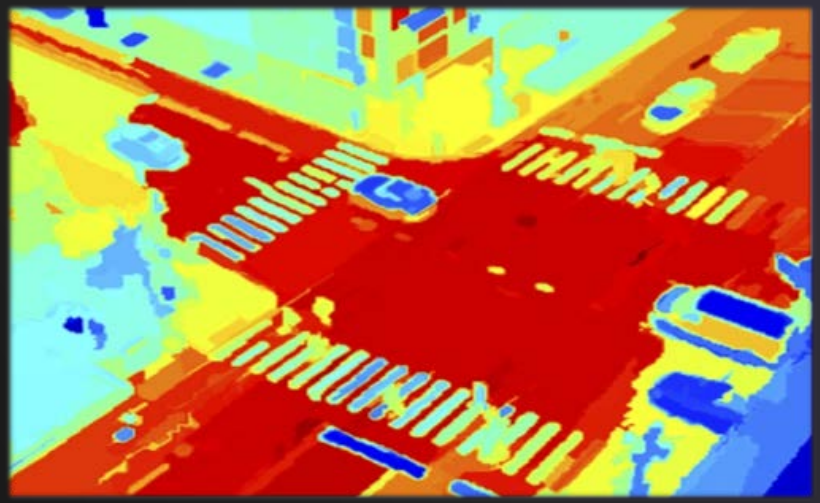
Planning



Training

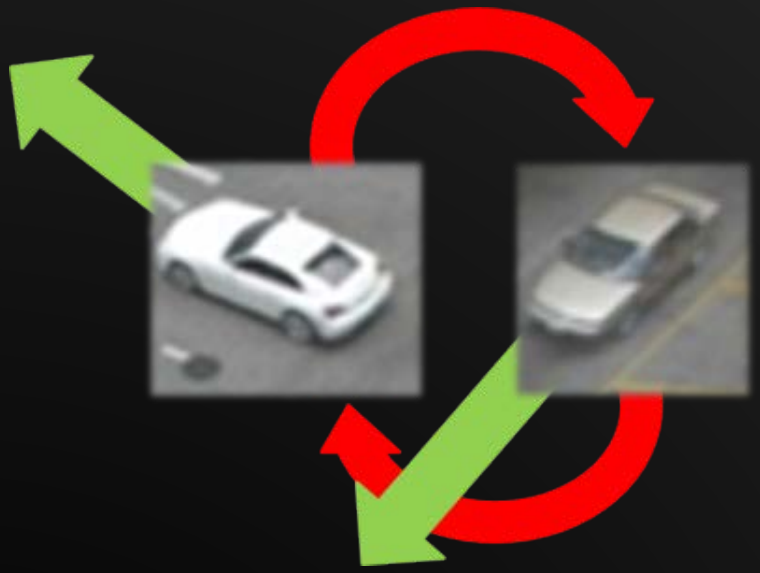


Transitions

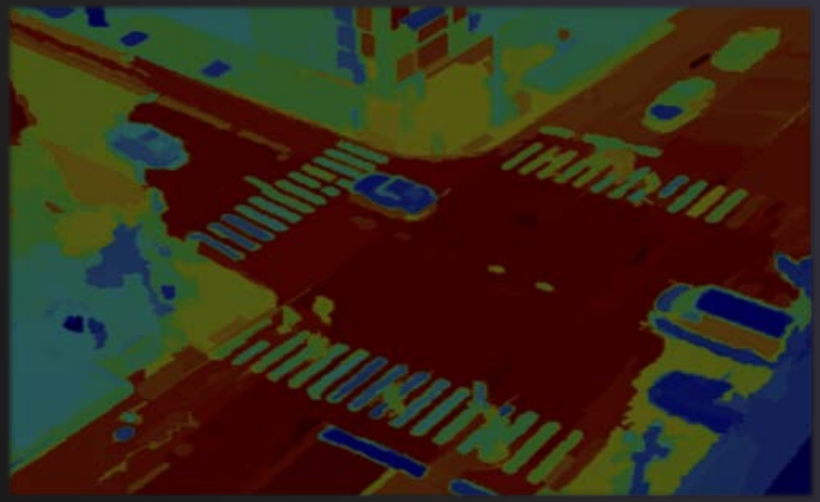


Scene Interaction

Training



