dolcit

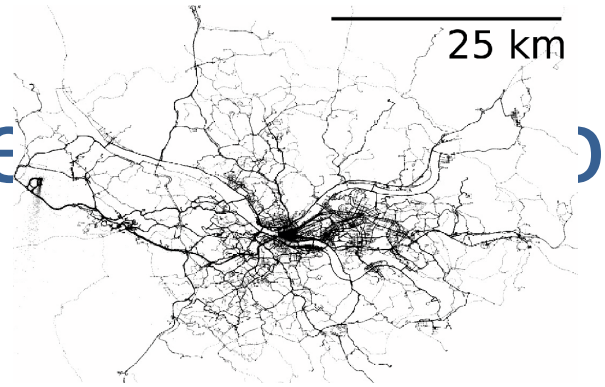
Caltech

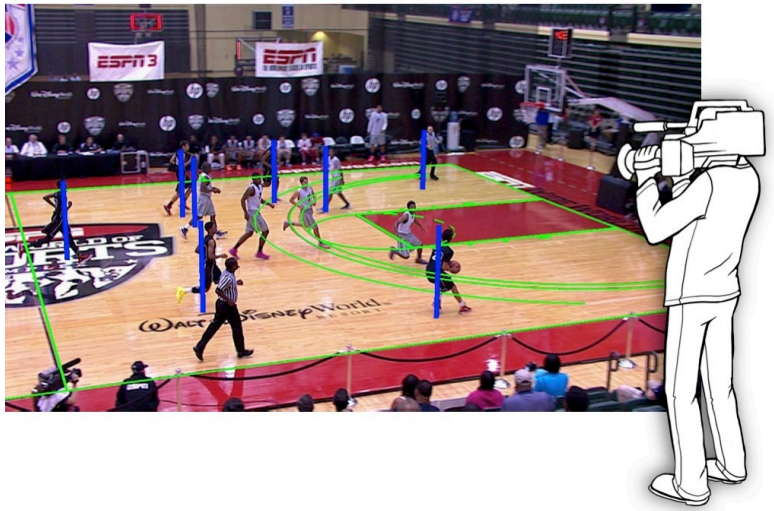
New Frontiers in Imitation Learning

Yisong Yue


LUNGE


Caltech
WindShape
GALCIT


25 km

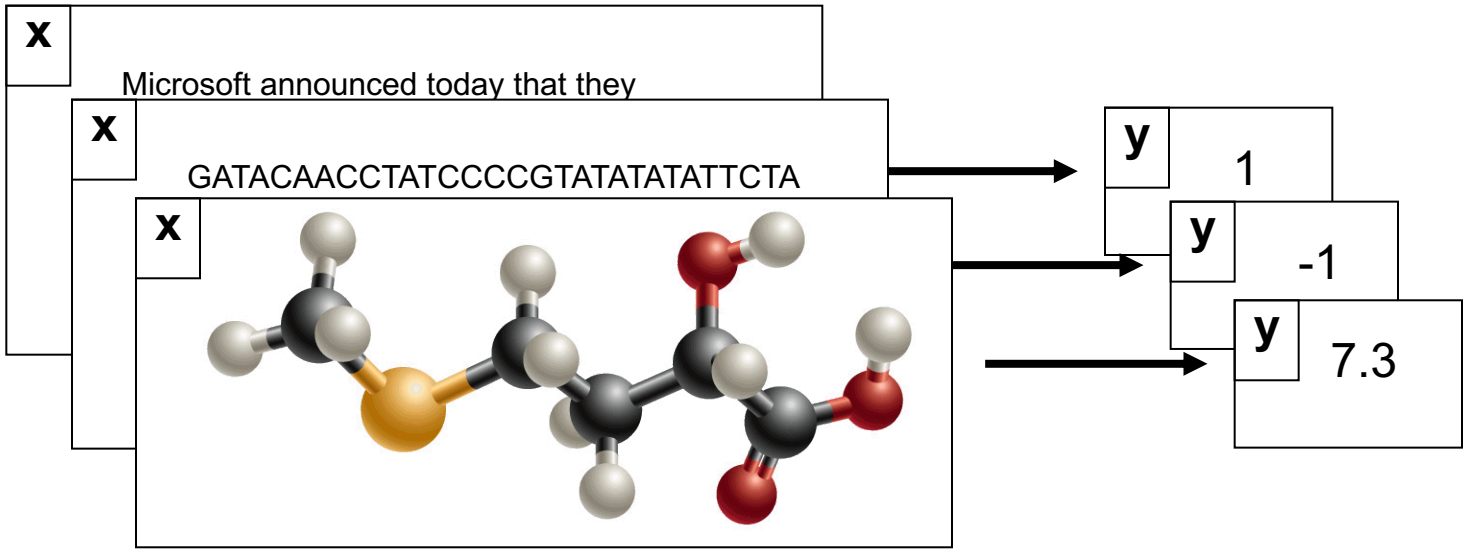

ESPN3
ESPN

Warm Up: Supervised Learning

- Find function from input space *X* to output space *Y*

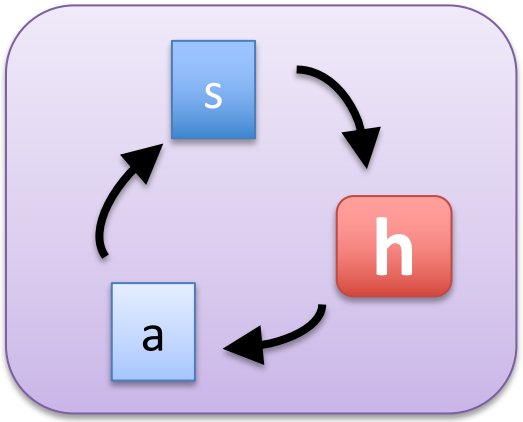
$$h : X \longrightarrow Y$$

such that the prediction error is low.



Imitation Learning

- Input:
  - Sequence of contexts/states:
- Predict:
  - Sequence of actions
- Learn Using:
  - Sequences of demonstrated actions

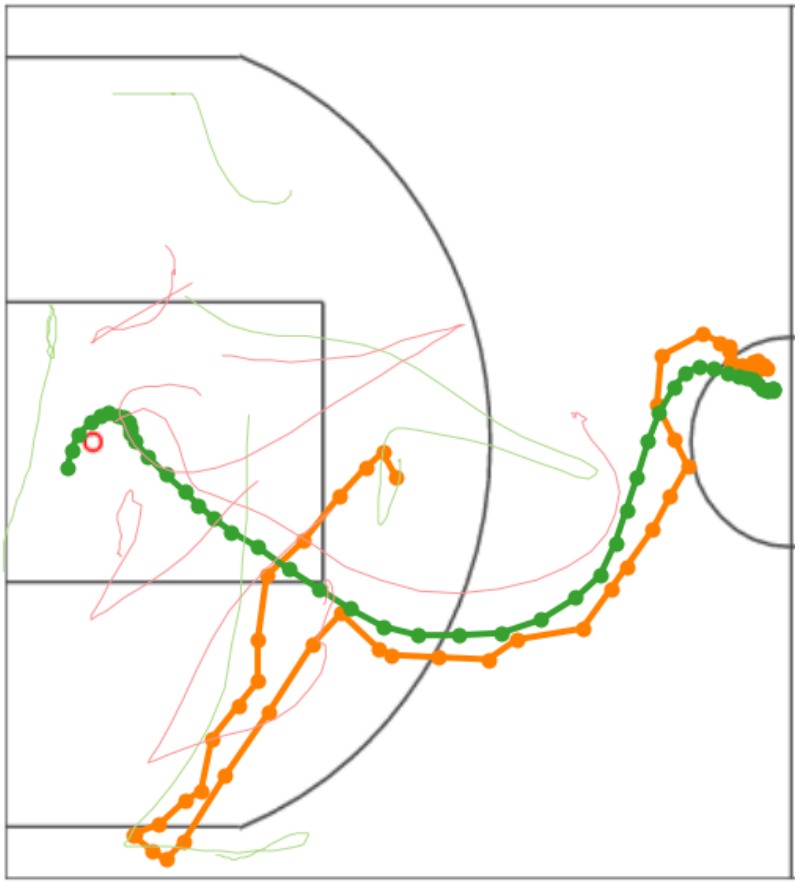


Example: Basketball Player Trajectories

- 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 $h(s) \rightarrow a$



Imitation Learning Tutorial

(ICML 2018)

