Caltech

Towards Real-World Imitation Learning

Animation, Sports Analytics, Robotics, and More

Yisong Yue

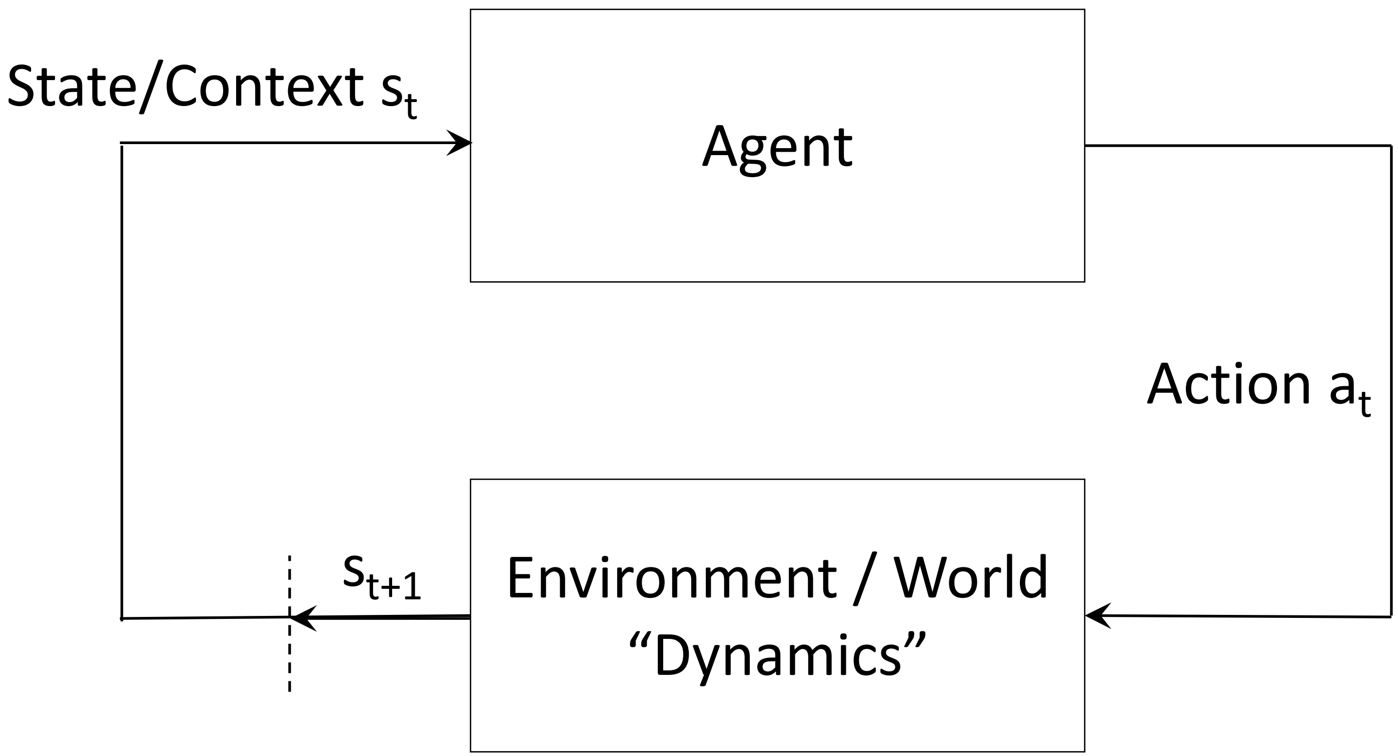
Policy/Controller Learning (Reinforcement & Imitation)

Goal: Find "Optimal" Policy

Imitation Learning:
Optimize imitation loss

Reinforcement Learning:
Optimize environmental reward

Learning-based Approach for Sequential Decision Making



Non-learning approaches include: optimal control, robust control, adaptive control, etc.

Imitation Learning Tutorial

https://sites.google.com/view/icml2018-imitation-learning/

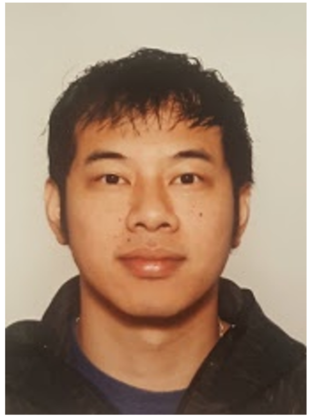
Yisong Yue

yyue@caltech.edu

@YisongYue

yisongyue.com

Hoang M. Le



Hoang.Le@Microsoft.com

@HoangMinhLe

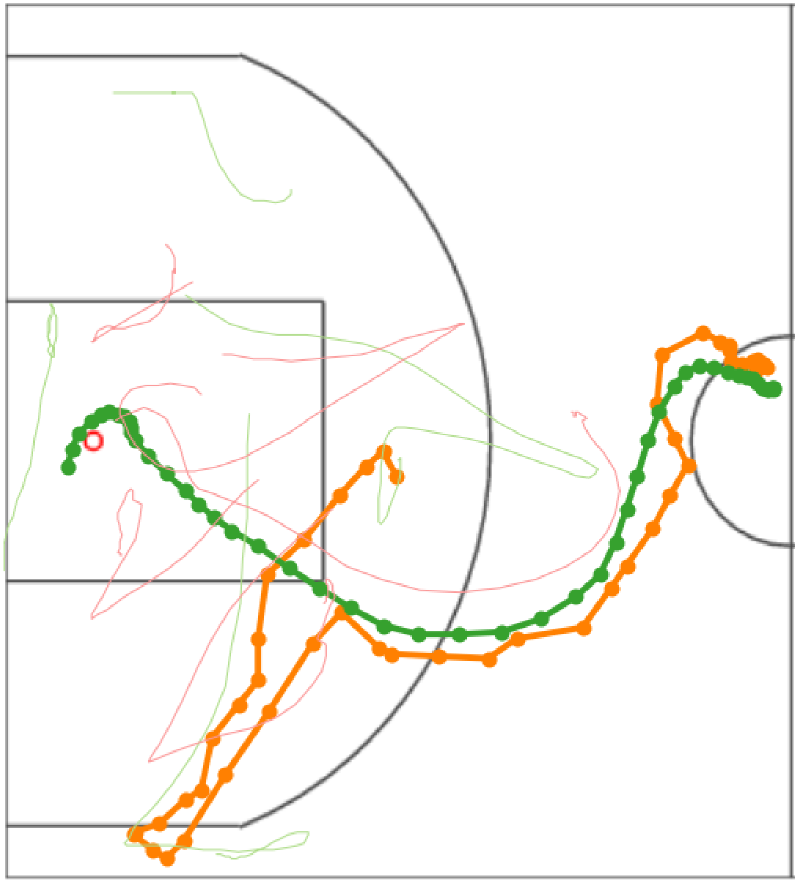
hoangle.info

Example #1: Basketball Player Trajectories

- s = location of players & ball
- a = next location of player

- Training set: $D = \{(\vec{s}, \vec{a})\}$
  - $\vec{s}$ = sequence of s
  - $\vec{a}$ = sequence of a

- **Goal:** learn $h(s) \rightarrow a$

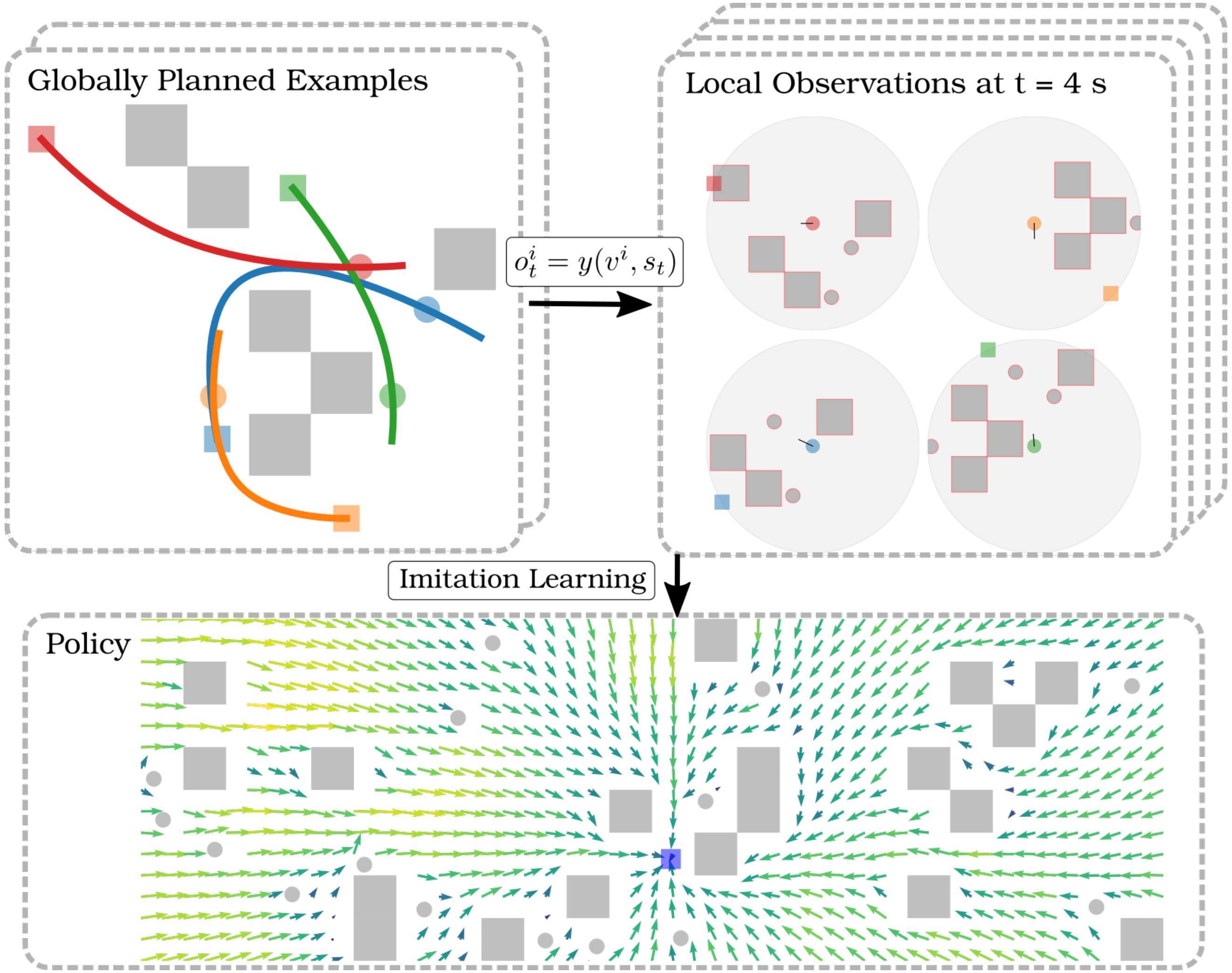


Example #2: Learning to Plan

- s = location of robots
- a = next location self robot

- Training set: $D = \{(\vec{s}, \vec{a})\}$
  - $\vec{s}$ = sequence of s
  - $\vec{a}$ = sequence of a

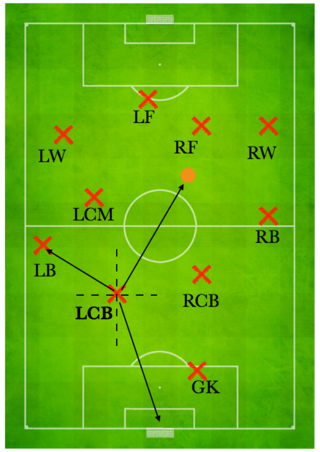
- **Goal:** learn $h(s) \rightarrow a$



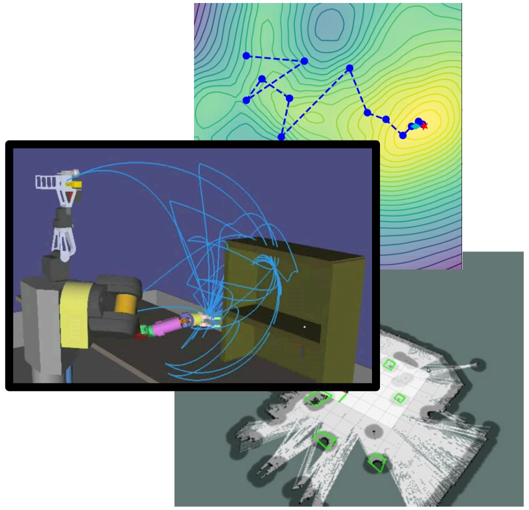
Towards Real-World Applications

Speech Animation

Side Constraints (Smoothness)