Transitions

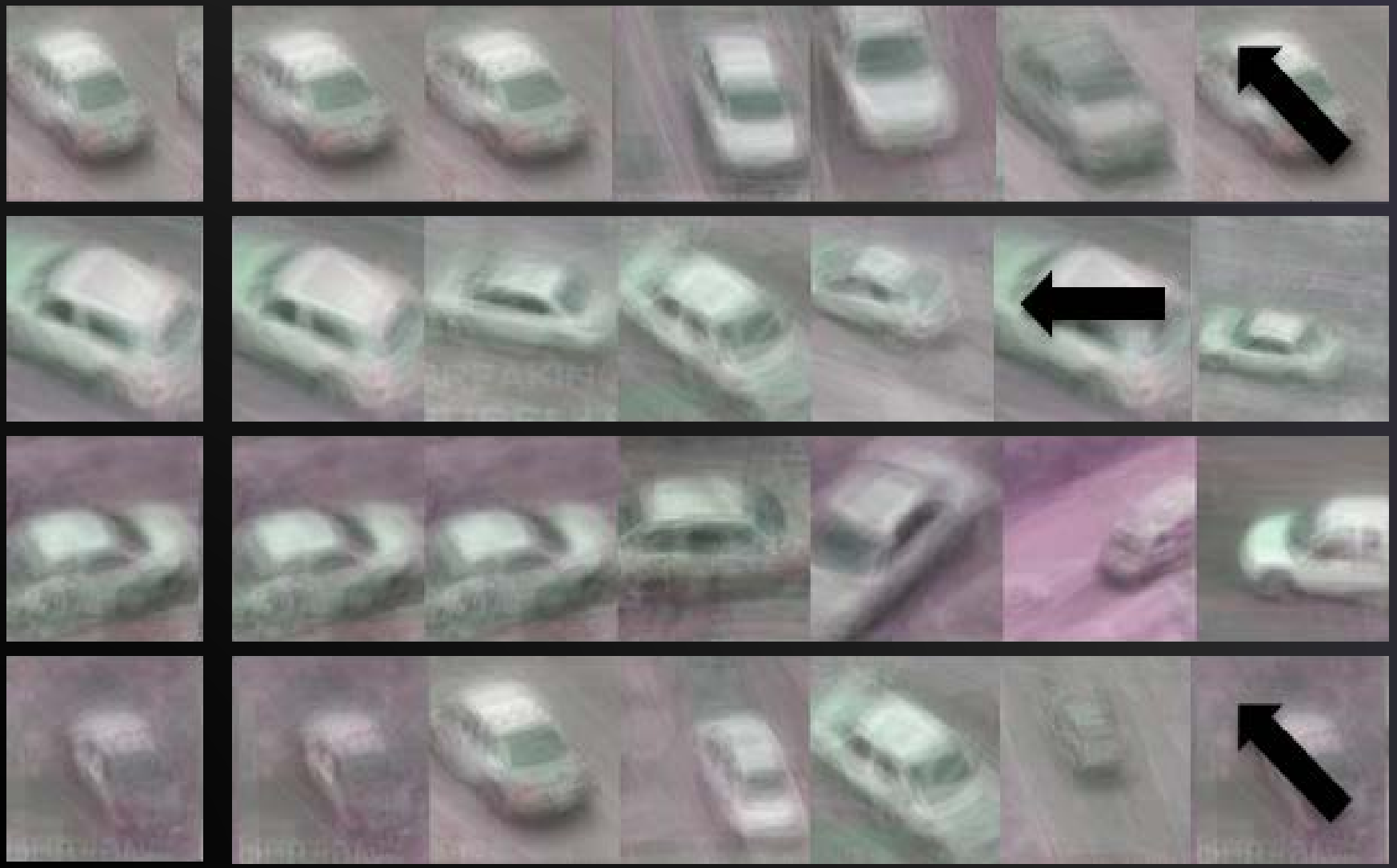


Scene Interaction

Training
Transitions

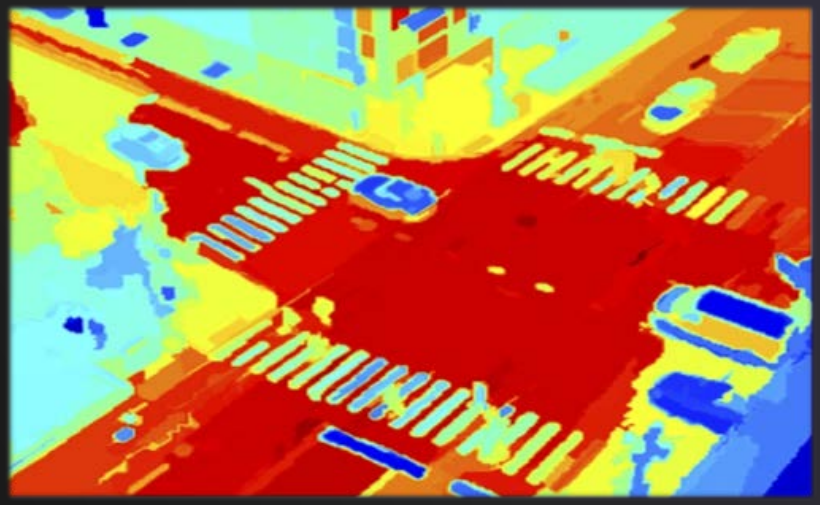
Training

Patch **Transitions**



Training