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

Yisong Yue

yyue@caltech.edu

@YisongYue

yisongyue.com

Hoang M. Le

hmle@caltech.edu

@HoangMinhLe

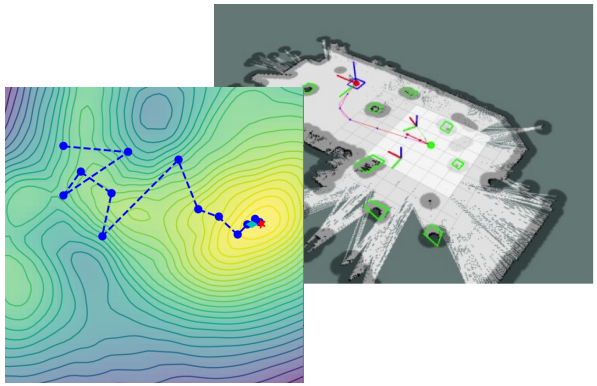
hoangle.info

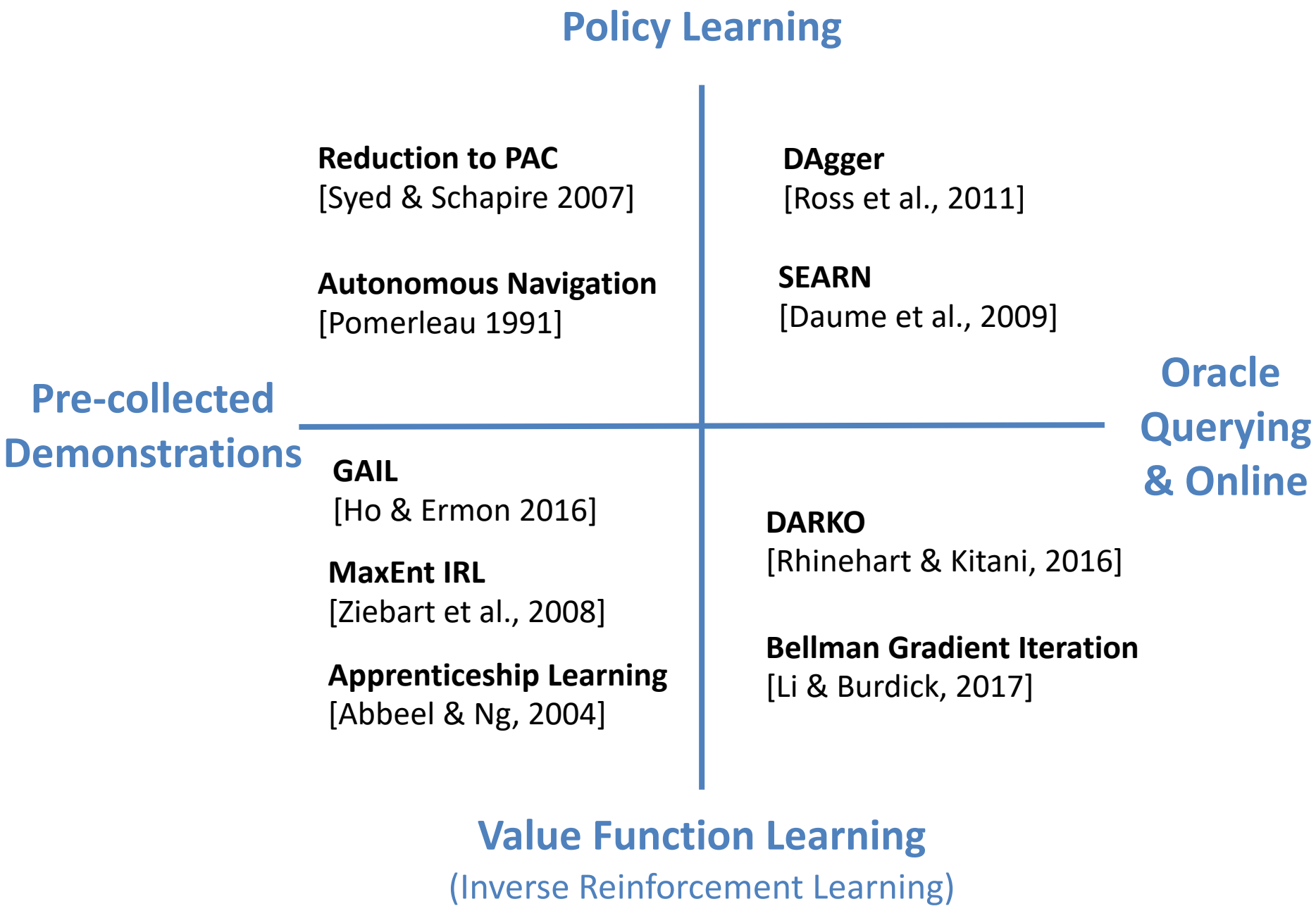
What to Imitate?

Human Demonstrations

Animal Demonstrations

Computational Oracle




Policy Learning
Reduction to PAC
[Syed & Schapire 2007]
Autonomous Navigation
[Pomerleau 1991]
DAgger
[Ross et al., 2011]
SEARN
[Daume et al., 2009]
Pre-collected Demonstrations
Oracle Querying & Online
GAIL
[Ho & Ermon 2016]
MaxEnt IRL
[Ziebart et al., 2008]
Apprenticeship Learning
[Abbeel & Ng, 2004]
DARKO
[Rhinehart & Kitani, 2016]
Bellman Gradient Iteration
[Li & Burdick, 2017]
Value Function Learning
(Inverse Reinforcement Learning)

Policy Learning

Reduction to PAC
[Syed & Schapire 2007]

Autonomous Navigation
[Pomerleau 1991]

DAgger
[Ross et al., 2011]

SEARN
[Daume et al., 2009]

Pre-colle… acle
Demonstra… erying
…nline

Previous (Deep Imitation) Work:

- Minimal assumptions
- Inefficient in complex & structured settings

[Rhinehart & Kitani, 2016]

MaxEnt IRL
[Ziebart et al., 2008]

Apprenticeship Learning
[Abbeel & Ng, 2004]

Bellman Gradient Iteration
[Li & Burdick, 2017]

Value Function Learning

(Inverse Reinforcement Learning)

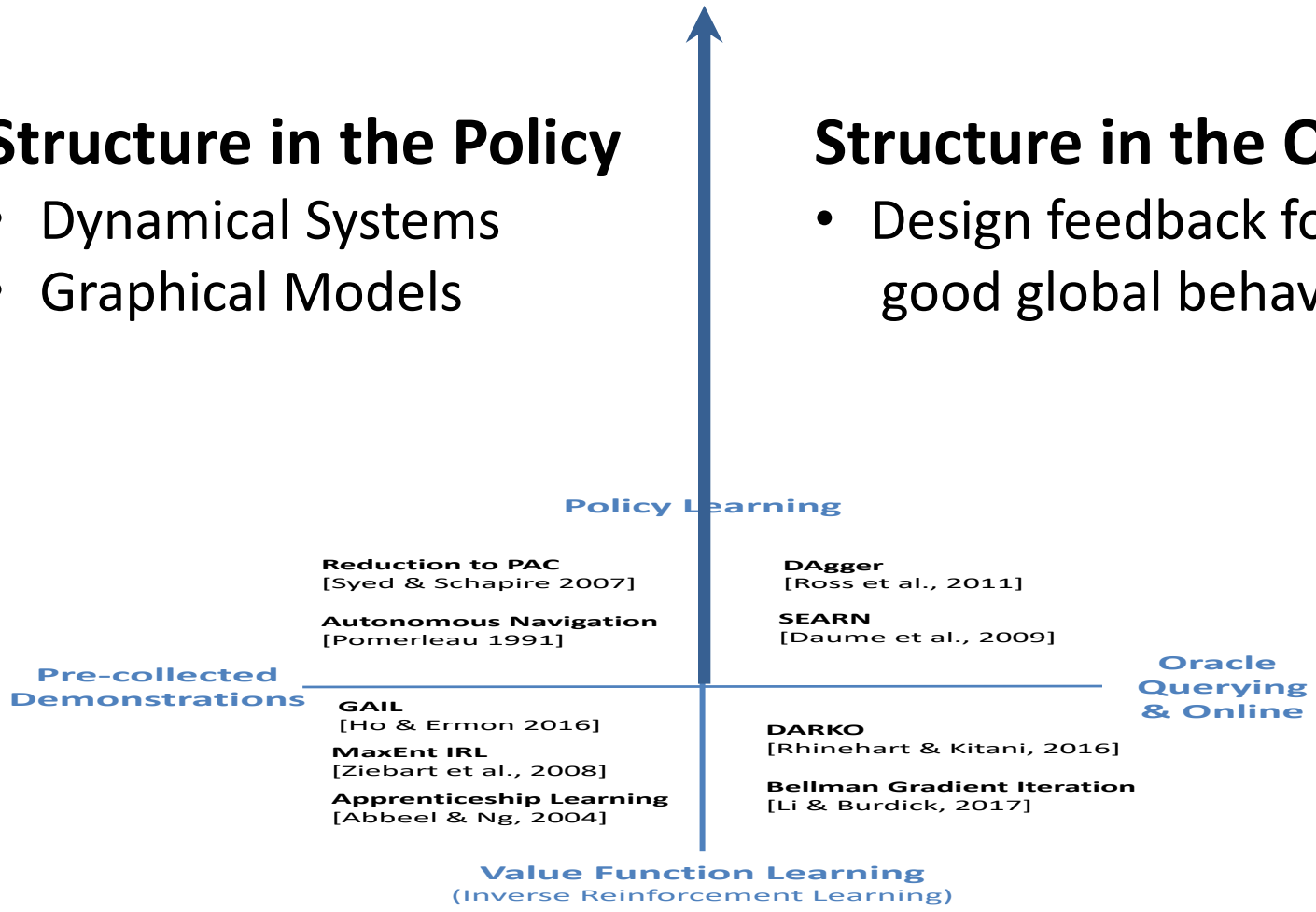
Structured Imitation Learning

Structure in the Policy

- Dynamical Systems
- Graphical Models

Structure in the Oracle

- Design feedback for good global behavior



Structured Imitation Learning

Structure in the Policy

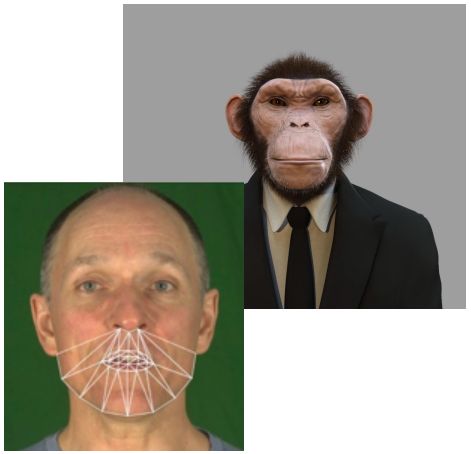
- Dynamical Systems
- Graphical Models

Structure in the Oracle

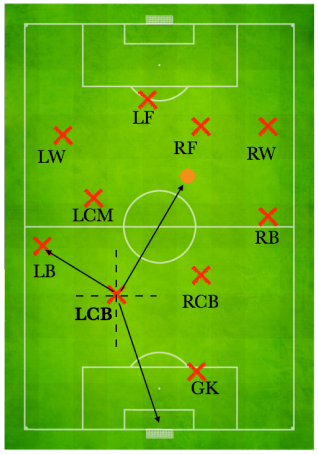
- Design feedback for good global behavior