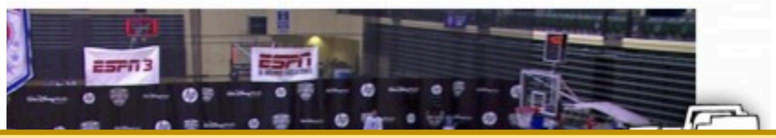
Coordinated Learning



Learning to Optimize

Towards Real-World Applications

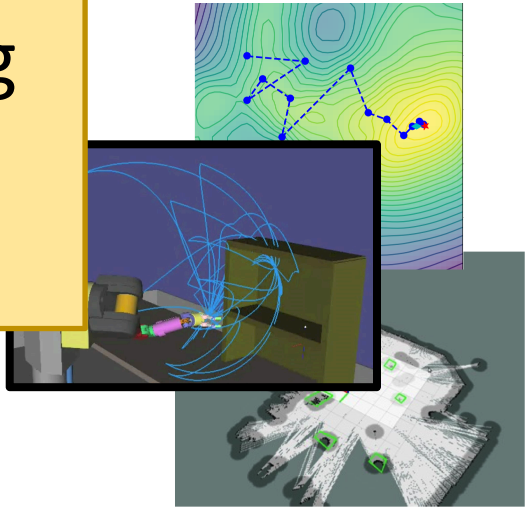
Speech Animation



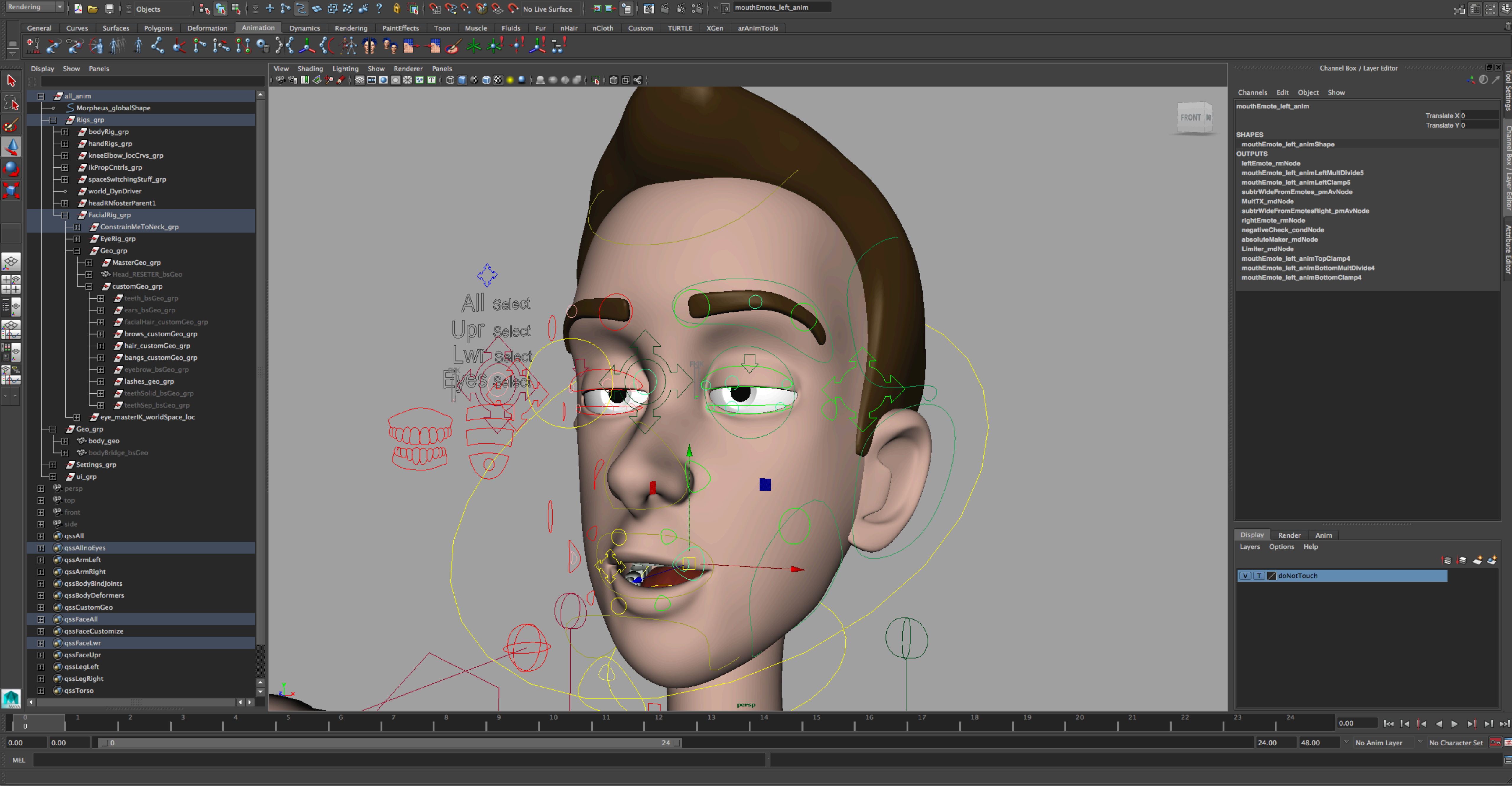
Research Goals:

- Abstractions for domain experts
- Better inductive bias
- Reductions to conventional learning
- Composable theoretical guarantees
- Works well in practice

Coordinated Learning



Learning to Optimize



Automatically generate high-frequency animations (lip syncing)

https://www.youtube.com/watch?v=9zL7qejW9fE

Prediction Task

Input sequence $X =< x_1, x_2, \ldots, x_{|x|} >$

Output sequence $Y =< y_1, y_2, \ldots, y_{|y|} > \ , y_t \in R^D$

Goal: learn predictor $h : X \rightarrow Y$

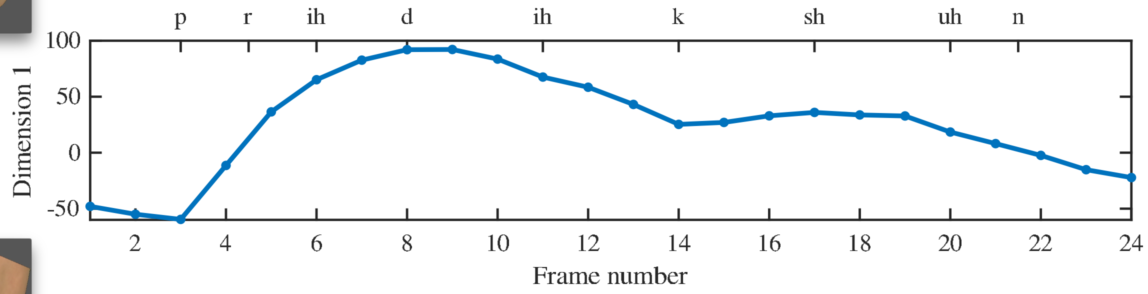
X

Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Token	-	p	p	r	ih	ih	d	d	ih	ih	ih	ih	k	k	sh	sh	sh	sh	uh	uh	n	-

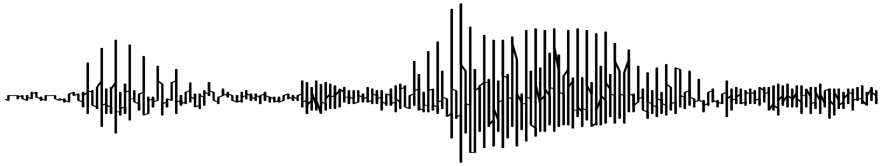
h

Phoneme sequence

Y



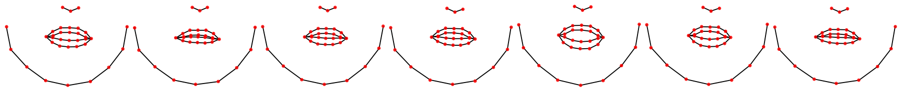
Sequence of face configurations



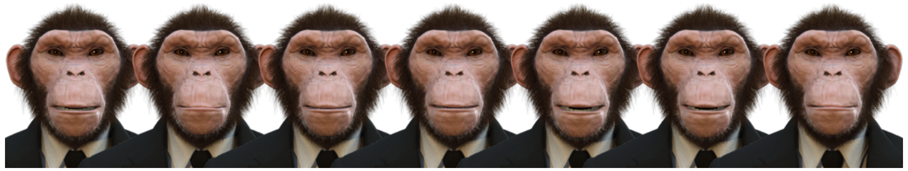
Input Audio

s s s s s ih ih ih g g r r ae ae ae ae f f f

Speech Recognition



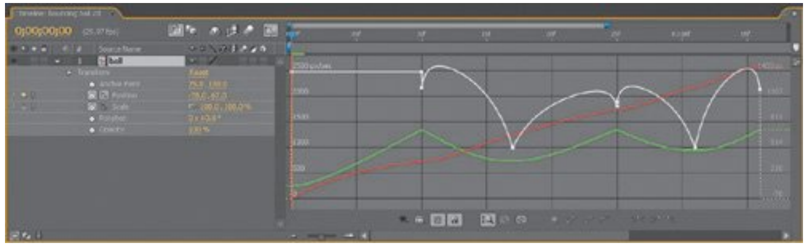
Speech Animation



(chimp rig courtesy of Hao Li)

Retargeting

E.g., [Sumner & Popovic 2004]



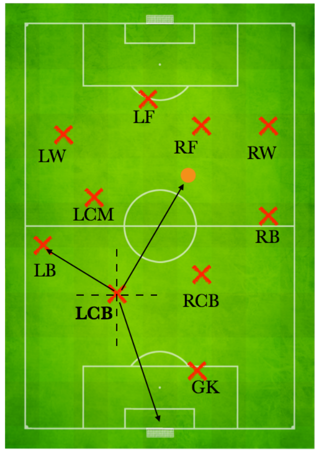
Editing

https://www.youtube.com/watch?v=9zL7qejW9fE

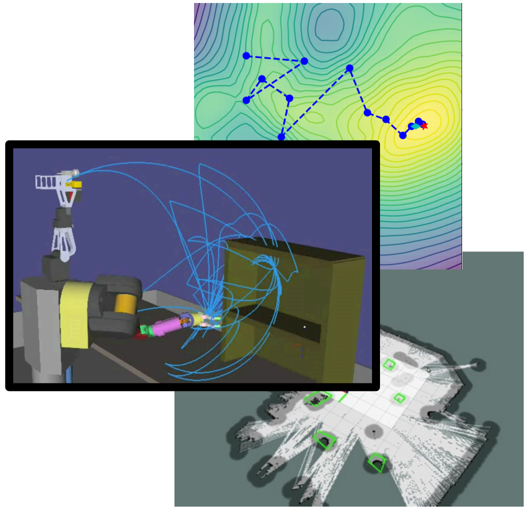
Towards Real-World Applications

Speech Animation

Side Constraints (Smoothness)