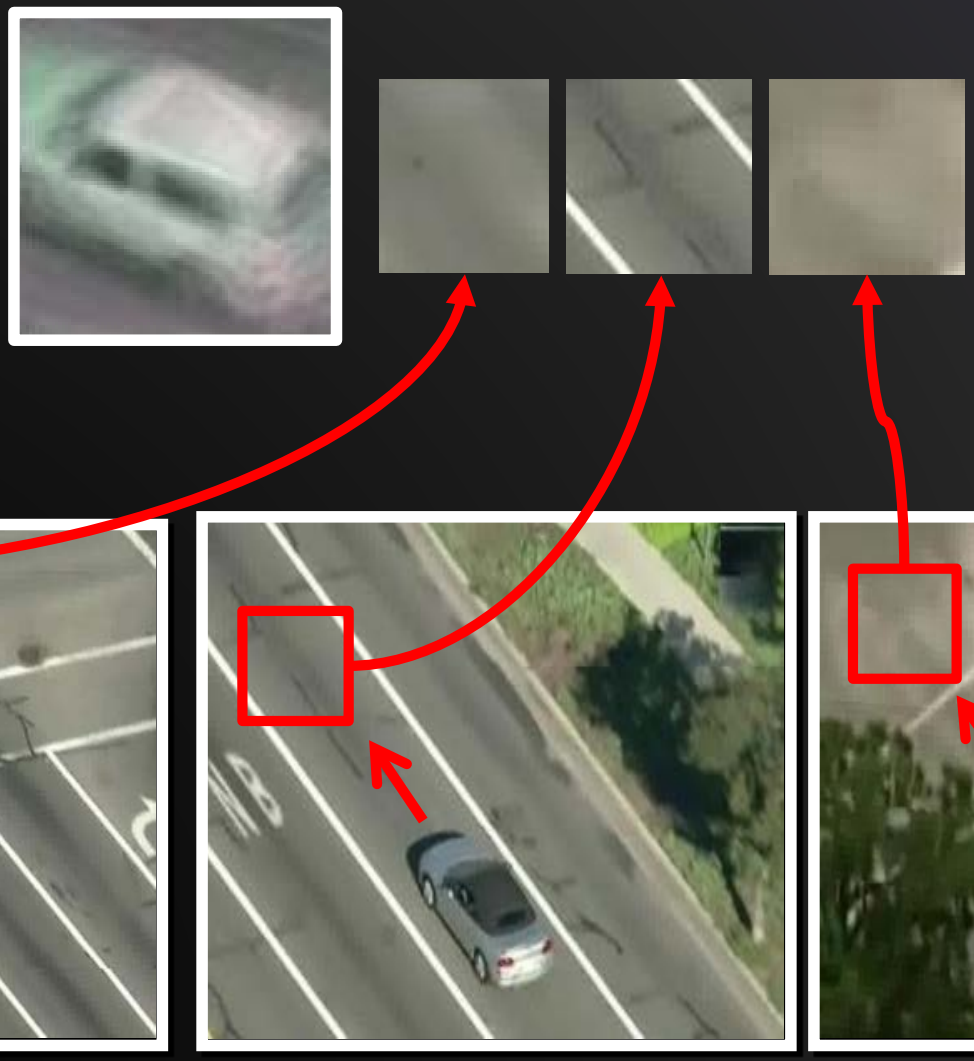
Transitions



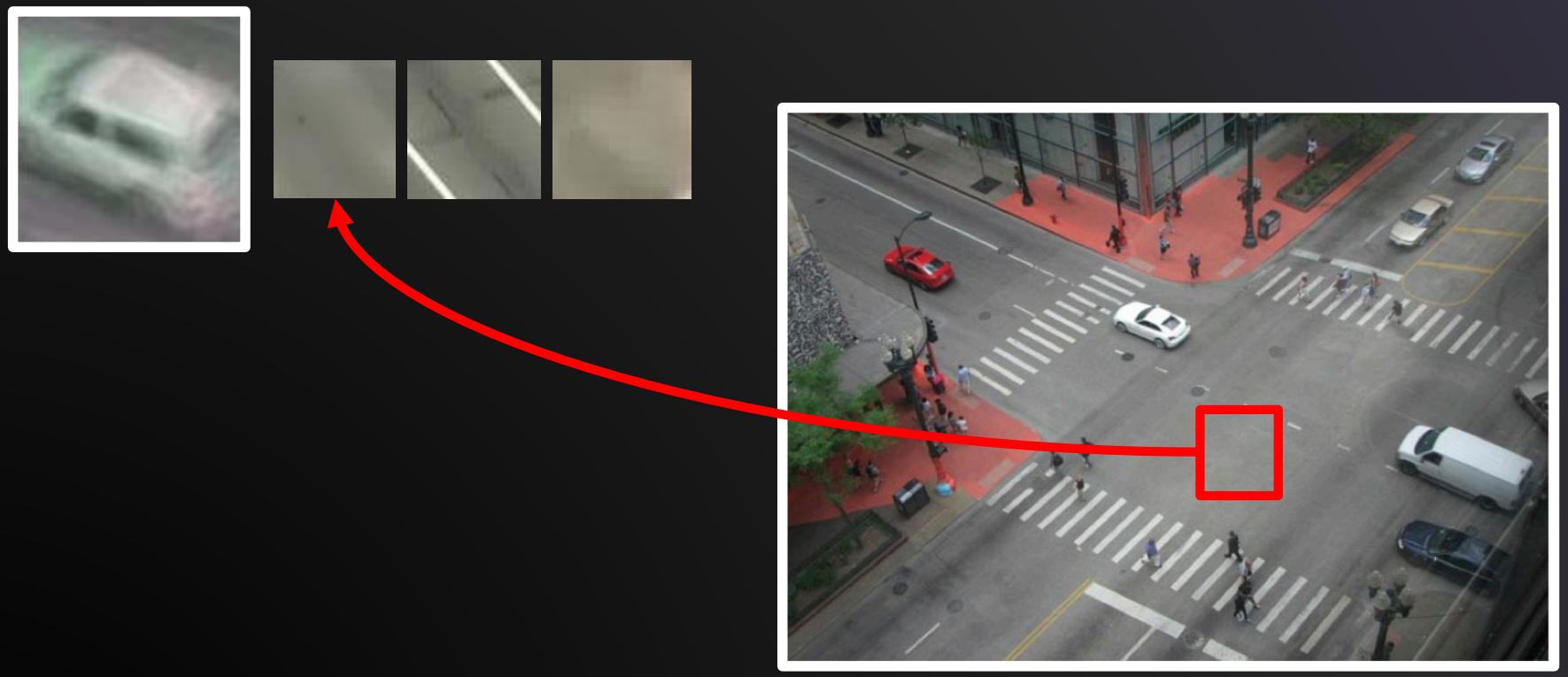
Scene Interaction

Training

Scene Interaction

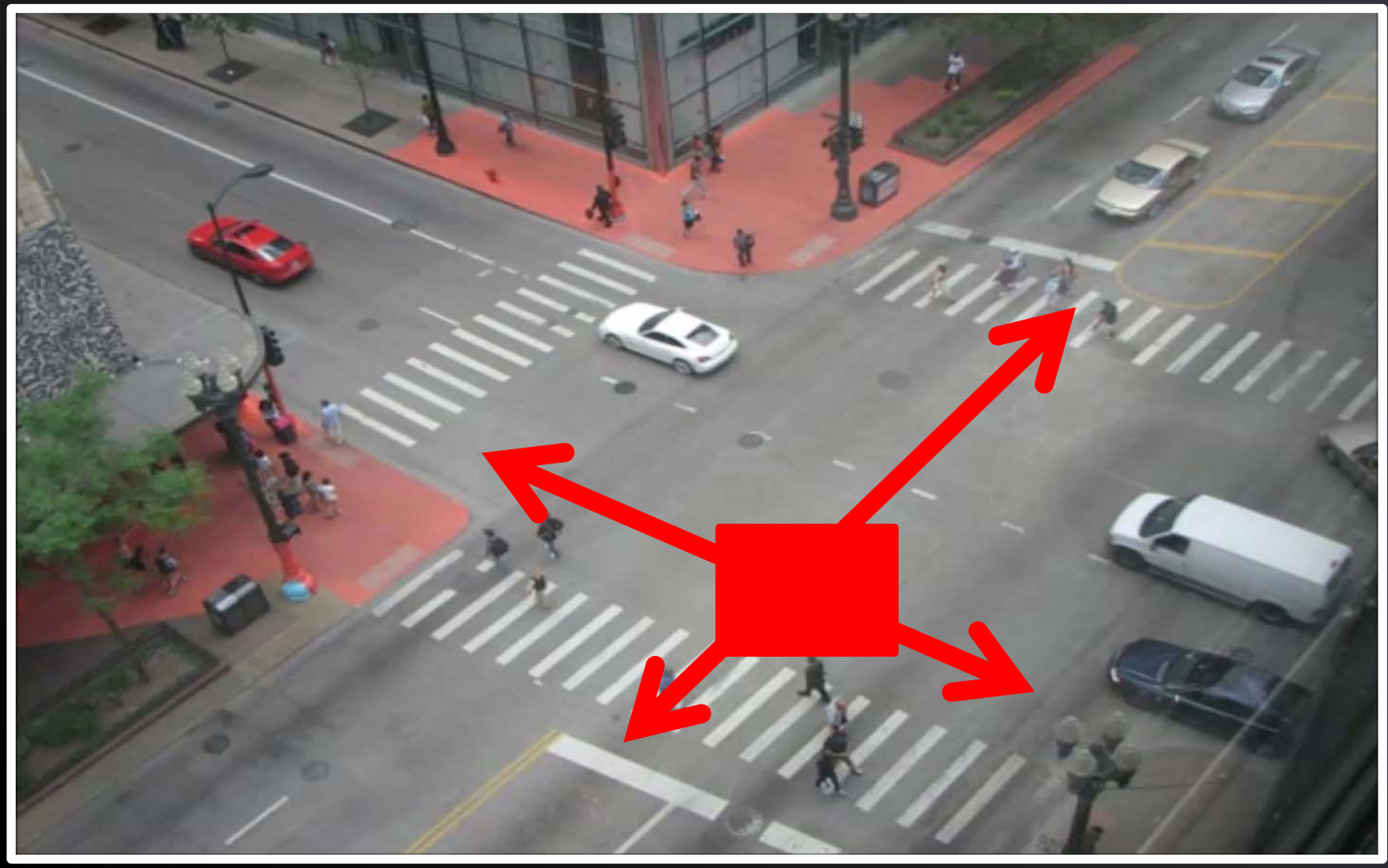