Benefits:

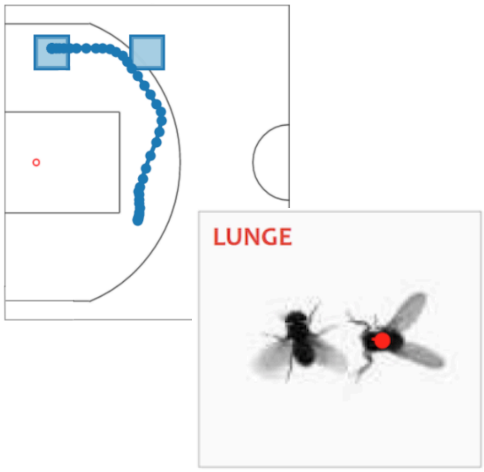
- Abstractions for domain experts
- Better inductive bias
- Reductions to conventional learning
- Composable theoretical guarantees



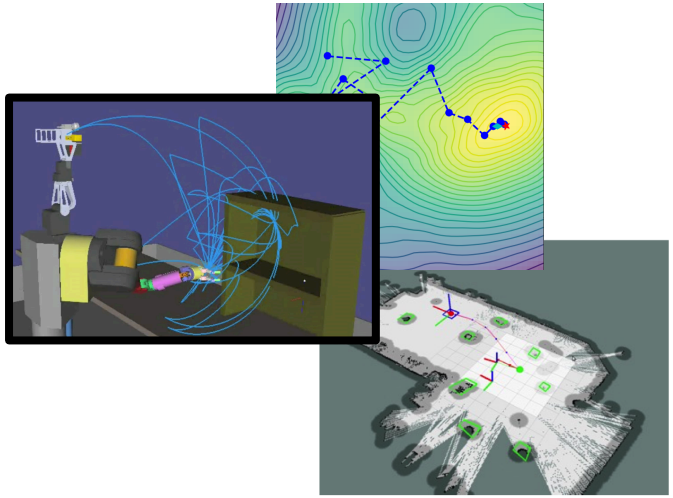
Speech Animation



Coordinated Learning

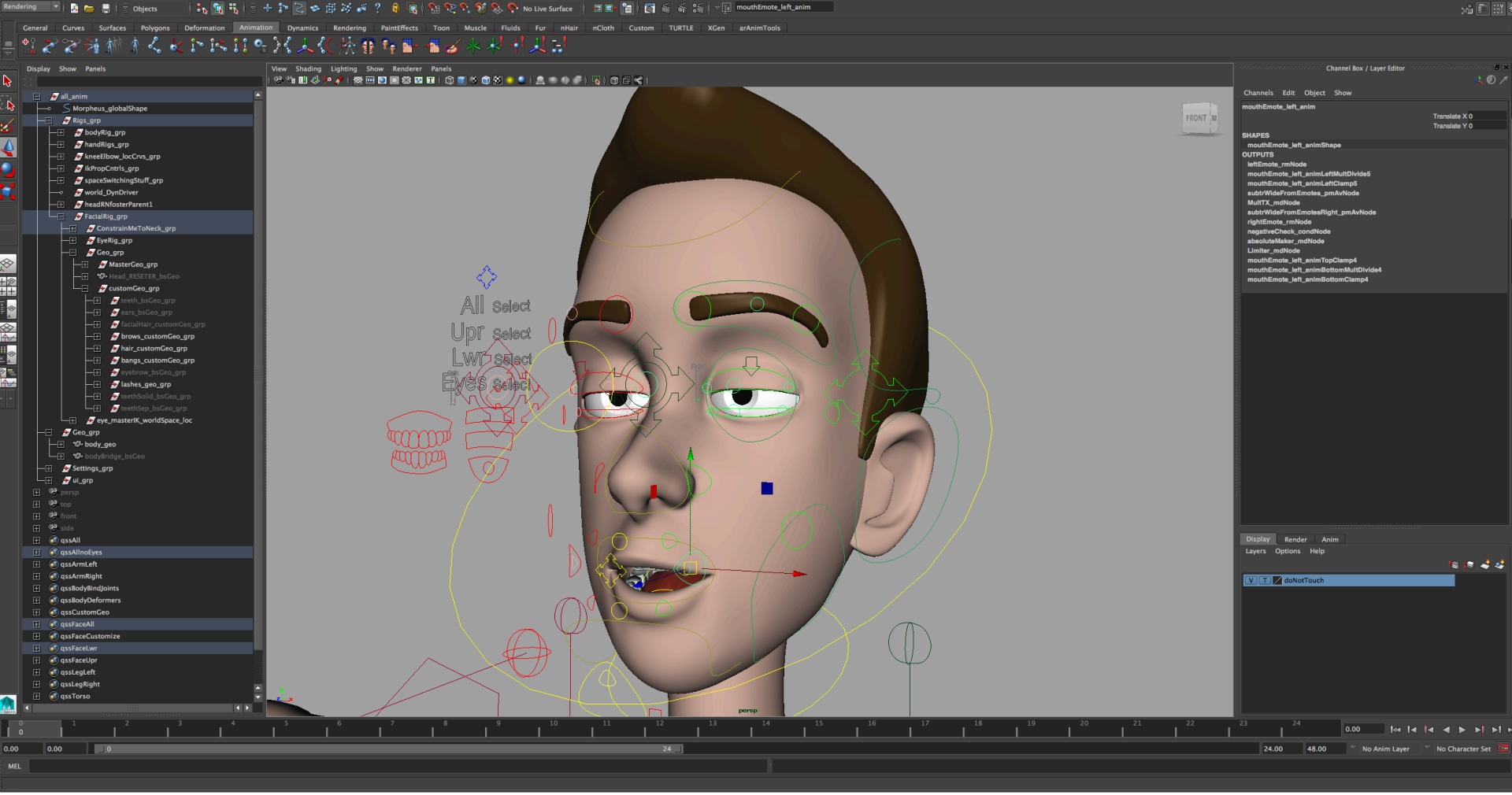


Hierarchical Behaviors (Generative)



Learning to Optimize

Side Constraints (Smoothness)

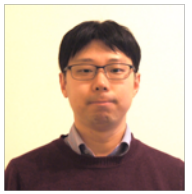


- **Animation artists spend ≥50% time on face**
  - Mostly eyes & mouth
  - Very tedious

We'll focus on mouth & speech.

Sarah Taylor

Prediction Task



Taehwan Kim

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 $h : X \rightarrow Y$

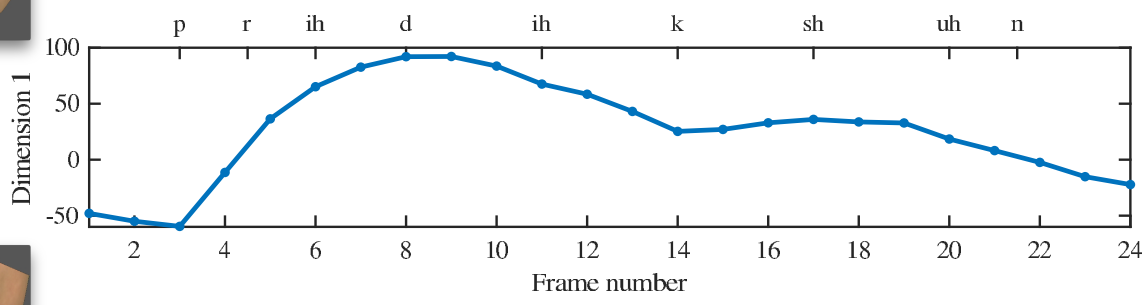
X

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	-

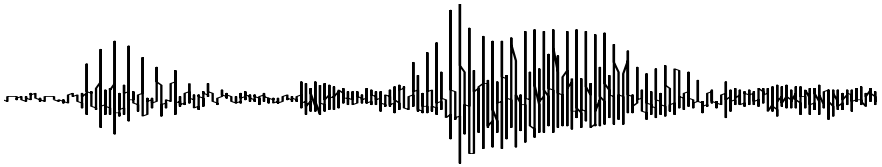
Phoneme sequence

h

Y



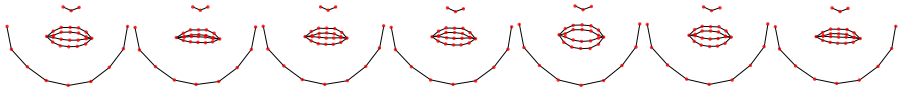
Sequence of face configurations



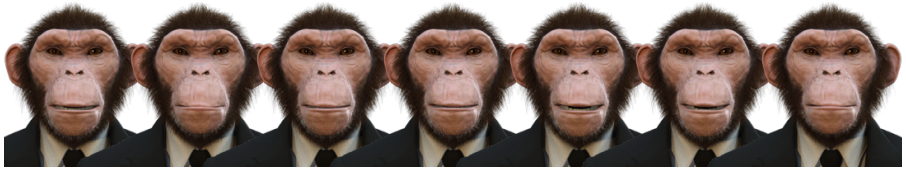
Input Audio

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

Speech Recognition



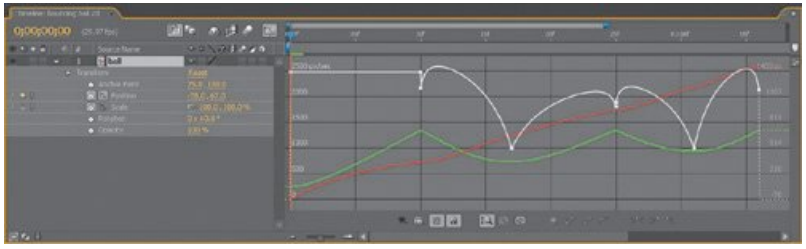
Speech Animation



(chimp rig courtesy of Hao Li)