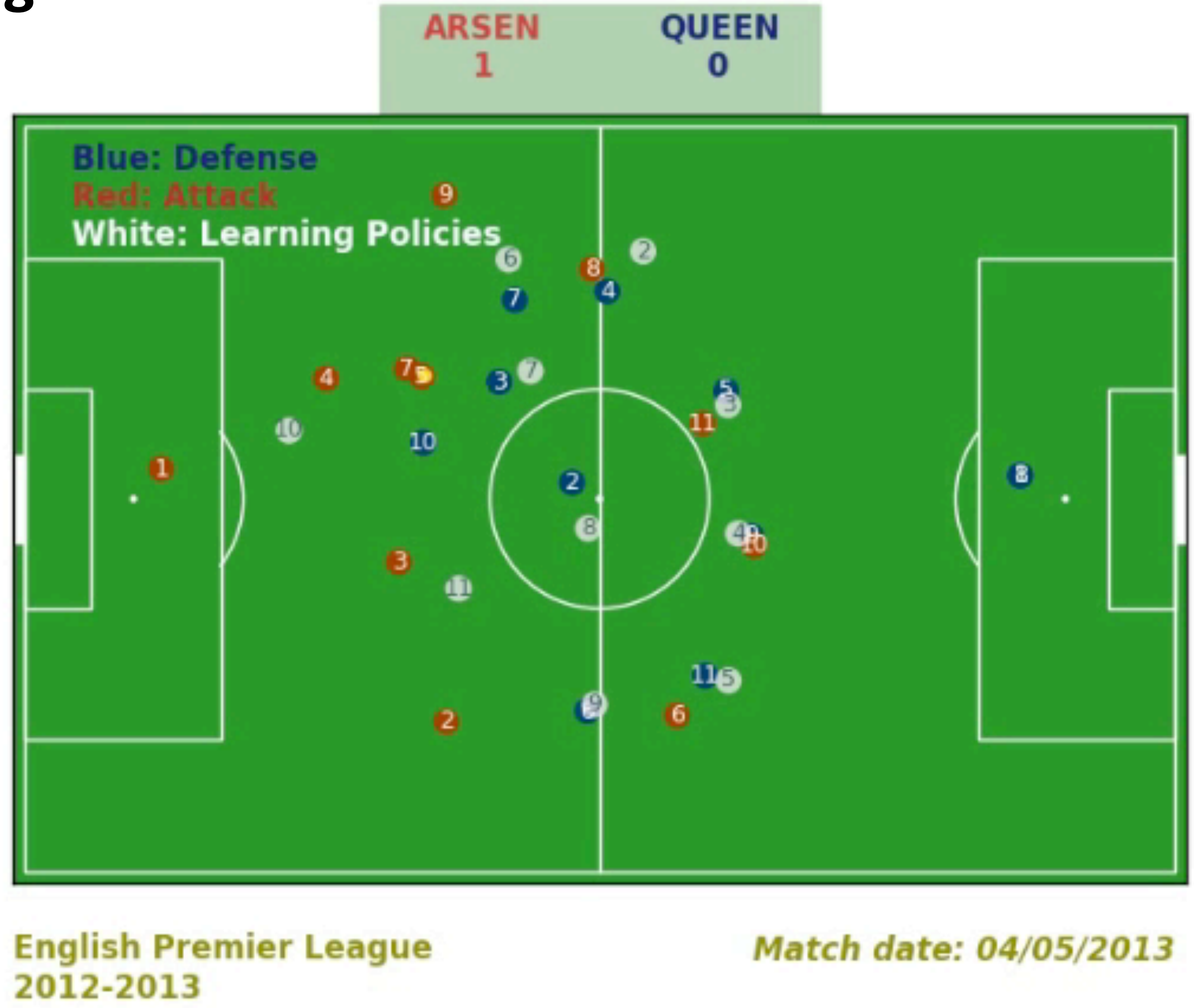


Coordinated Learning



Learning to Optimize

Data-Driven Ghosting

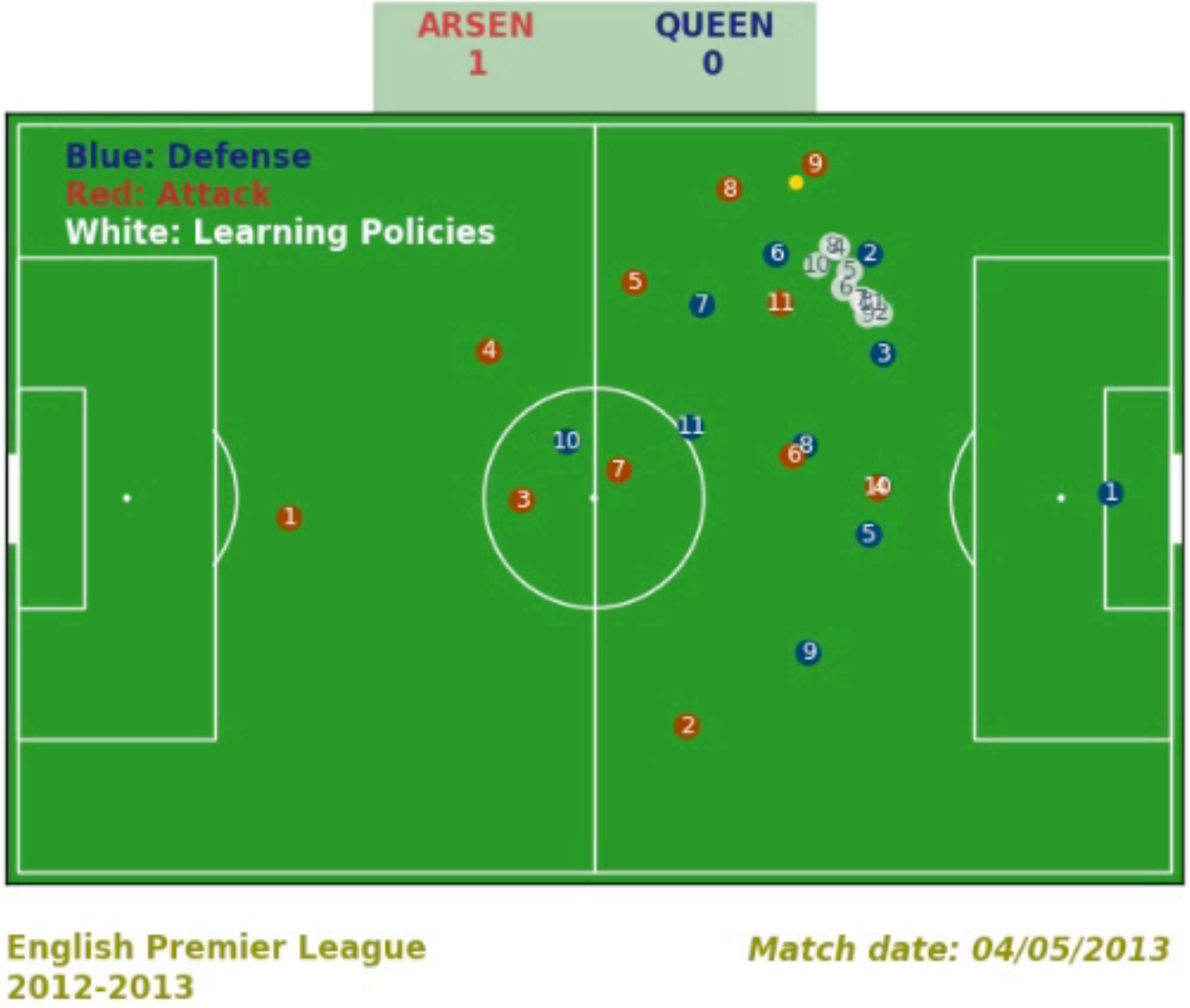


Hoang
Le

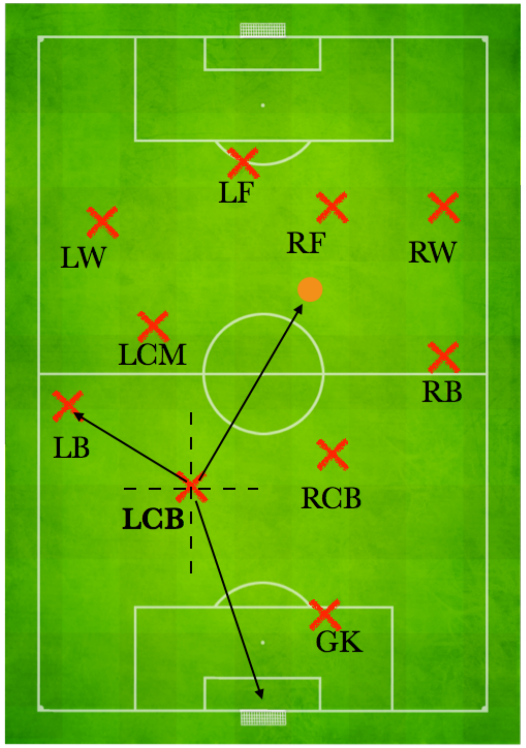
Data-Driven Ghosting using Deep Imitation Learning
Hoang Le et al. SSAC 2017

https://www.youtube.com/watch?v=WI-WL2cj0CA

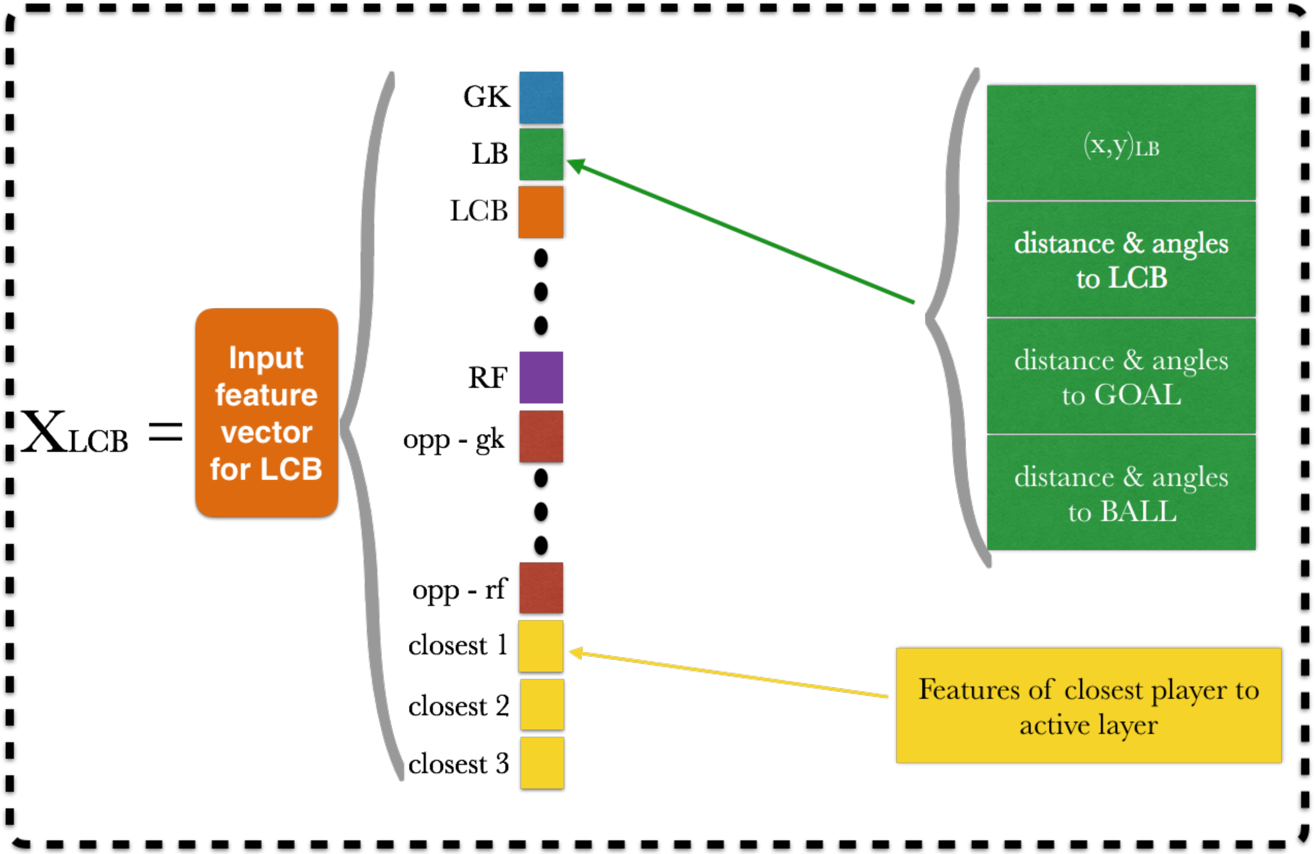
Naïve Baseline



State Representation



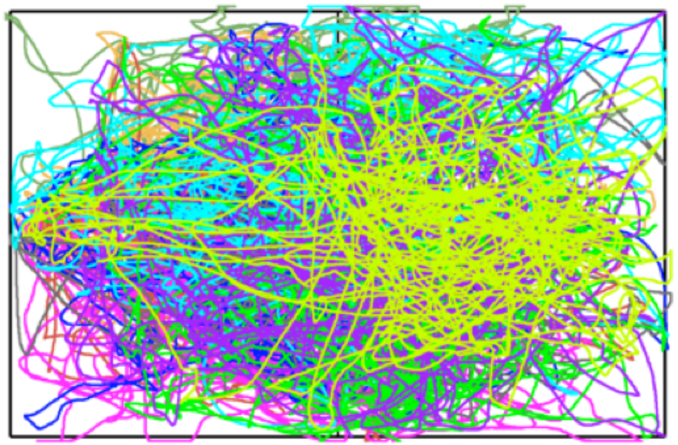
Geometric features computed



Data-Driven Ghosting using Deep Imitation Learning
Hoang Le, Peter Carr, Yisong Yue, Patrick Lucey. SSAC 2017