Training
Scene Interaction



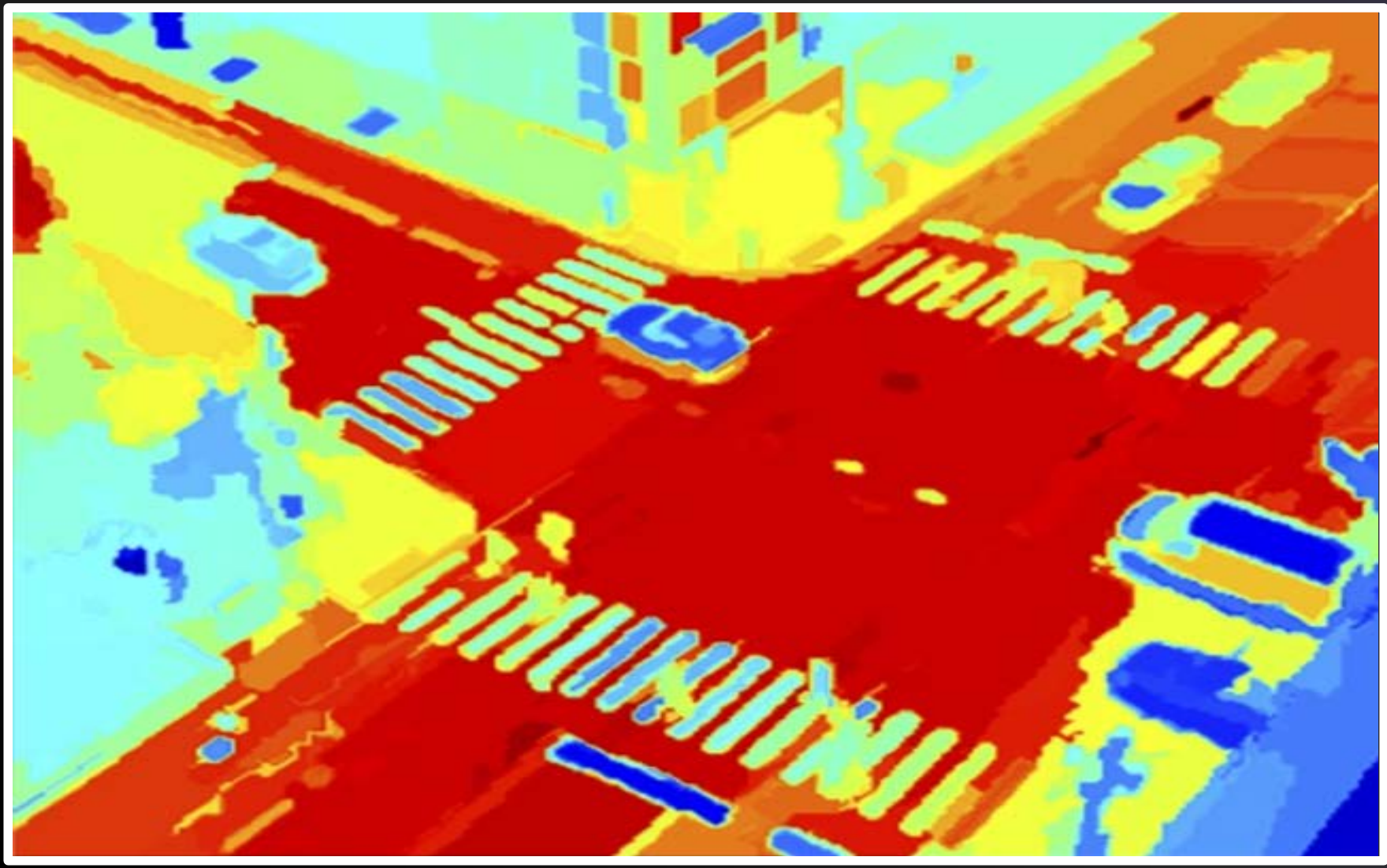
Training

Scene Interaction



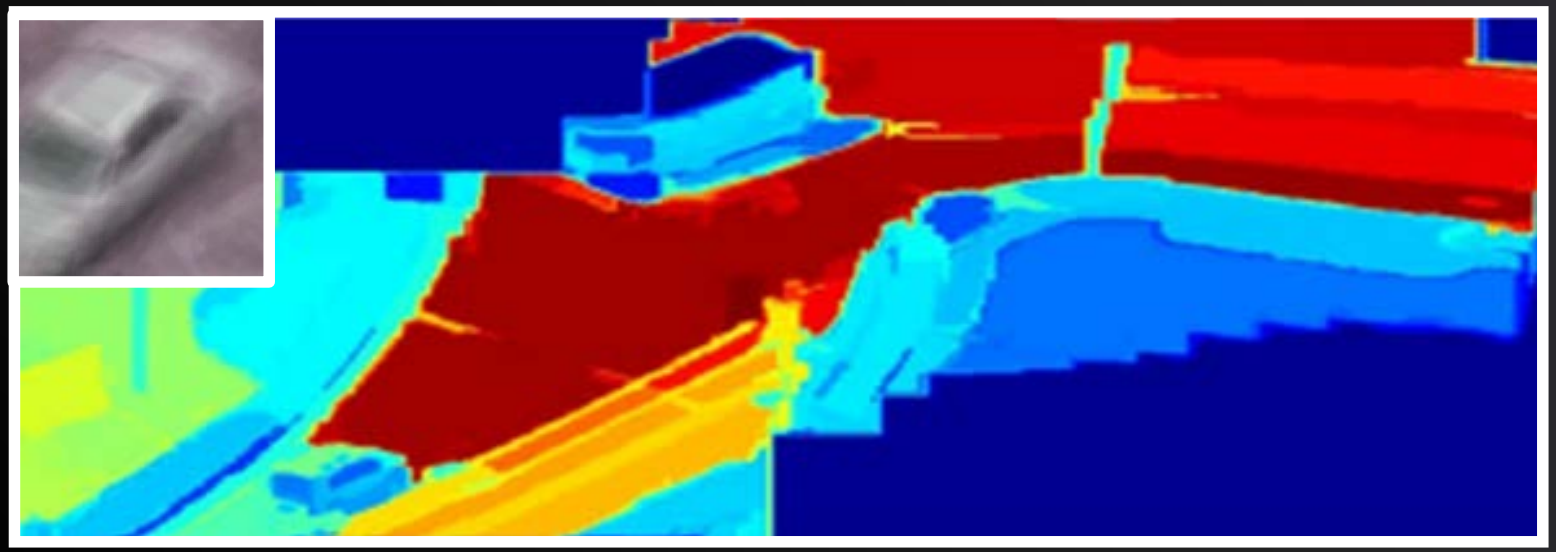
Training

Scene Interaction