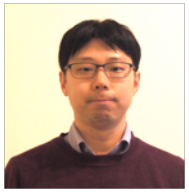
Retargeting

E.g., [Sumner & Popovic 2004]



Editing

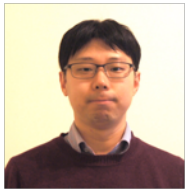

Sarah Taylor


Taehwan Kim

A Decision Tree Framework for Spatiotemporal Sequence Prediction
Taehwan Kim, Yisong Yue, Sarah Taylor, Iain Matthews. KDD 2015
A Deep Learning Approach for Generalized Speech Animation
Sarah Taylor, Taehwan Kim, Yisong Yue, et al. SIGGRAPH 2017

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

Sarah Taylor

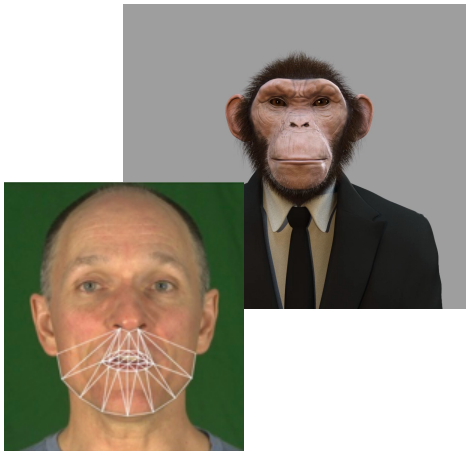


Taehwan Kim

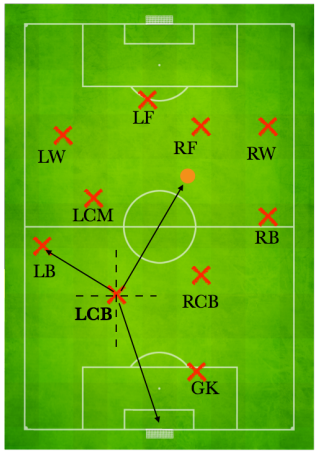
A Decision Tree Framework for Spatiotemporal Sequence Prediction
Taehwan Kim, Yisong Yue, Sarah Taylor, Iain Matthews. KDD 2015

A Deep Learning Approach for Generalized Speech Animation
Sarah Taylor, Taehwan Kim, Yisong Yue, et al. SIGGRAPH 2017

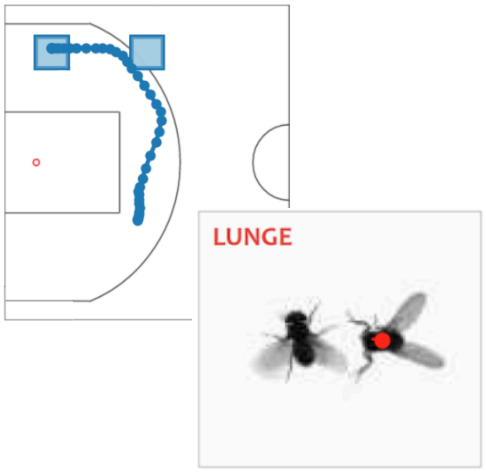
https://www.youtube.com/watch?v=9zL7qejW9fE



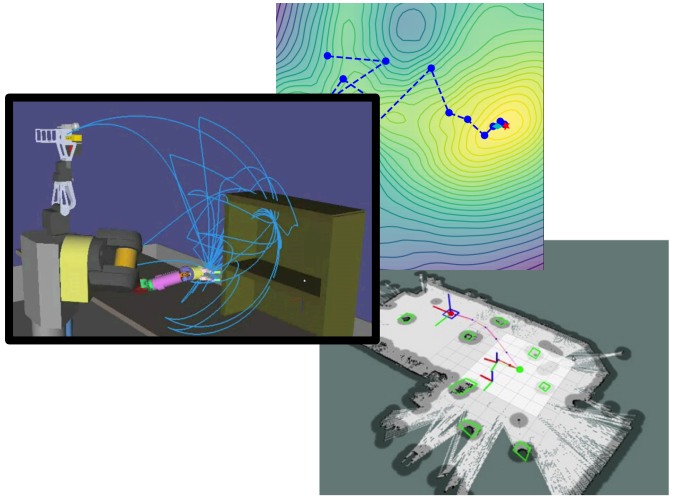
Speech Animation



Coordinated Learning



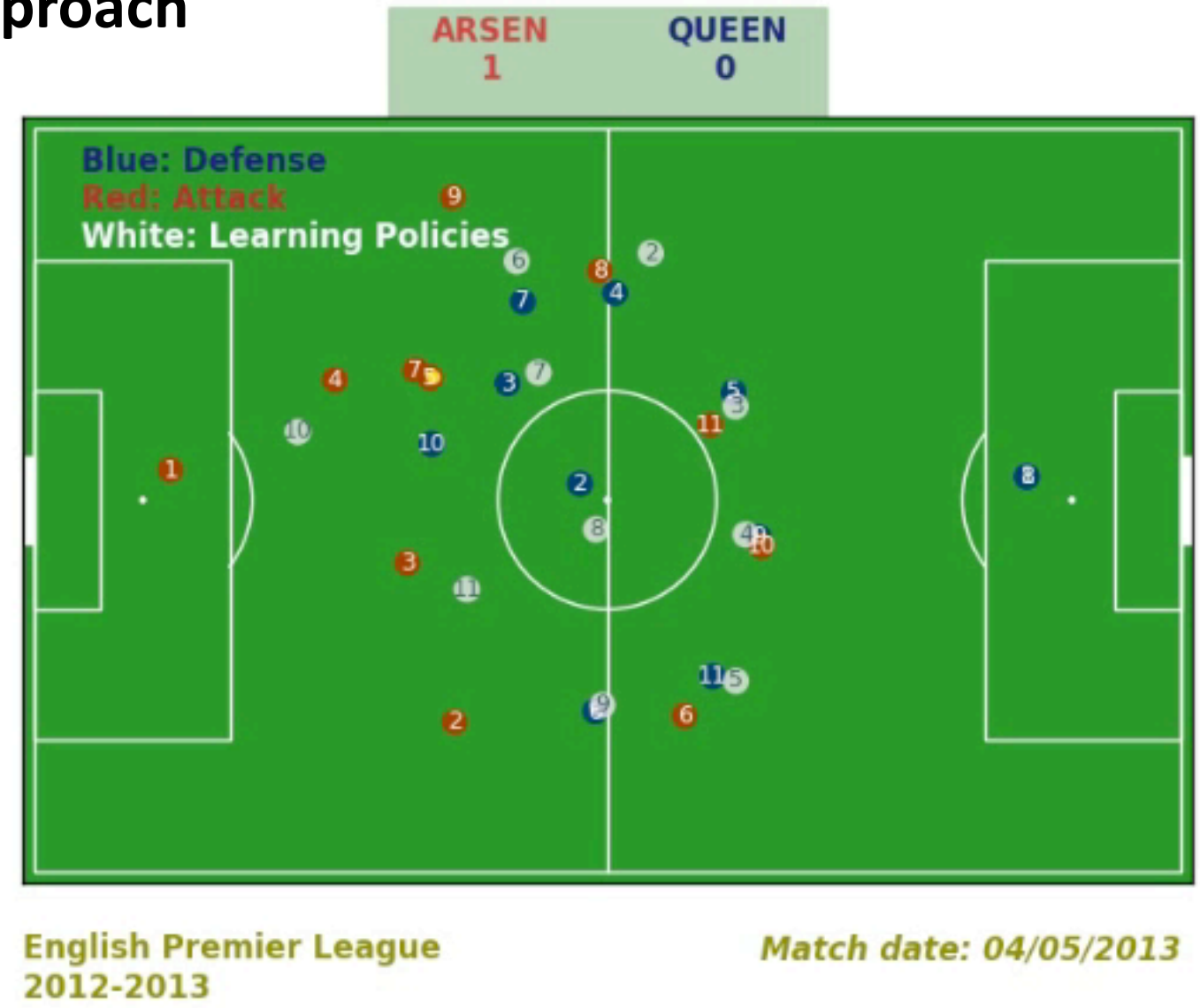
Hierarchical Behaviors (Generative)