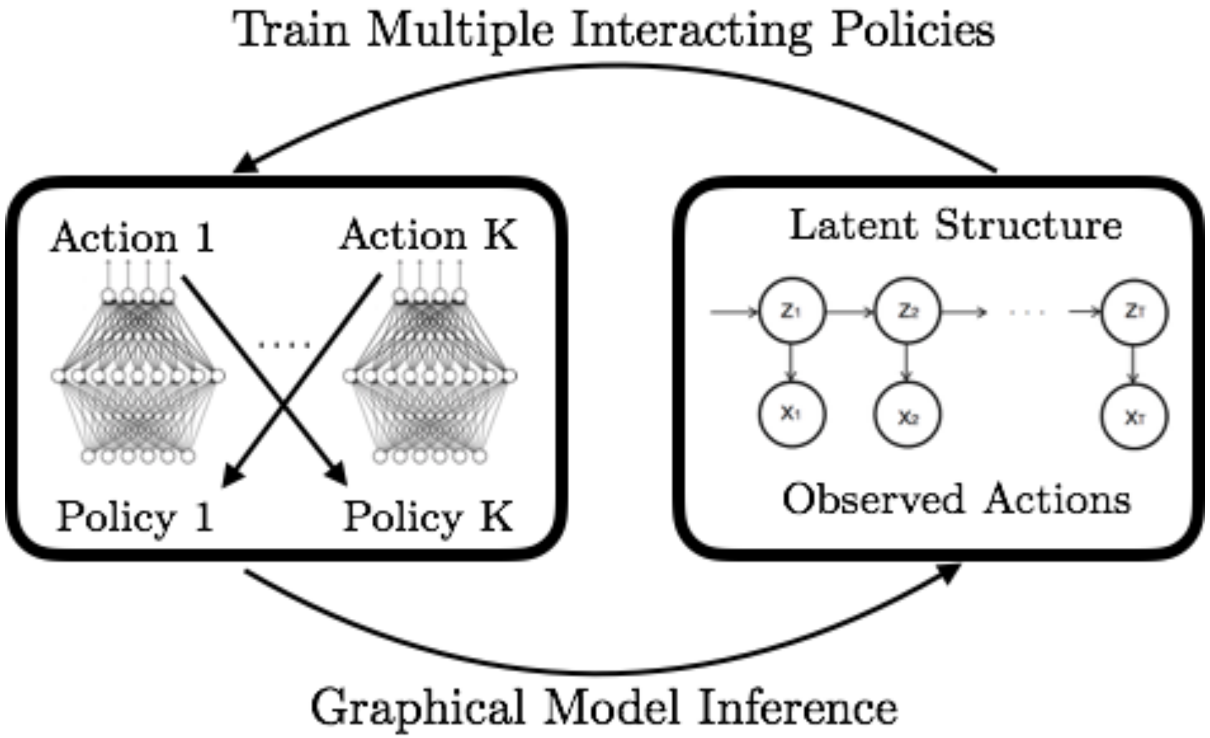
But Who Plays Which Role?

- All we get are trajectories!
  - Don’t know which belongs to which role.



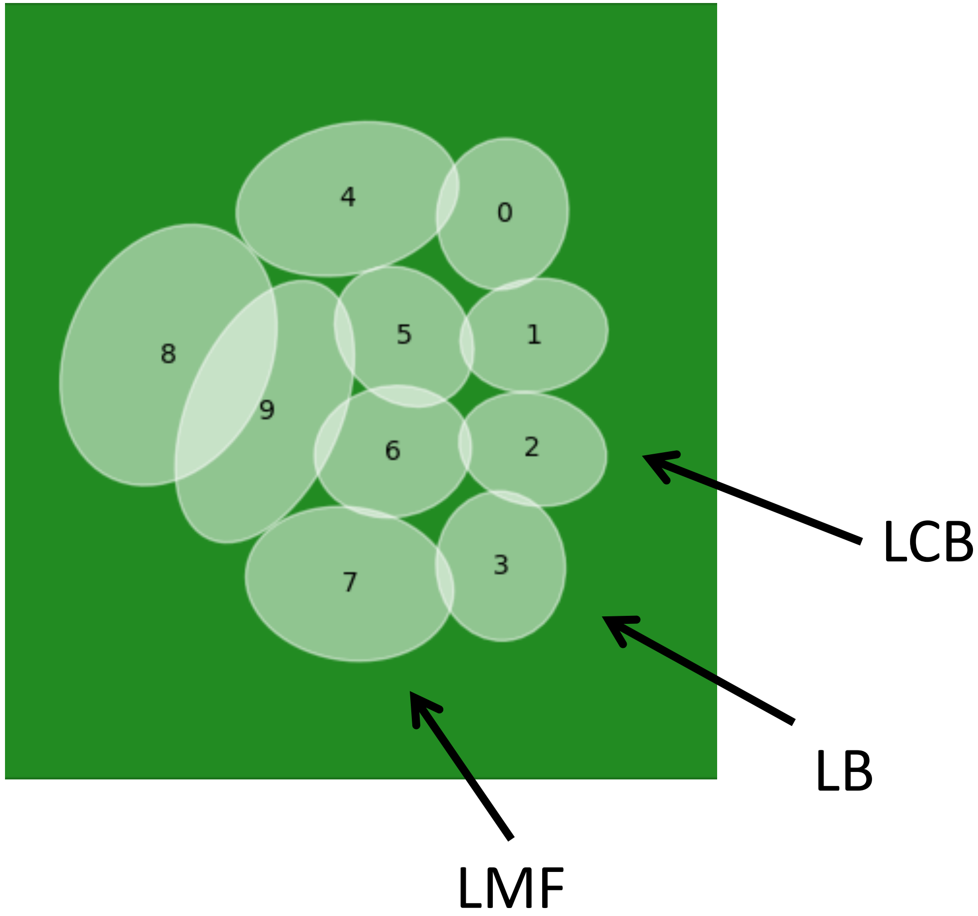
- Need to solve a permutation problem
  - **Naïve baseline ignores this!**

Coordination Model



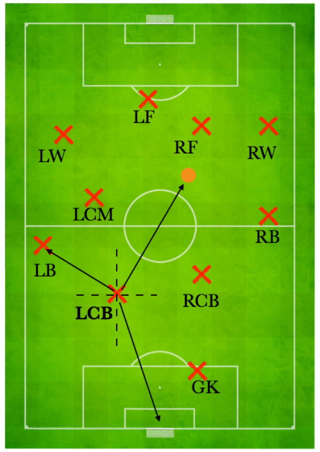
Coordinated Multi-Agent Imitation Learning
Hoang Le, Yisong Yue, Peter Carr, Patrick Lucey. ICML 2017

Learned Roles



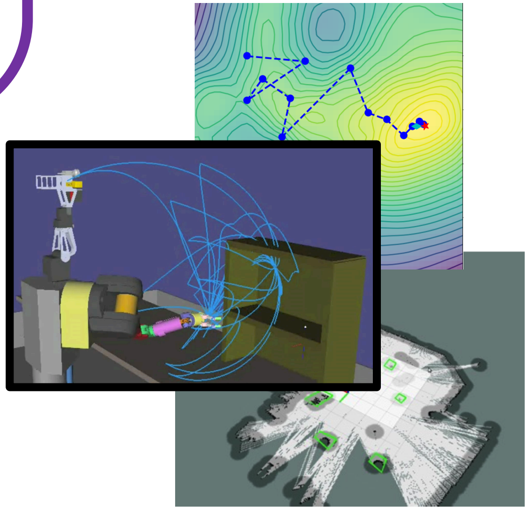
Towards Real-World Applications

Speech Animation

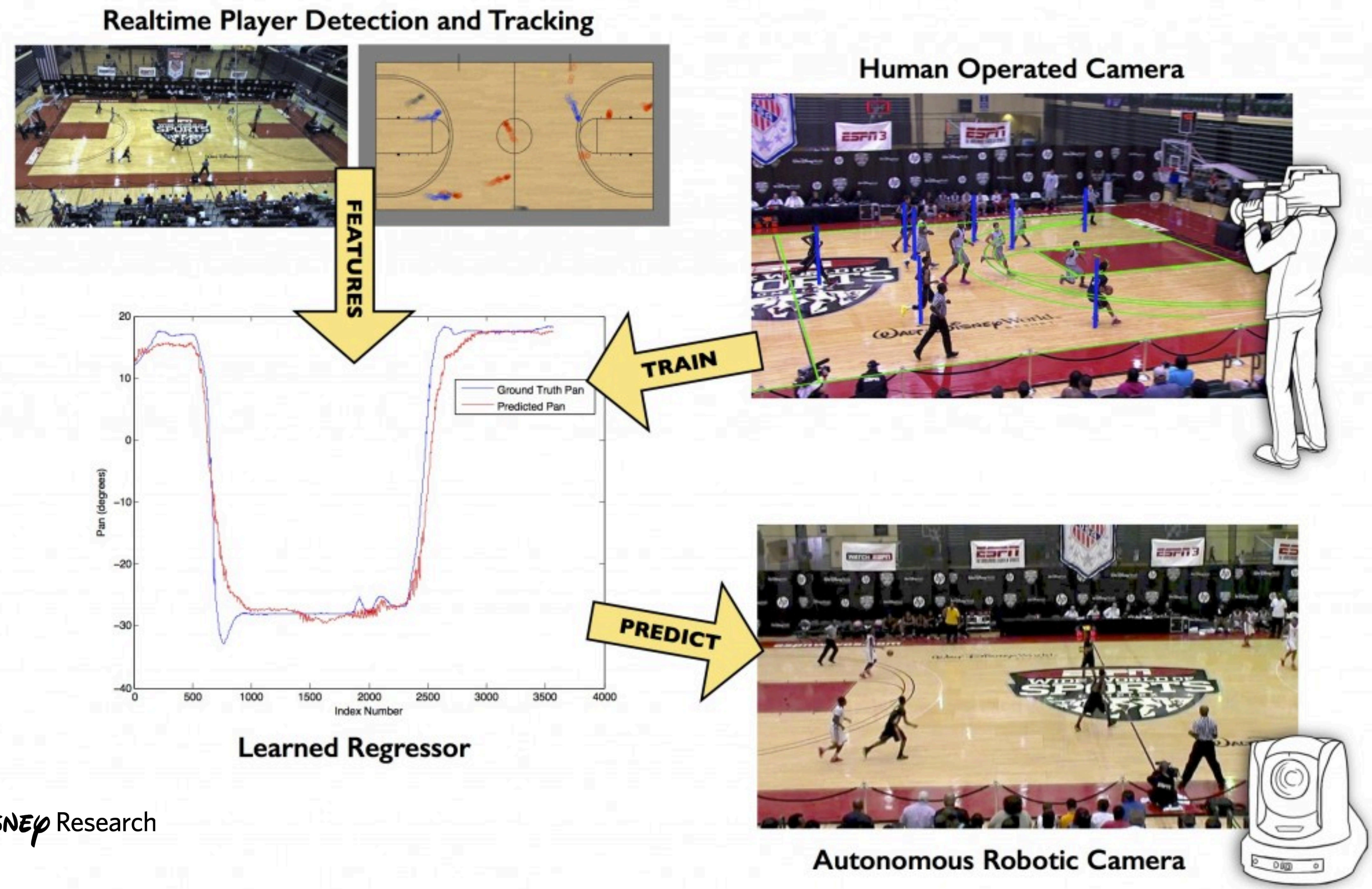


Coordinated Learning

Side Constraints (Smoothness)



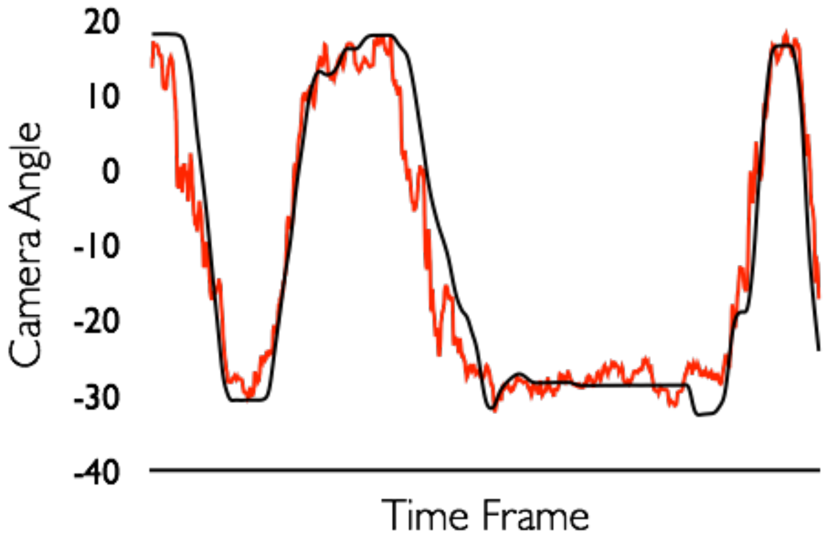
Learning to Optimize



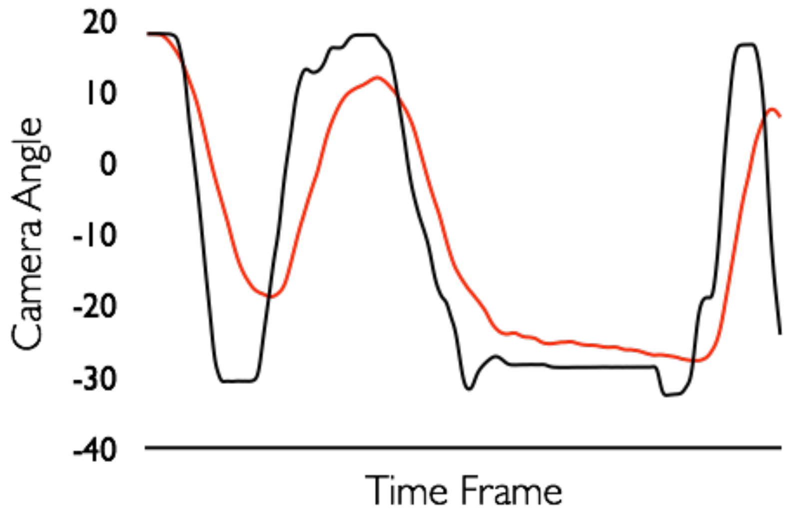
Disney Research

Naïve Approach

- Supervised learning of demonstration data
  - Train predictor per frame
  - Predict per frame