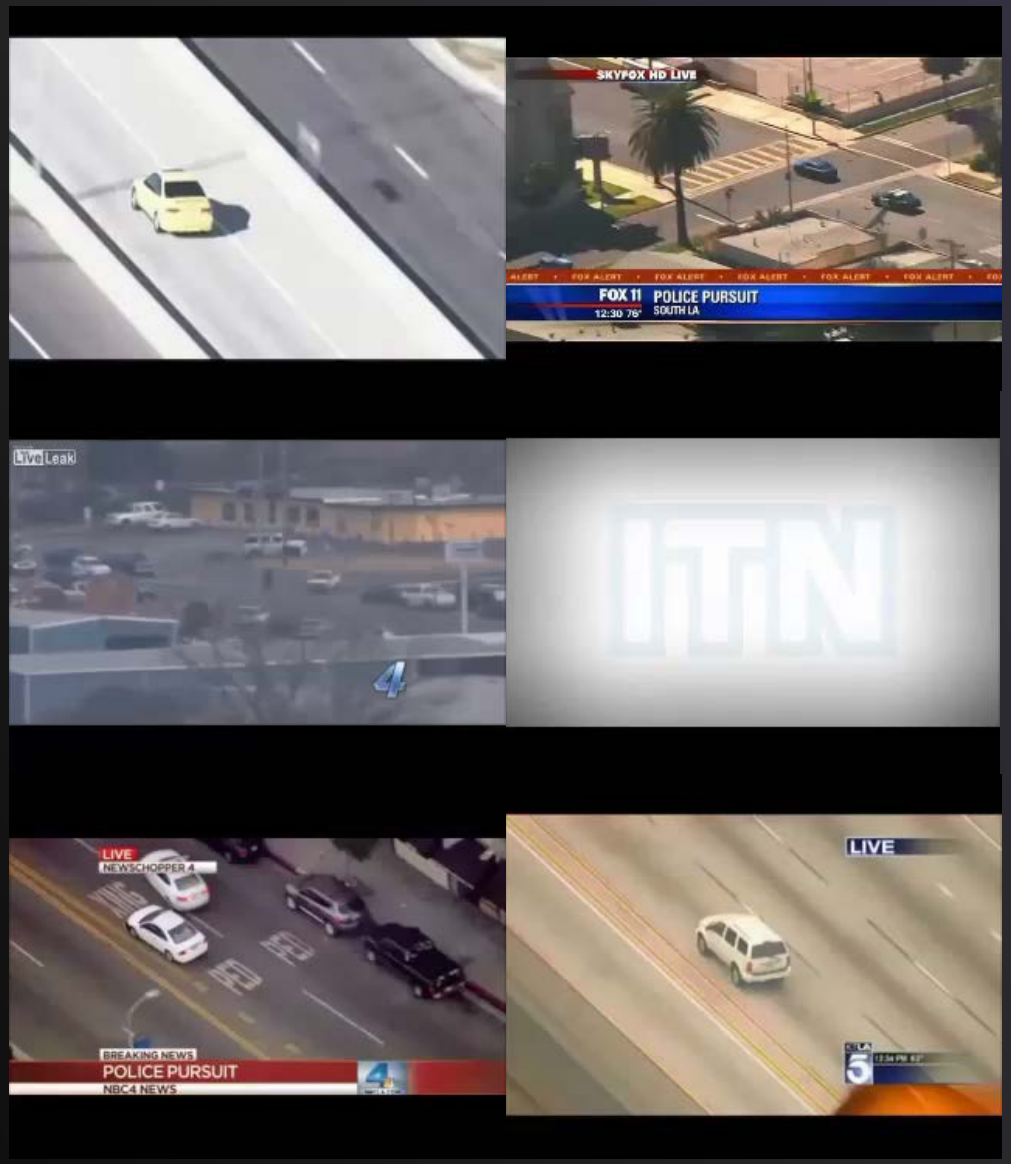
Training

Scene Interaction

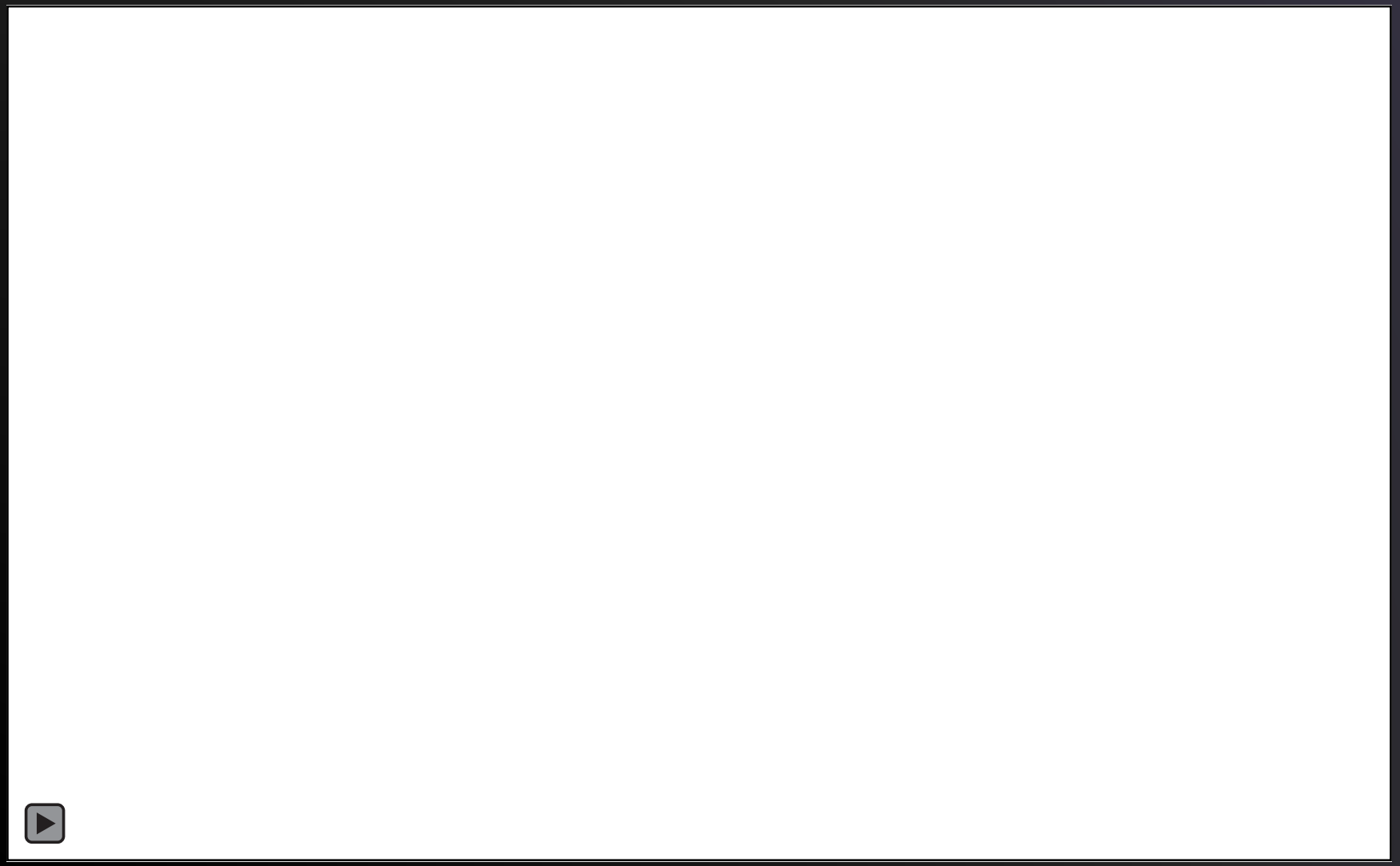


Datasets

- **183 Videos**
- **139 Training**
- **44 Testing**
- **~300 Minutes**