Learning to Optimize

Smooth Imitation Learning

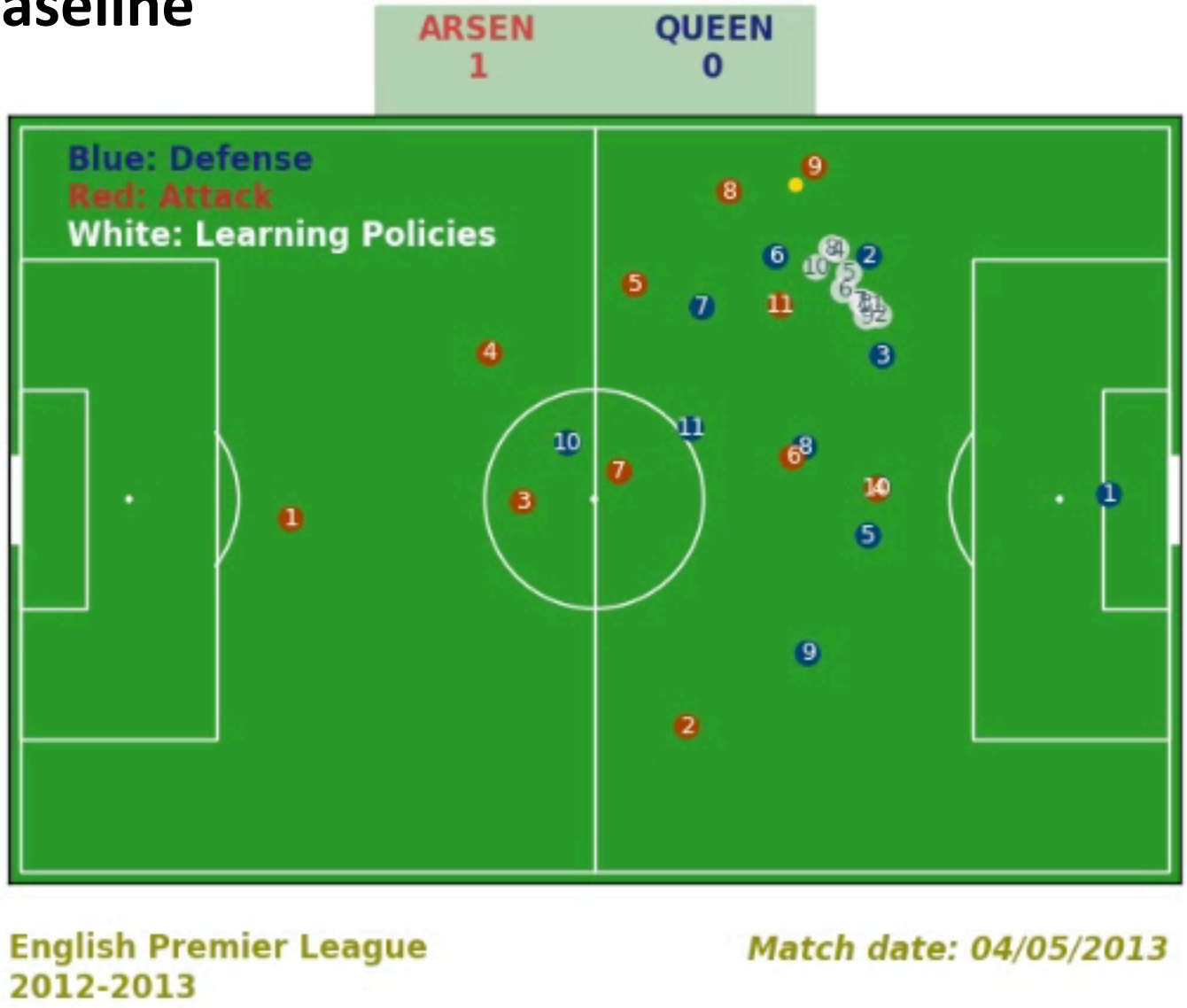
Our Approach



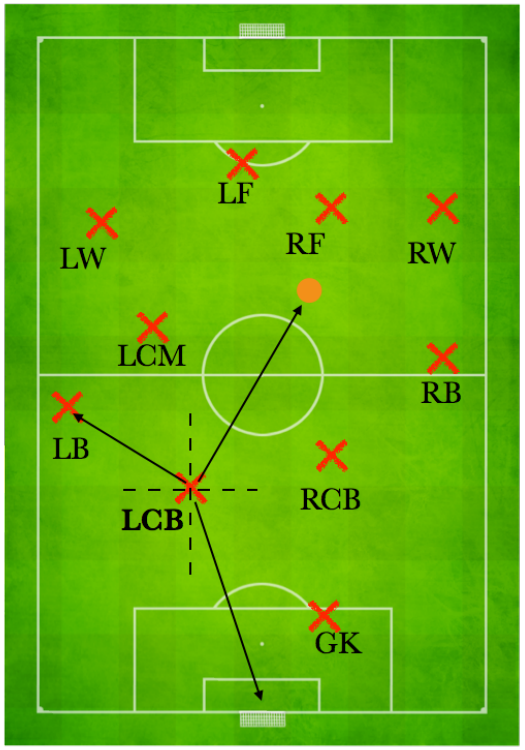
Data-Driven Ghosting using Deep Imitation Learning,
Hoang Le et al. SSAC 2017

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

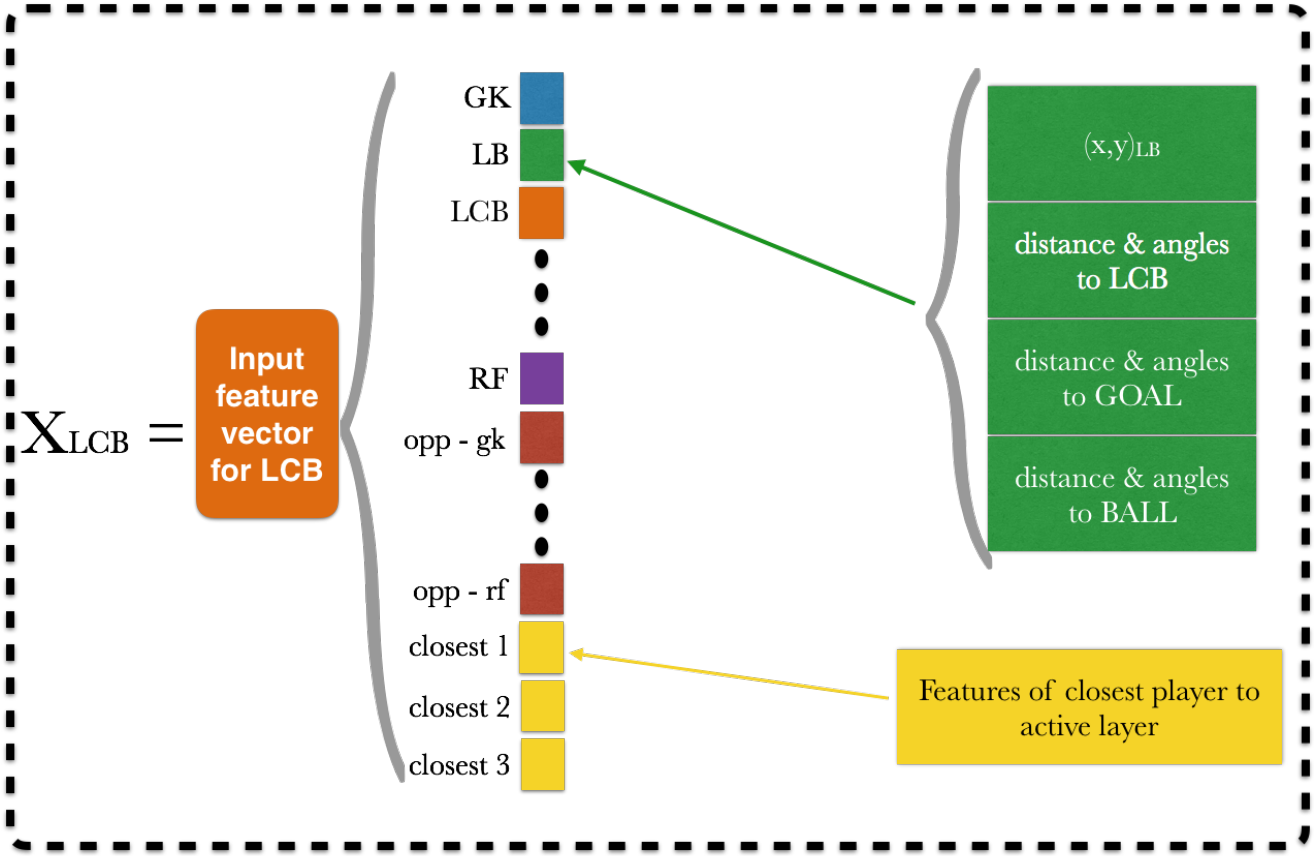
Naïve Baseline



State Representation



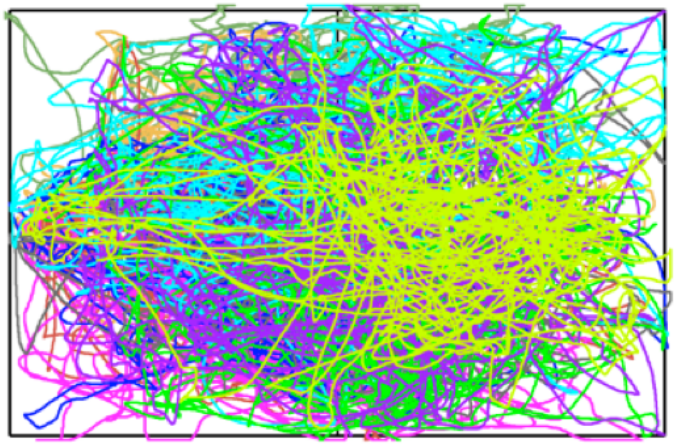
Geometric features computed



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

But Who Plays Which Role?