In practice, 2-step smoothing:

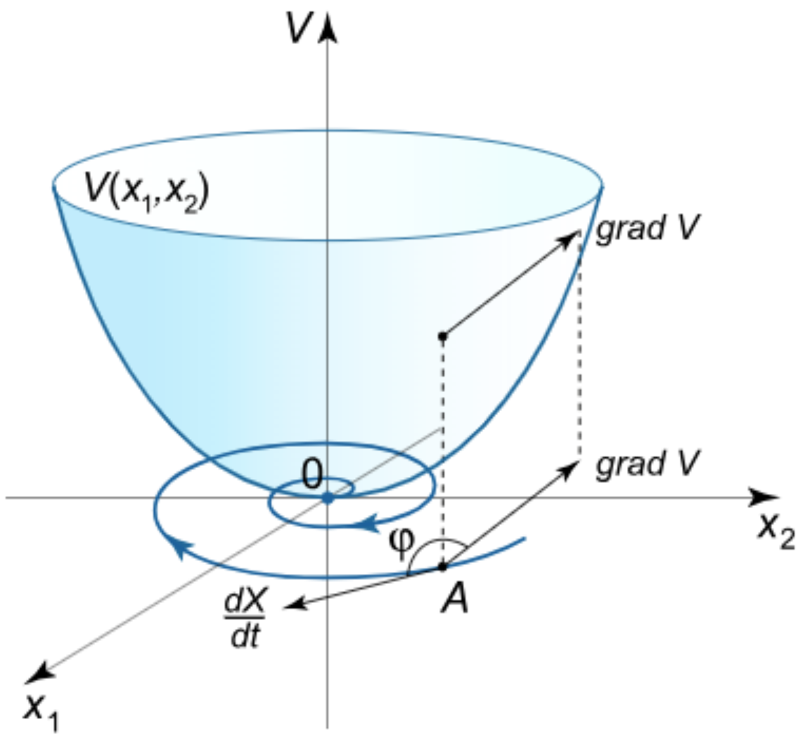


"I want to use deep learning to optimize the design, manufacturing and operation of our aircrafts. But I need some guarantees." -- Aerospace Director

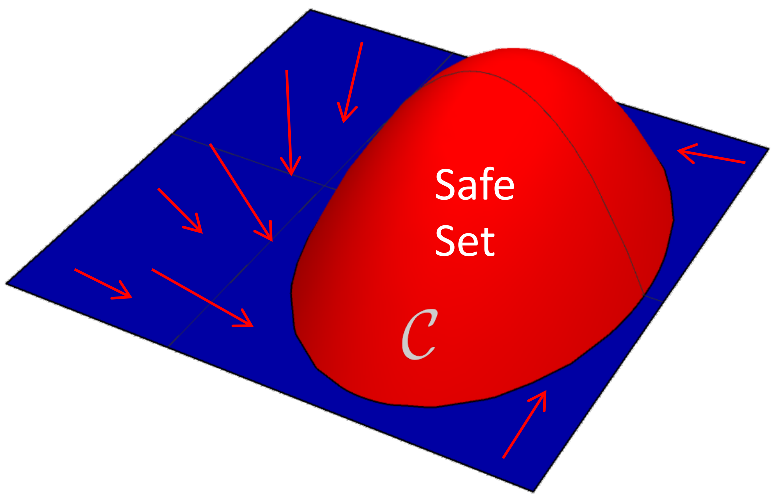
Behavioral Guarantees

Possibly Others:

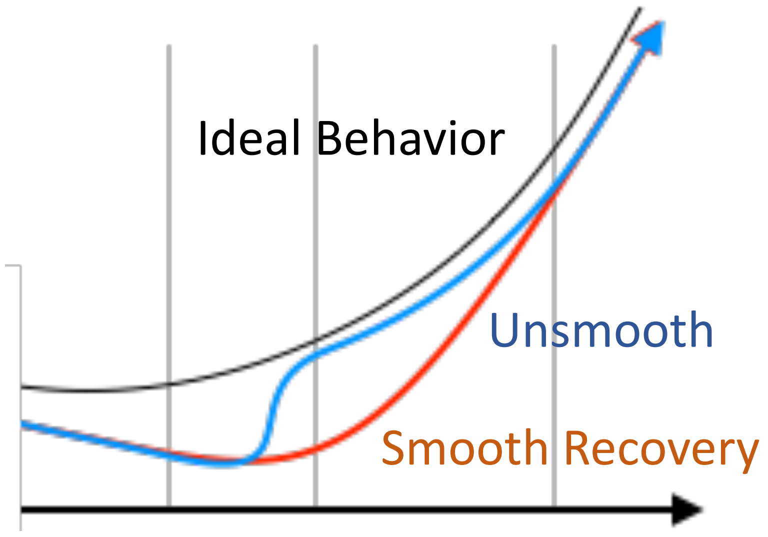
- Fairness
- Low-risk
- Temporal logic
- Etc…



Stability



Safety



Smoothness

Blended Policy Class (solution concept)

Hoang Le

Richard Cheng

Policy

Black Box Predictor

Model-Based Controller

w/ Certified Guarantees

$$argmin_{h=(f,g)} L(h) \quad \text{s.t.} \quad h(s) = argmin_{a'} (f(s) - a')^2 + \lambda (g(s) - a')^2$$

$$= \frac{f(s)+\lambda g(s)}{1+\lambda}$$

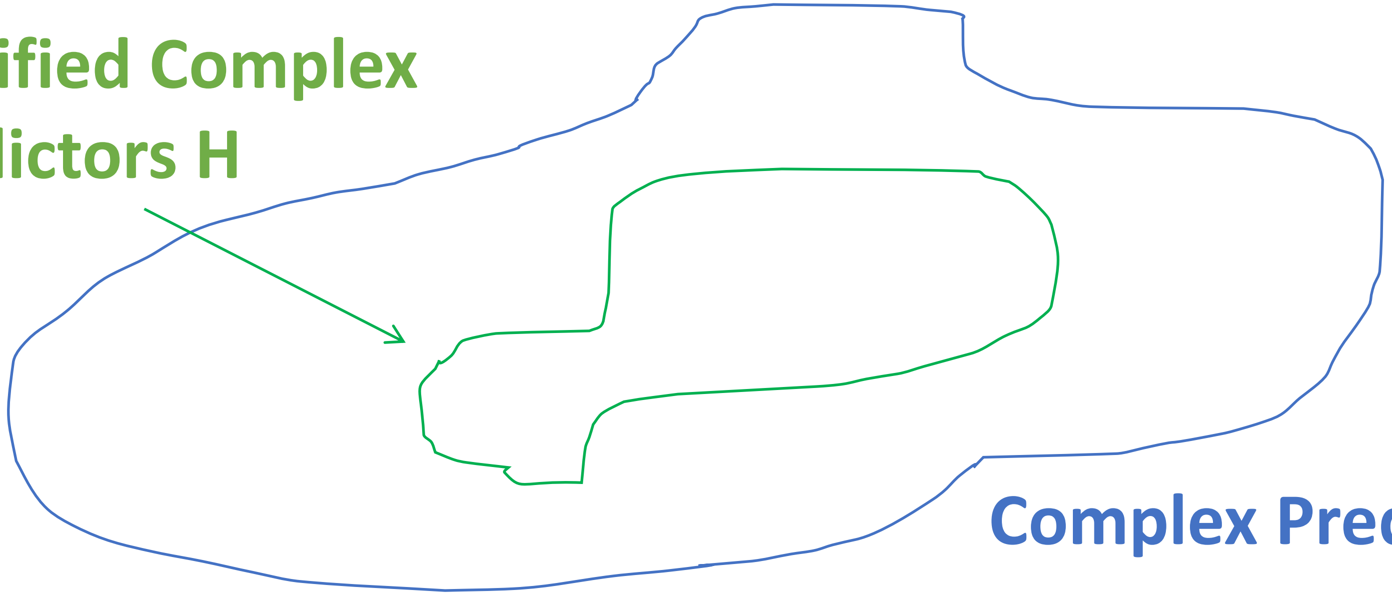
Smooth Imitation Learning for Online Sequence Prediction, Hoang Le, Andrew Kang, Yisong Yue, Peter Carr. ICML 2016

Control Regularization for Reduced Variance Reinforcement Learning, Richard Cheng, Abhinav Verma, et al. ICML 2019

Test-Time Functional Regularization

Hoang Le

Certified Complex Predictors H



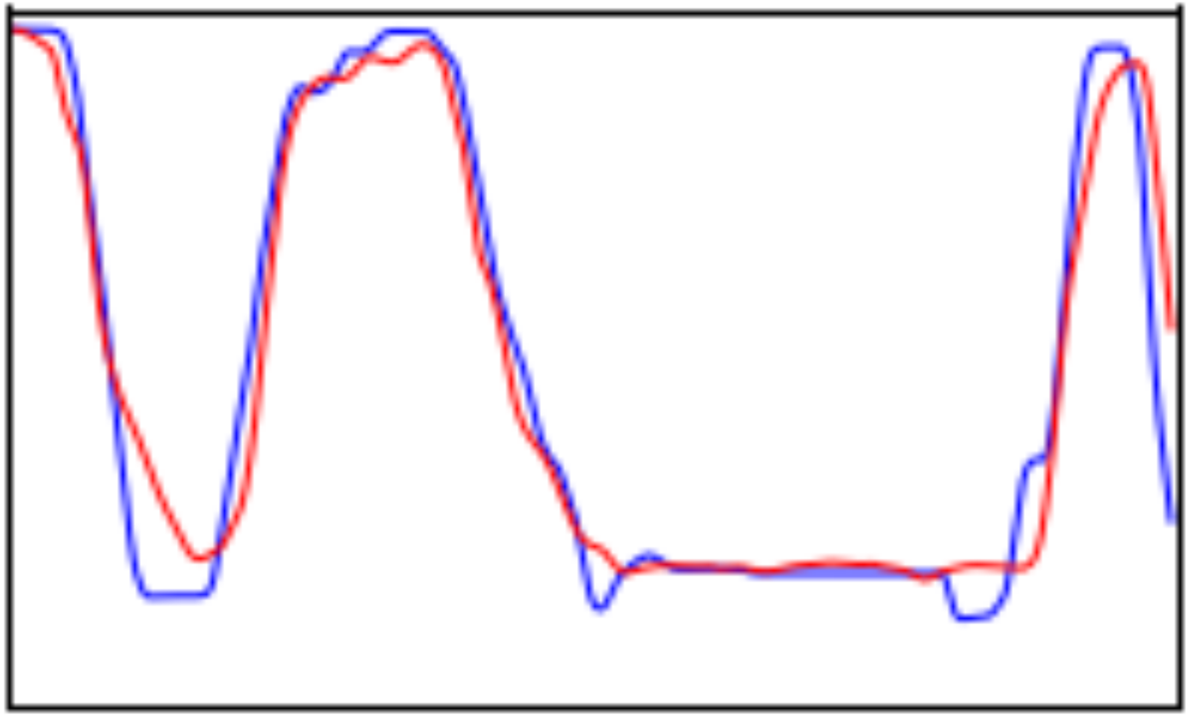
Complex Predictors F

$$\operatorname{argmin}_{h=(f,g)} L(h) \quad \text{s.t.} \quad h(s) = \operatorname{argmin}_{a'} (f(s) - a')^2 + \lambda (g(s) - a')^2$$