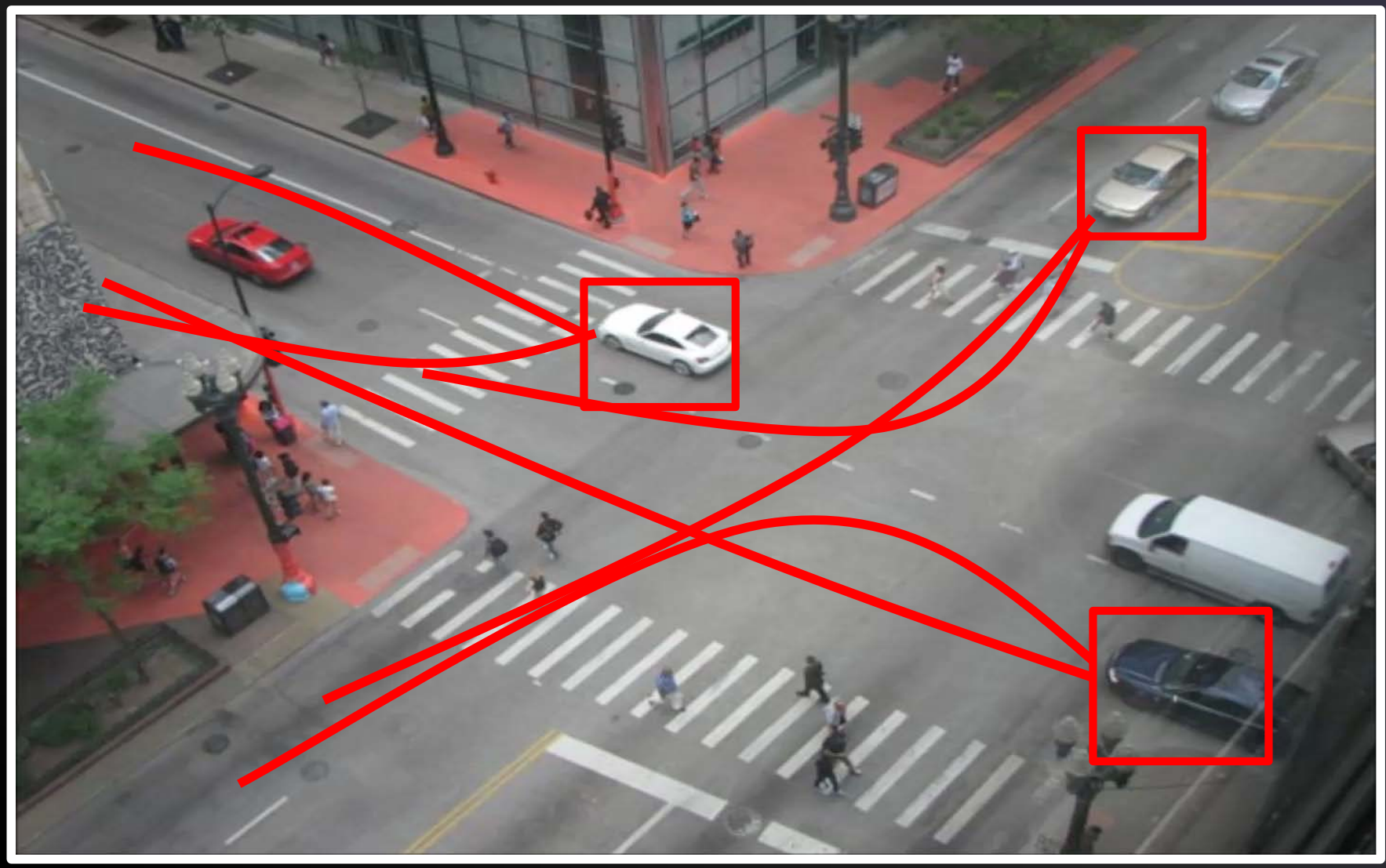


Qualitative Results



Qualitative Results

Quantitative Results



Quantitative Results

Data-Driven Active Entity

Error (Top 6)	NN + Sift-Flow	Ours
Mean	22.34	**14.38**
Median	16.68	**10.91**

Quantitative Results

Human-Chosen Active Entity

Error (Top 1)	NN+Sift-Flow	Kitani et al.	Ours
Mean	27.55	37.94	**21.55**
Median	23.77	30.23	**14.98**

VIRAT

Second Dataset

Conclusion

- Unsupervised method for prediction
- No explicit modeling of semantics
- Models appearance changes
- Code will be available!

Thank You!