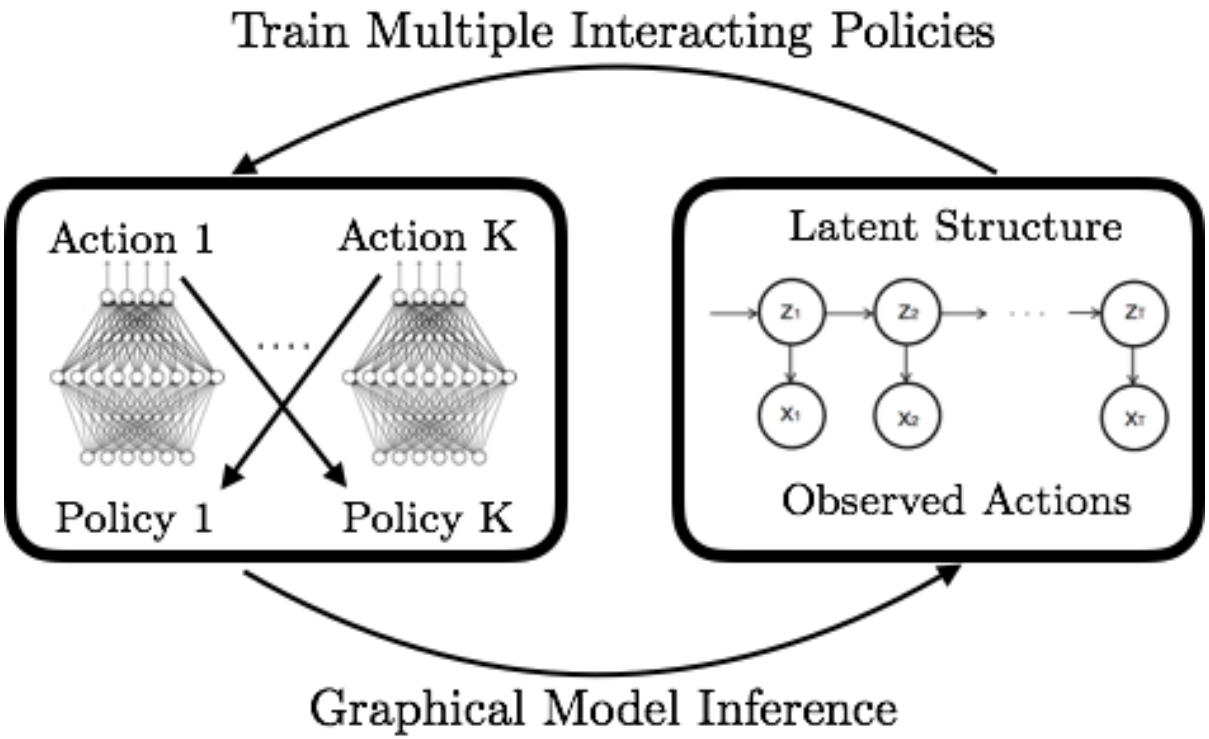
- All we get are trajectories!
  - Don’t know which belongs to which role.



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

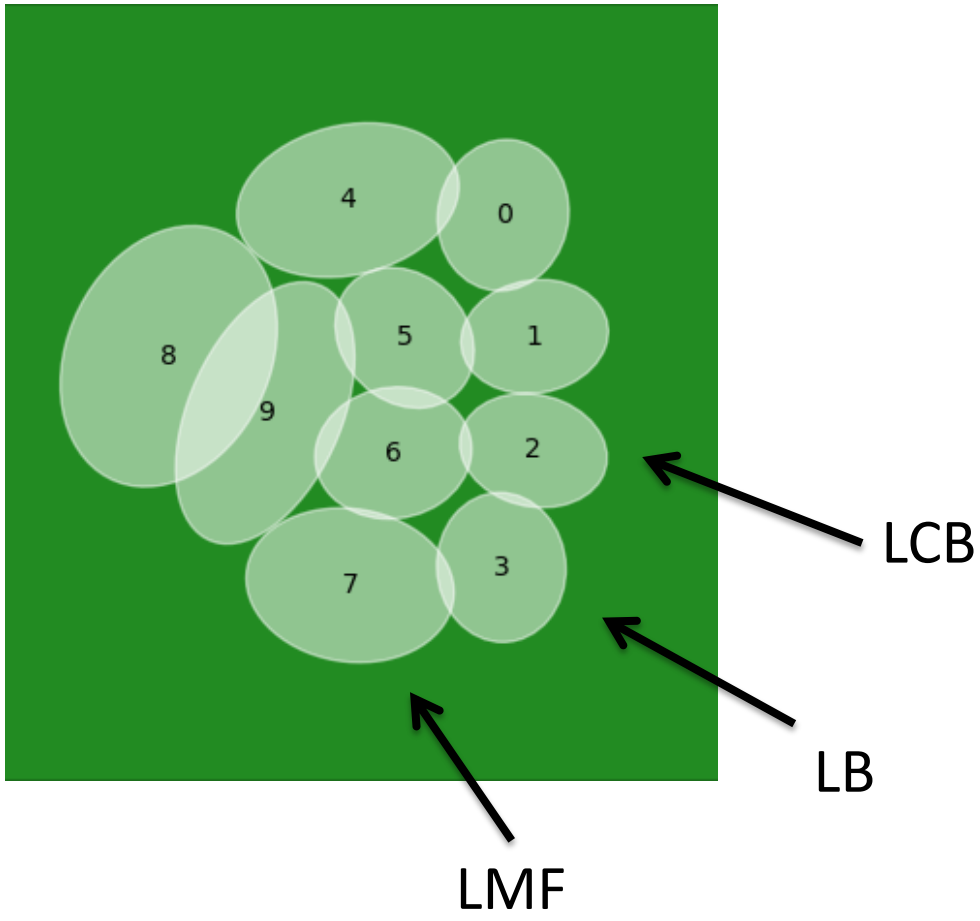
Hoang Le

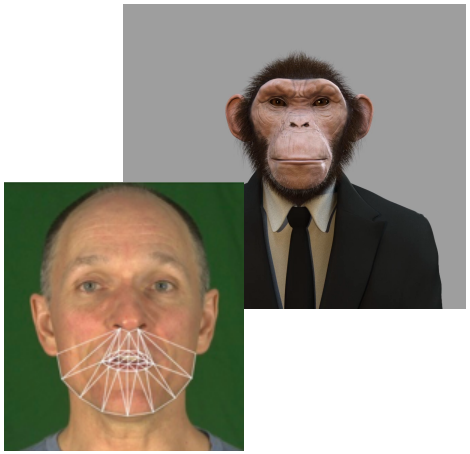
Coordination Model



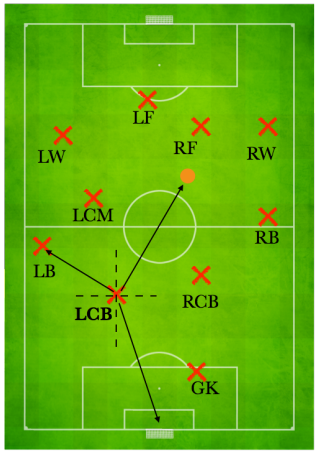
Coordinated Multi-Agent Imitation Learning
Hoang Le, Yisong Yue, Peter Carr, Patrick Lucey. ICML 2017

Learned Roles

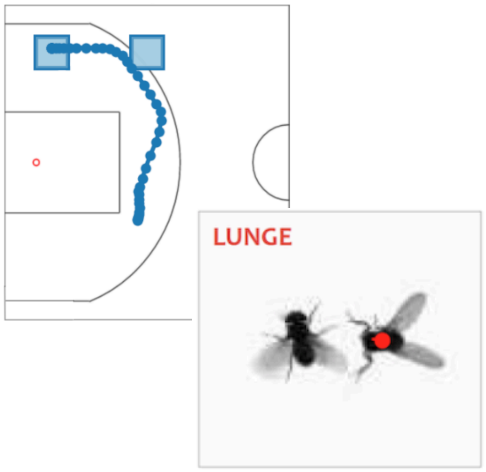




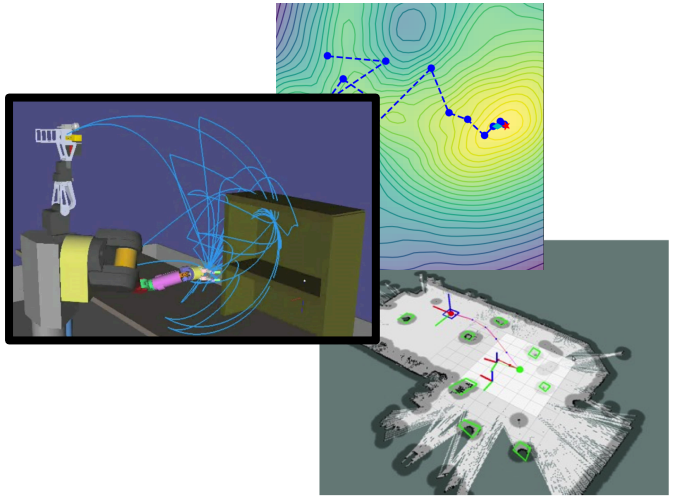
Speech Animation



Coordinated Learning



Hierarchical Behaviors (Generative)