$$= \frac{f(s) + \lambda g(s)}{1+\lambda}$$

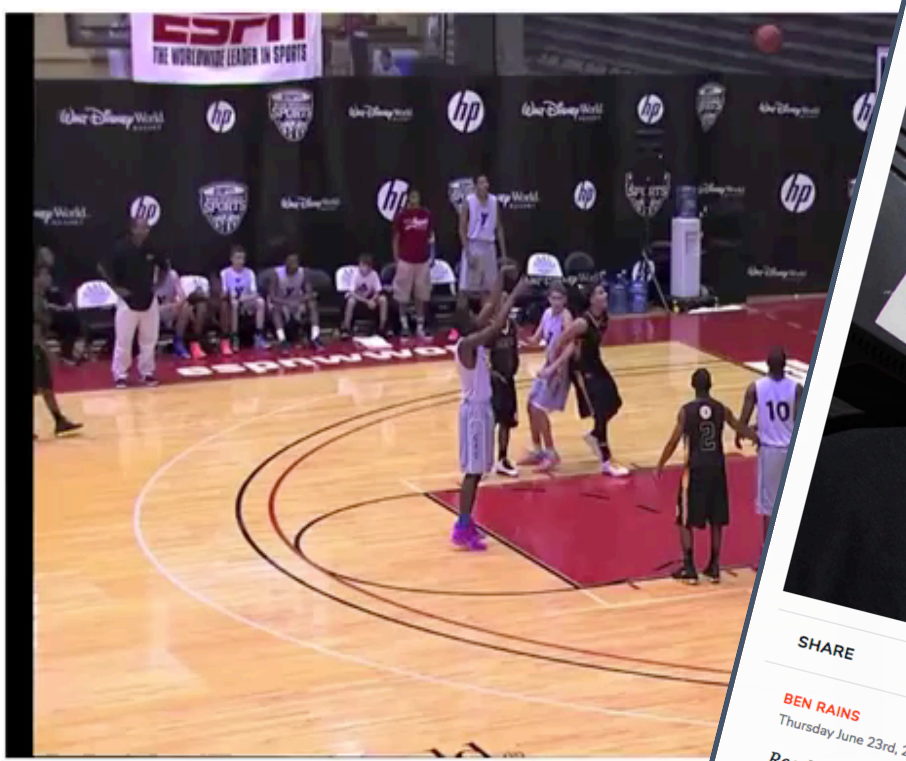
Smooth Imitation Learning for Online Sequence Prediction
Hoang Le, Andrew Kang, Yisong Yue, Peter Carr. ICML 2016

Our Results

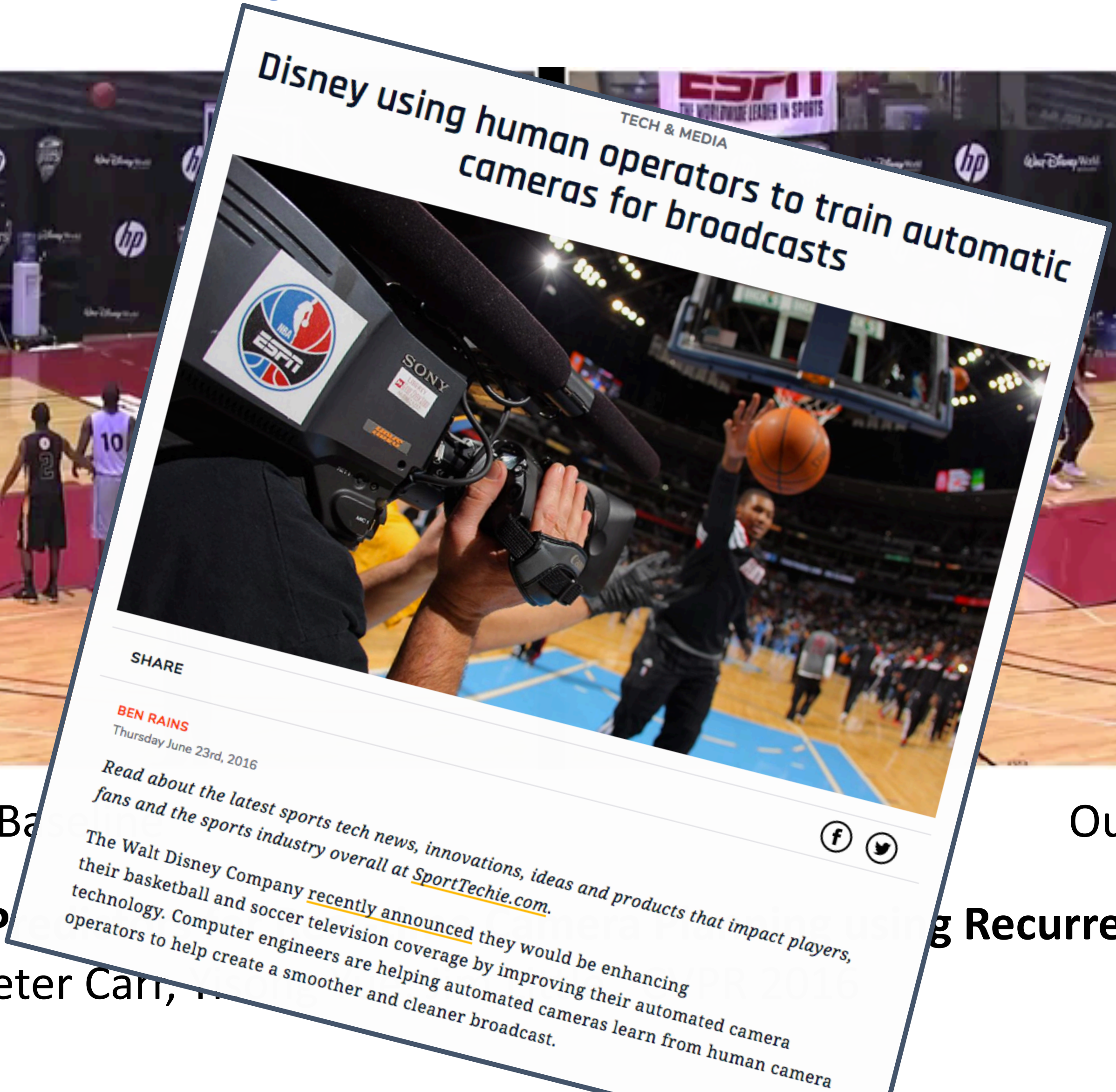


Smooth Imitation Learning for Online Sequence Prediction
Hoang Le, Andrew Kang, Yisong Yue, Peter Carr. ICML 2016

Qualitative Comparison



2-Step Ba



TECH & MEDIA

Disney using human operators to train automatic cameras for broadcasts

SHARE

BEN RAINS

Thursday June 23rd, 2016

Read about the latest sports tech news, innovations, ideas and products that impact players, fans and the sports industry overall at SportTechie.com.

The Walt Disney Company recently announced they would be enhancing their basketball and soccer television coverage by improving their automated camera technology. Computer engineers are helping automated cameras learn from human camera operators to help create a smoother and cleaner broadcast.

Our Approach

Learning Online Smooth P … g Recurrent Decision Trees

Jianhui Chen, Hoang Le, Peter Carr, …

Summary: Functional Regularization (cont.)

- Control methods => analytic guarantees (side guarantees)
- Blend w/ learning => improve precision/flexibility (real-world improvements)
- Preserve side guarantees (possibly relaxed)
- Interpret as functional regularization (speeds up learning)
- Other directions:

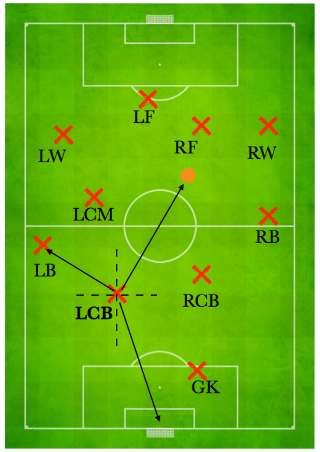
  Batch Policy Learning under Constraints
  Hoang Le, Cameron Voloshin, Yisong Yue. ICML 2019 (offline learning)

  Imitation-Projected Programmatic Reinforcement Learning
  Abhinav Verma, Hoang Le, Yisong Yue, Swarat Chaudhuri. NeurIPS 2019 (programmatic controllers)

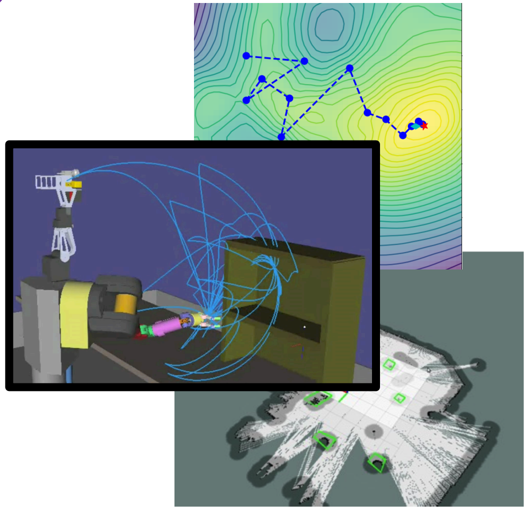
Towards Real-World Applications

Speech Animation

Side Constraints (Smoothness)



Coordinated Learning



Learning to Optimize

Optimization as Sequential Decision Making

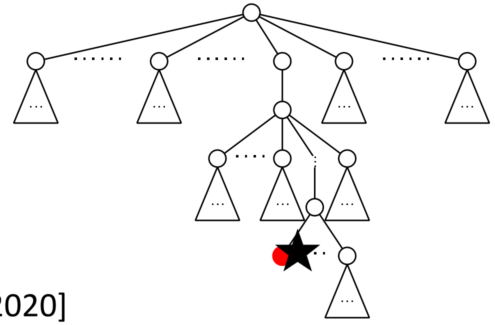
- Many Solvers are Sequential
  - Tree-Search
  - Greedy
  - Gradient Descent

- Can view solver as “agent” or “policy”
  - State = intermediate solution
  - Find a state with high reward (solution)
  - **Learn better local decision making**

Optimization as Sequential Decision Making

Learning to Search/Plan

- Discrete Optimization (Tree Search), Sparse Rewards
  - **Learning to Search via Retrospective Imitation** [arXiv]
  - **Co-training for Policy Learning** [UAI 2019]
  - **GLAS: Global-to-Local Safe Autonomy Synthesis** [RA-L 2020]
  - **A General Large Neighborhood Search Framework for Solving Integer Programs** [NeurIPS 2020]

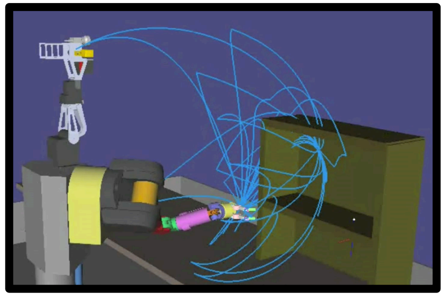


Jialin Song

Ben Riviere

Contextual Submodular Maximization

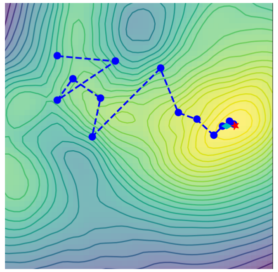
- Discrete Optimization (Greedy), Dense Rewards
  - **Learning Policies for Contextual Submodular Prediction** [ICML 2013]

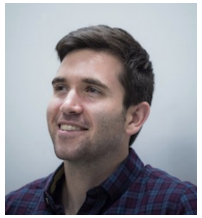


Stephane Ross

Learning to Infer

- Continuous Optimization (Gradient-style), Dense Rewards
  - **Iterative Amortized Inference** [ICML 2018]
  - **A General Method for Amortizing Variational Filtering** [NeurIPS 2018]
  - **Iterative Amortized Policy Optimization** [arXiv]



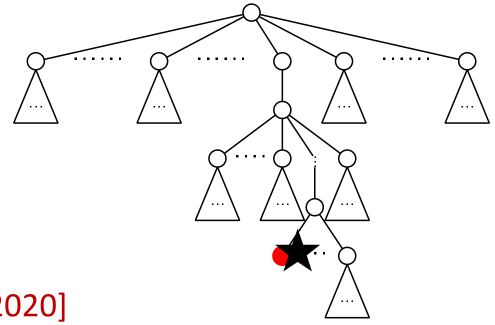


Joe Marino

Optimization as Sequential Decision Making

Learning to Search/Plan

- Discrete Optimization (Tree Search), Sparse Rewards
- **Learning to Search via Retrospective Imitation** [arXiv]
- **Co-training for Policy Learning** [UAI 2019]
- **GLAS: Global-to-Local Safe Autonomy Synthesis** [RA-L 2020]
- **A General Large Neighborhood Search Framework for Solving Integer Programs** [NeurIPS 2020]