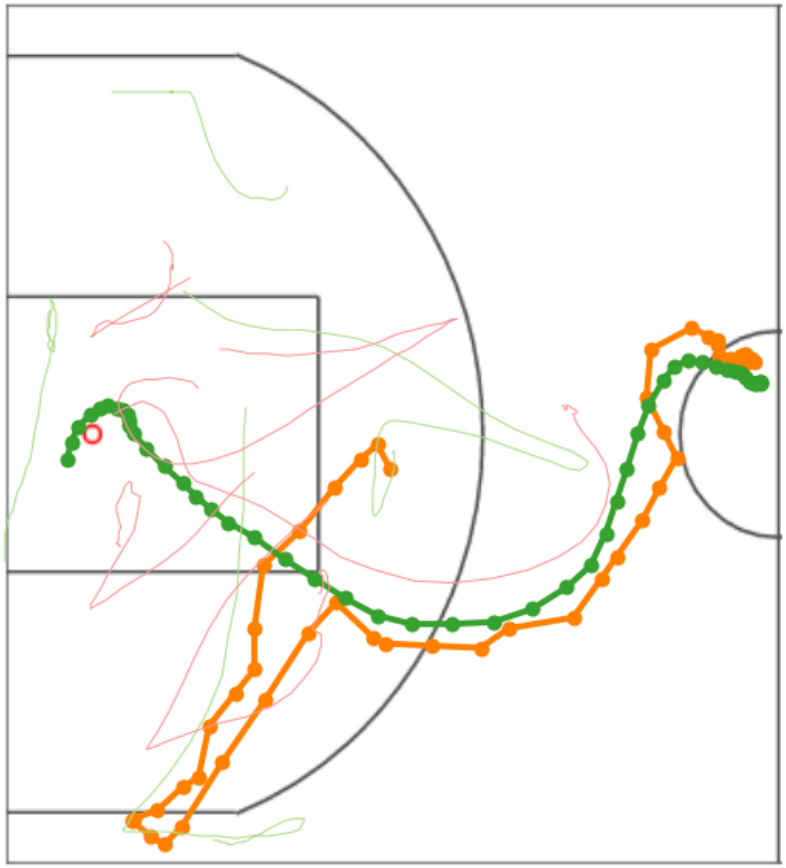


Learning to Optimize

Smooth Imitation Learning

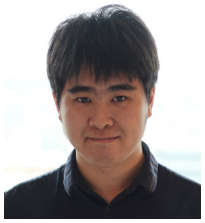
Strategy vs Tactics

- Long-term Goal:
  - Curl around basket
- Tactics
  - Drive left w/ ball
  - Pass ball
  - Cut towards basket



Stephan Zheng

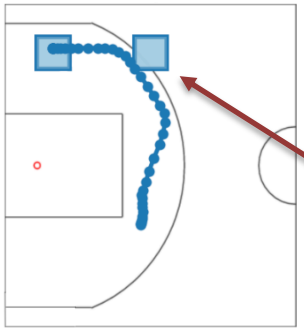
Eric Zhan

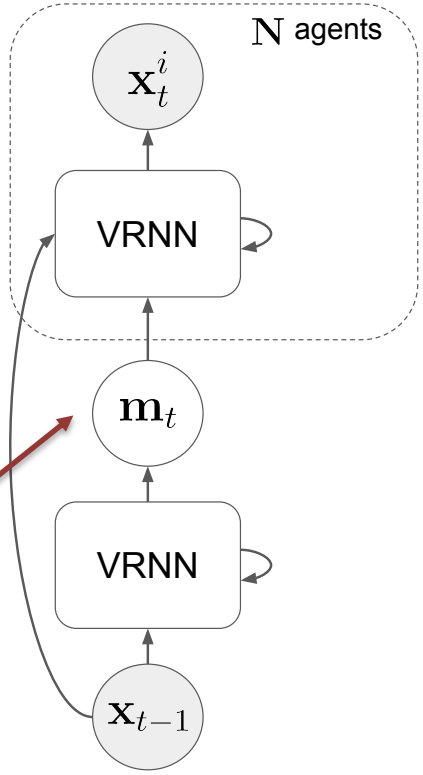


Yukai Liu

Generative + Hierarchical

- **Generative Imitation Learning**
  - No single “correct” action
- **Hierarchical**
  - Make predictions at multiple resolutions



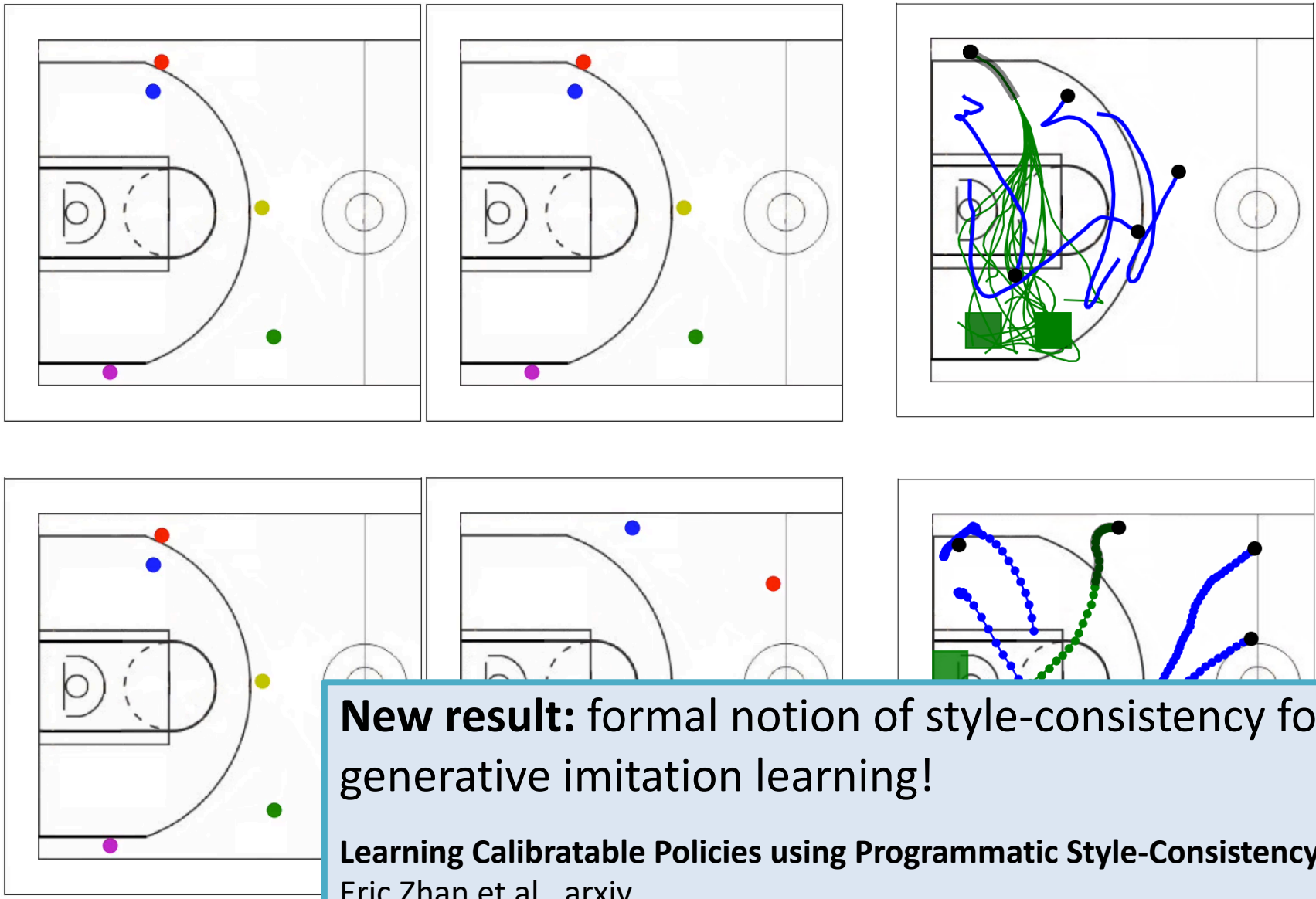


Macro-goals

Generating Long-term Trajectories using Deep Hierarchical Networks
Stephan Zheng, Yisong Yue, Patrick Lucey. NeurIPS 2016

Generating Multi-agent Trajectories using Programmatic Weak Supervision
Eric Zhan, Stephan Zheng, Yisong Yue, Long Sha, Patrick Lucey. ICLR 2019

NAOMI: Non-Autoregressive Multiresolution Sequence Imputation
Yukai Liu, Rose Yu, Stephan Zheng, Eric Zhan, Yisong Yue. NeurIPS 2019