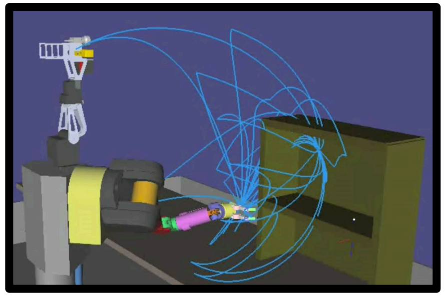


Jialin Song

Ben Riviere

Contextual Submodular Maximization

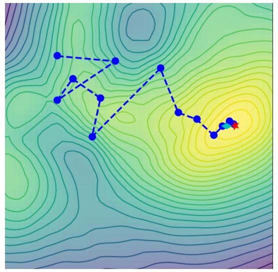
- Discrete Optimization (Greedy), Dense Rewards
- **Learning Policies for Contextual Submodular Prediction** [ICML 2013]

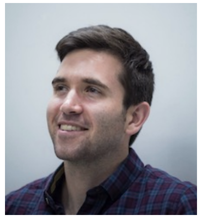


Stephane Ross

Learning to Infer

- Continuous Optimization (Gradient-style), Dense Rewards
- **Iterative Amortized Inference** [ICML 2018]
- **A General Method for Amortizing Variational Filtering** [NeurIPS 2018]
- **Iterative Amortized Policy Optimization** [arXiv]

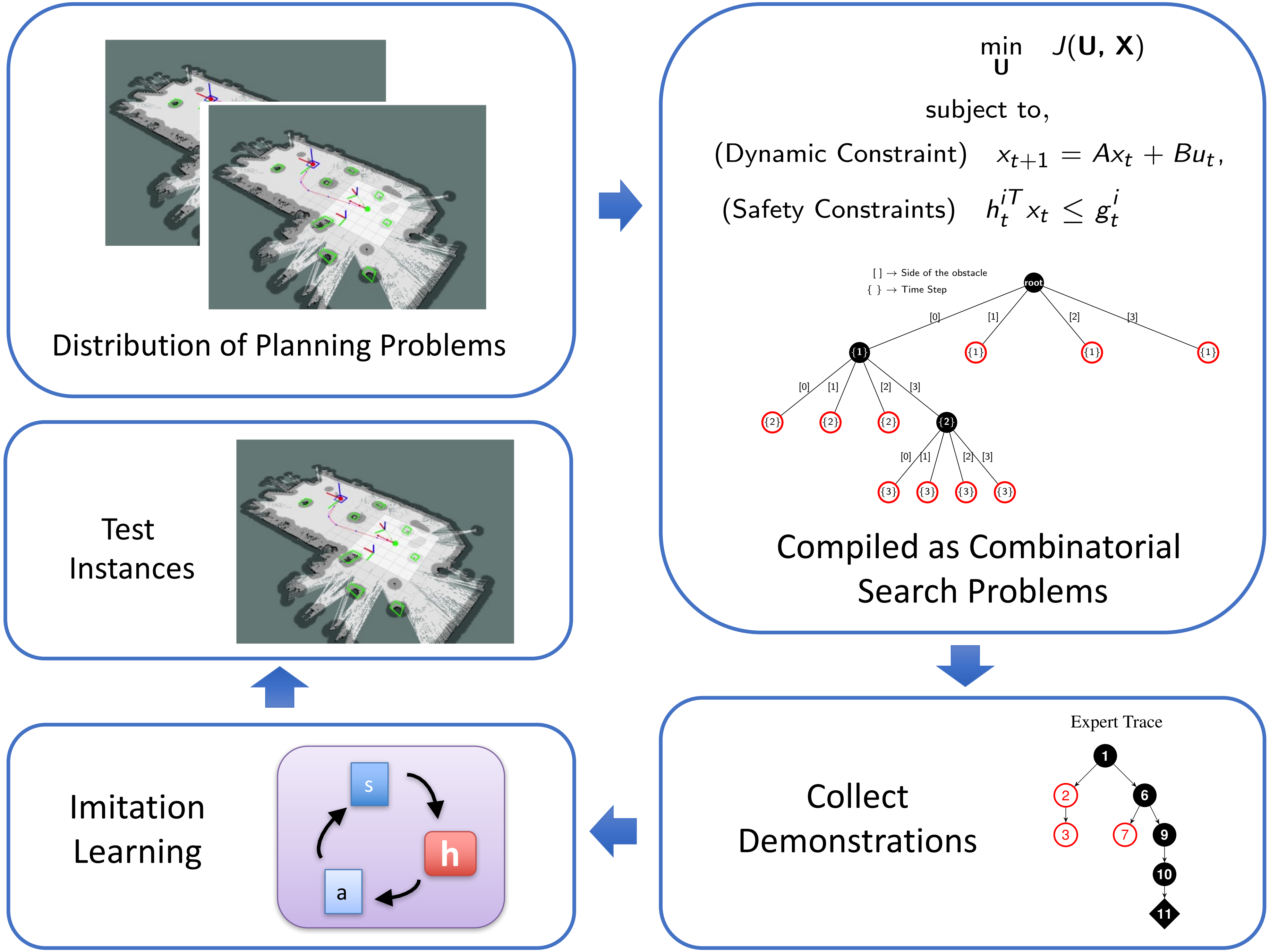




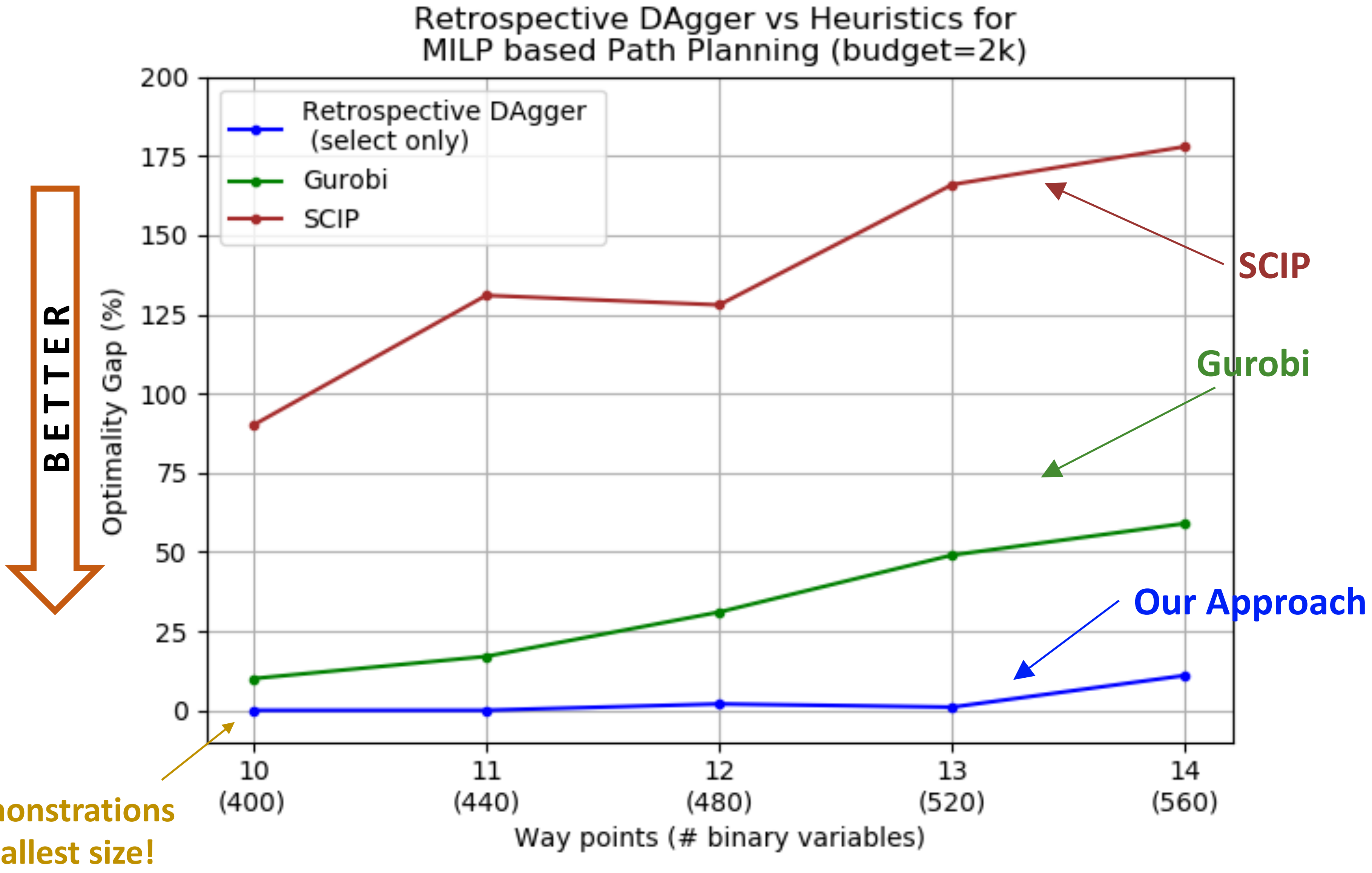
Joe Marino

Ravi Lanka

Jialin Song



Ravi Lanka

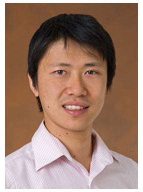


Jialin Song

Learning to Search via Retrospective Imitation, Jialin Song, Ravi Lanka, et al., arXiv

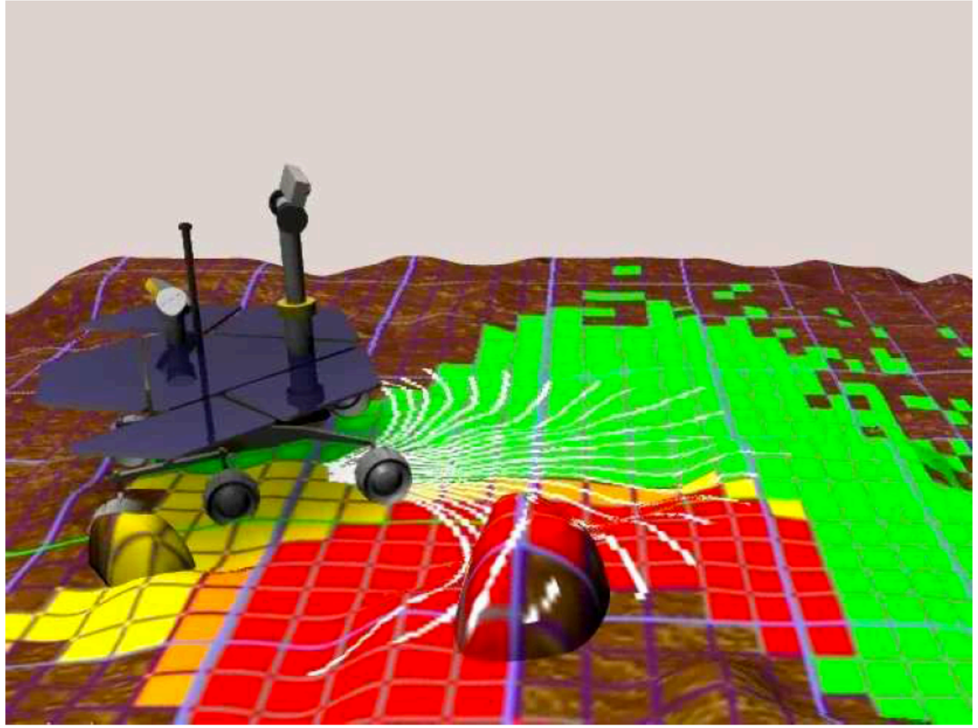
Ongoing: Integration with ENav

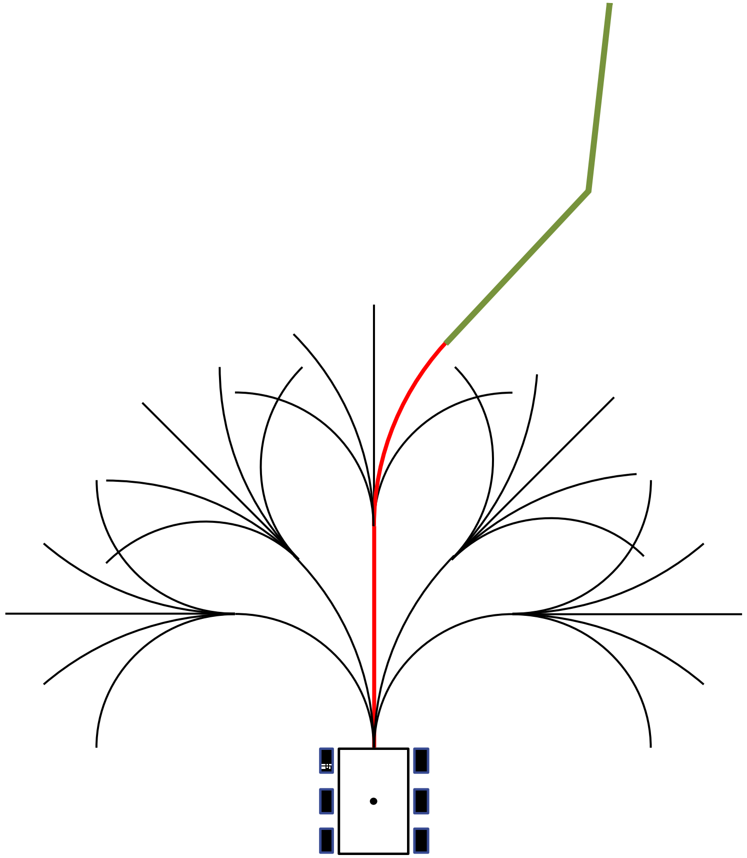

Shreyansh Daftry


Hiro Ono


Olivier Toupet


Neil Abcouwer





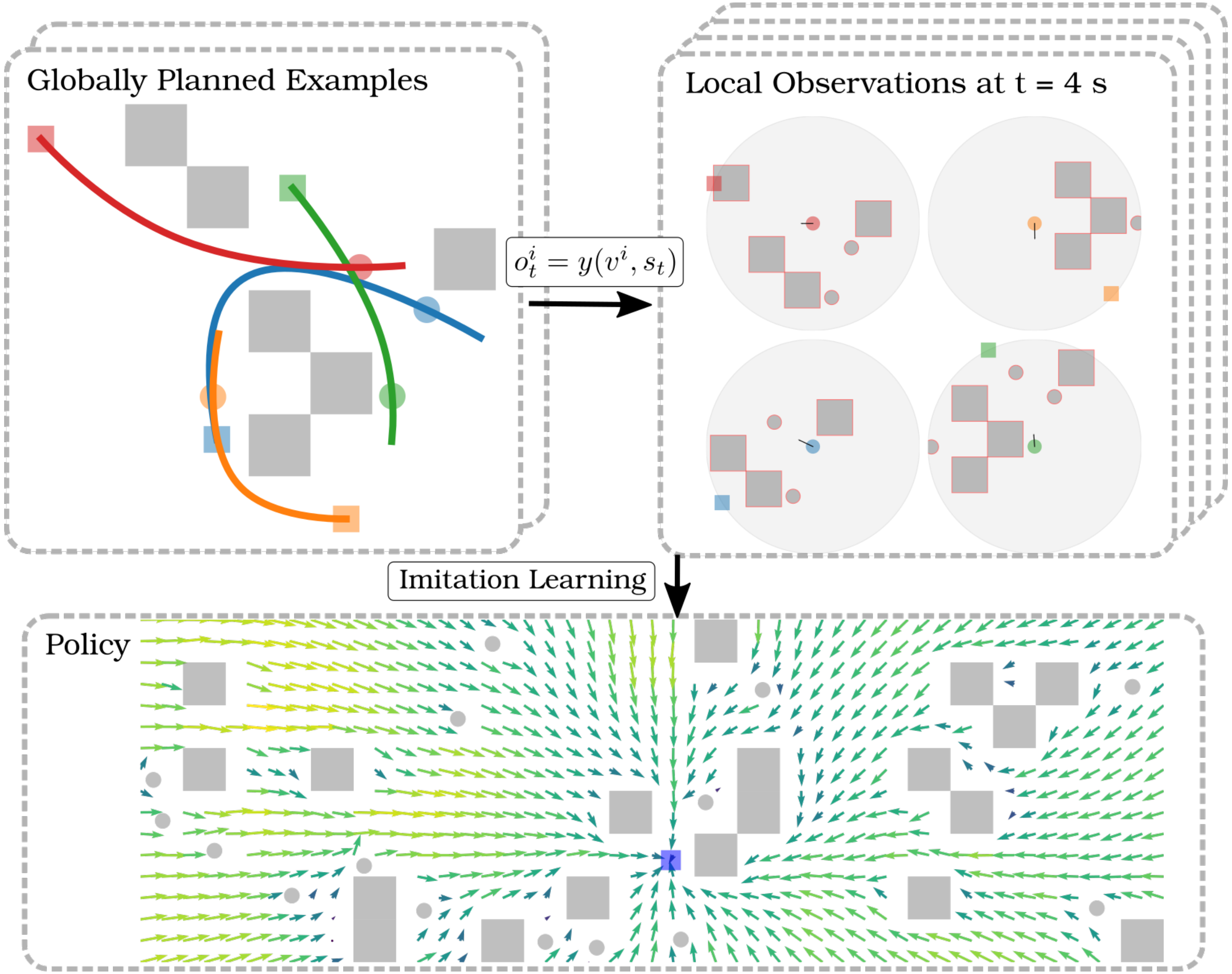
Machine Learning Based Path Planning for Improved Rover Navigation, Neil Abcouwer et al., (under review)

Learned Decentralized Planner

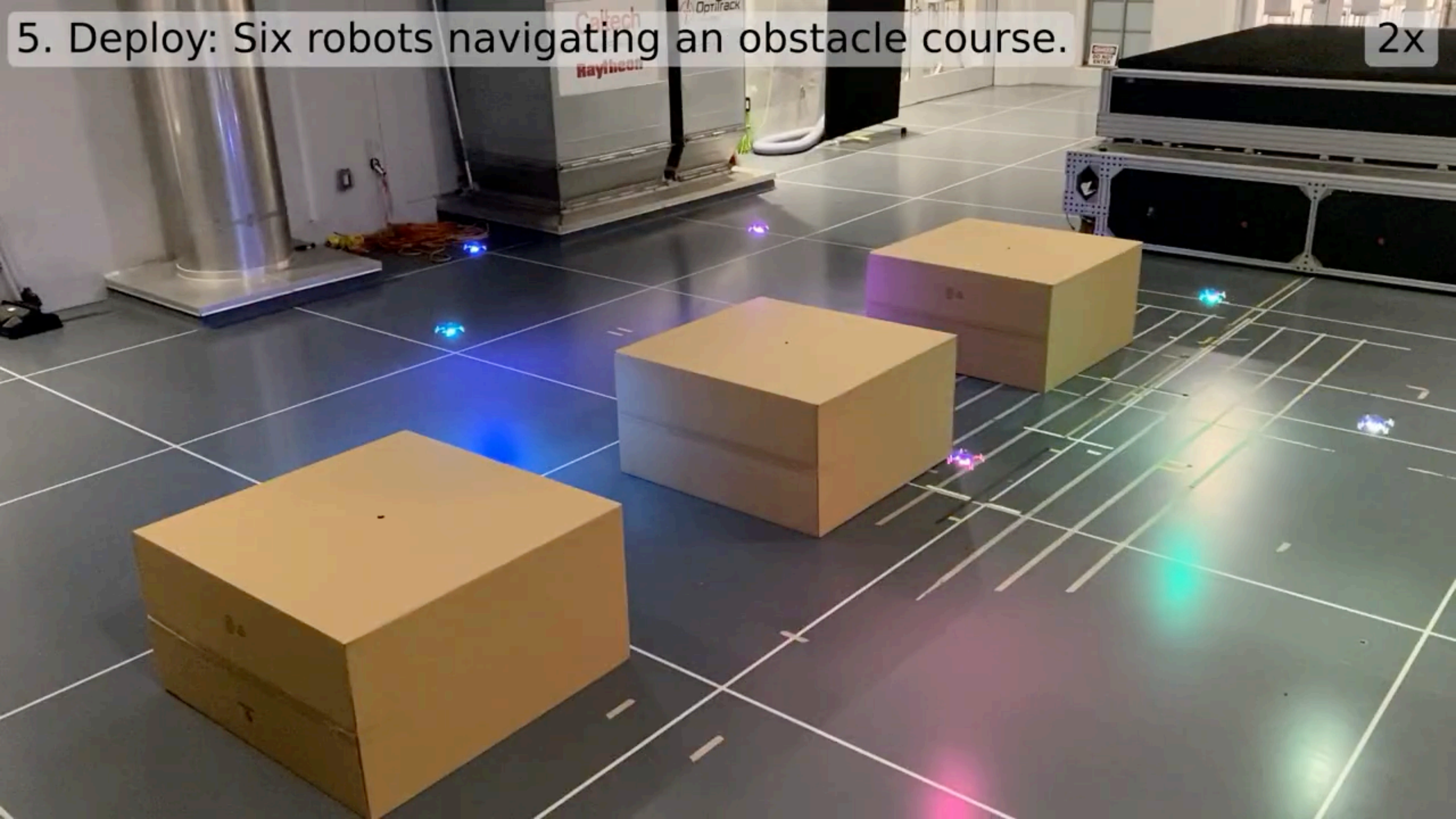
(enforcing safety)

Ben Riviere

Wolfgang Hoenig



GLAS: Global-to-Local Safe Autonomy Synthesis for Multi-Robot Motion Planning with End-to-End Learning,
Benjamin Rivière, et al., R-AL 2020


5. Deploy: Six robots navigating an obstacle course.
2x

Collaborators


Jialin Song

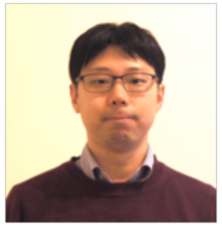

Ravi Lanka


Joe Marino


Hoang Le


Shreyansh Daftry


Wolfgang Hoenig


Taehwan Kim

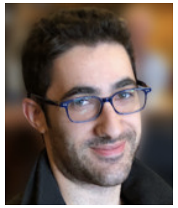

Sarah Taylor

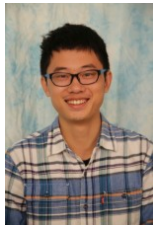

Abhinav Verma


Neil Abcouwer


Ben Riviere

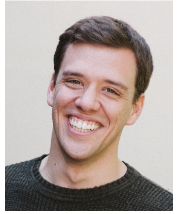

Stephane Ross

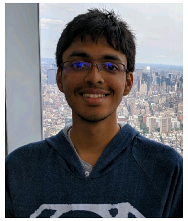

Cameron Voloshin

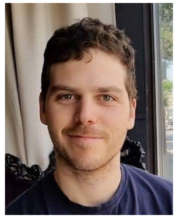

Robin Zhou


Jimmy Chen


Andrew Kang


Milan Cvitkovic


Siddarth Venkatraman


Alex Piche


Aadyot Bhatnagar


Albert Zhao


Debadeepta Dey


Alessandro Ialongo


Richard Cheng


Soon-Jo Chung


Joel Burdick

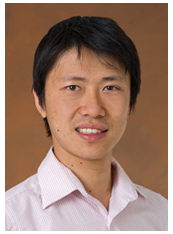

Gabor Orosz


Swarat Chaudhuri


Iain Matthews


Stephan Mandt


Drew Bagnell


Hiro Ono


Jim Little


Olivier Toupet


Peter Carr


Bistra Dilkina

References

A Decision Tree Framework for Spatiotemporal Sequence Prediction, Taehwan Kim et al., KDD 2015

A Deep Learning Approach for Generalized Speech Animation, Sarah Taylor et al., SIGGRAPH 2017`

Smooth Imitation Learning for Online Sequence Prediction, Hoang Le, et al., ICML 2016

Control Regularization for Reduced Variance Reinforcement Learning, Richard Cheng et al. ICML 2019

Batch Policy Learning under Constraints, Hoang Le, et al. ICML 2019

Learning Smooth Online Predictors for Real-Time Camera Planning using Recurrent Decision Trees, Jianhui Chen, et al., CVPR 2016

Imitation-Projected Programmatic Reinforcement Learning, Abhinav Verma, Hoang Le, et al., NeurIPS 2019

Learning to Search via Retrospective Imitation, Jialin Song, Ravi Lanka, et al., arXiv

Co-Training for Policy Learning, Jialin Song, Ravi Lanka, et al., UAI 2019

A General Large Neighborhood Search Framework for Solving Integer Programs, Jialin Song, et al., NeurIPS 2020

GLAS: Global-to-Local Safe Autonomy Synthesis for Multi-Robot Motion Planning with End-to-End Learning, Benjamin Rivière, et al., R-AL 2020

Learning Policies for Contextual Submodular Optimization, Stephane Ross et al., ICML 2013

Iterative Amortized Inference, Joe Marino et al., ICML 2018

A General Framework for Amortizing Variational Filtering, Joe Marino et al, NeurIPS 2018

Iterative Amortized Policy Optimization, Joe Marino et al., arXiv

Machine Learning Based Path Planning for Improved Rover Navigation, Neil Abcouwer et al., (under review)