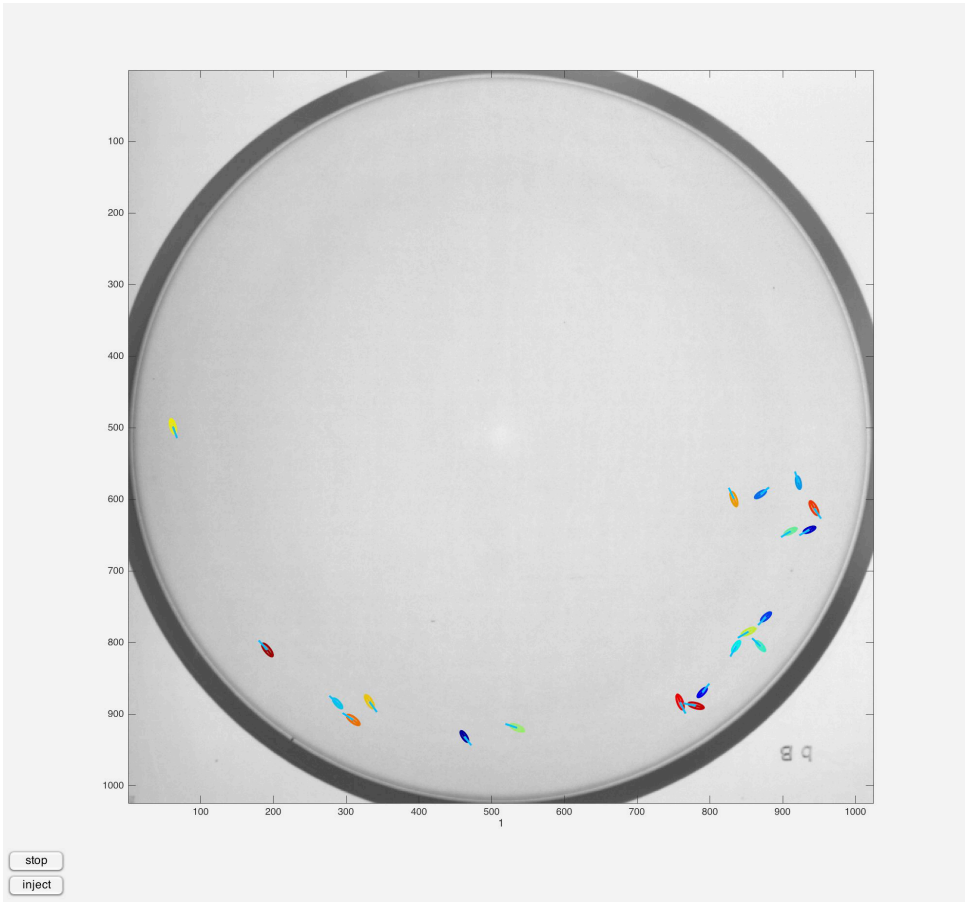
New result: formal notion of style-consistency for generative imitation learning!

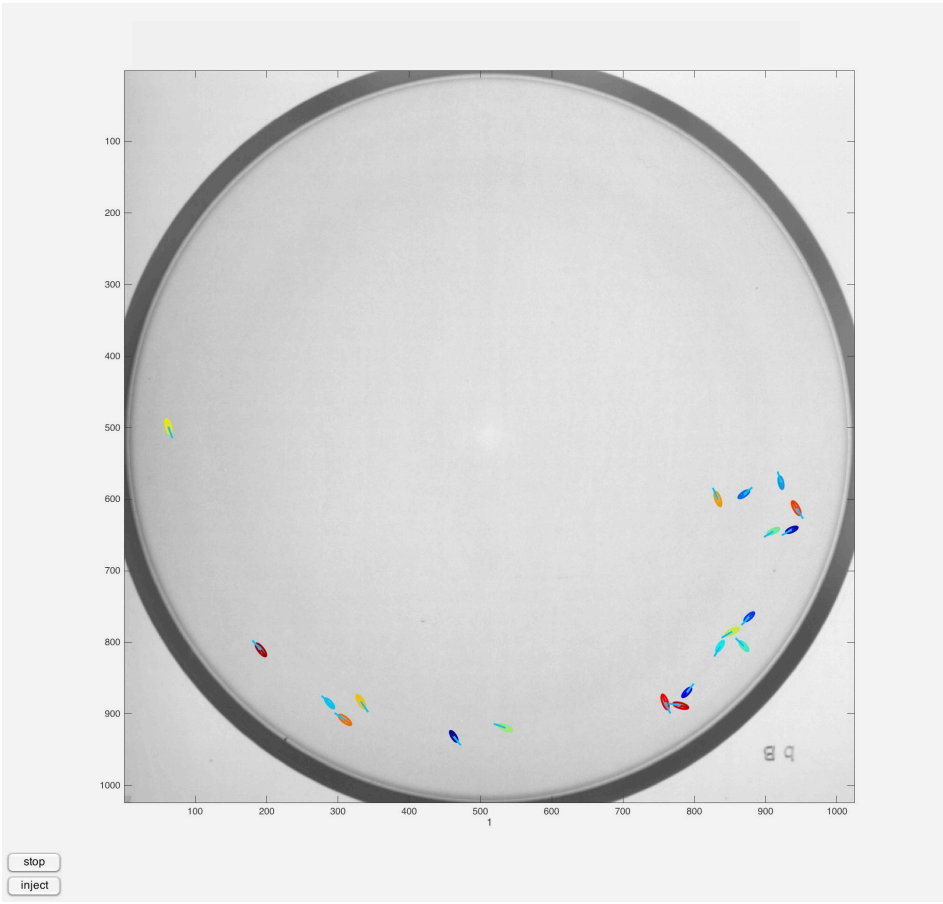
Learning Calibratable Policies using Programmatic Style-Consistency, Eric Zhan et al., arxiv

https://www.youtube.com/watch?v=0q1j22yMipY

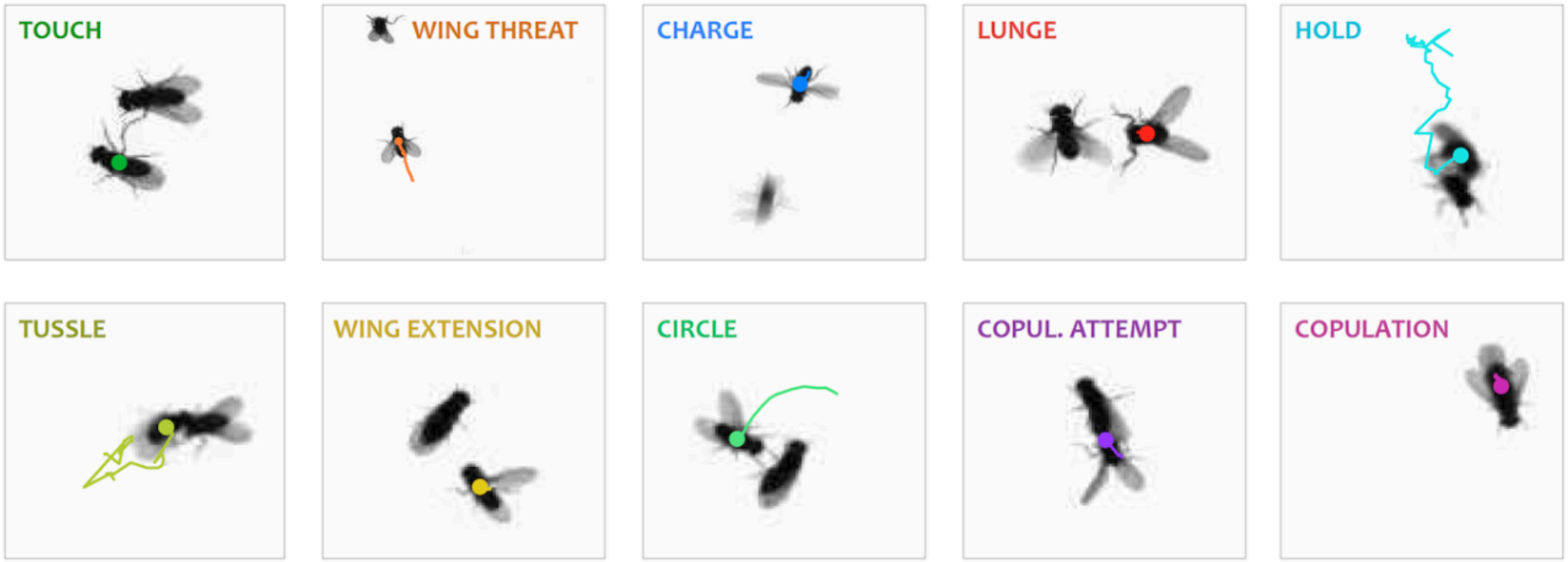
Eyrun
Eyolfsdottir

Drosophila Behavior





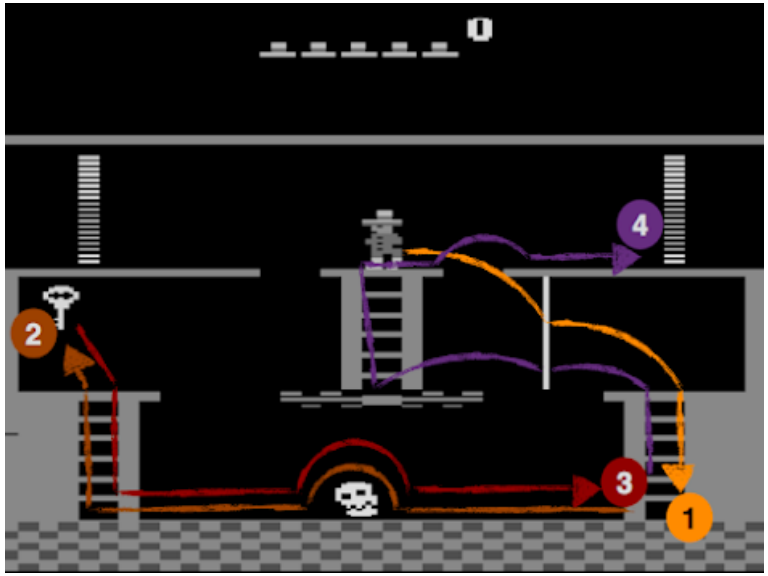
Activity Labels

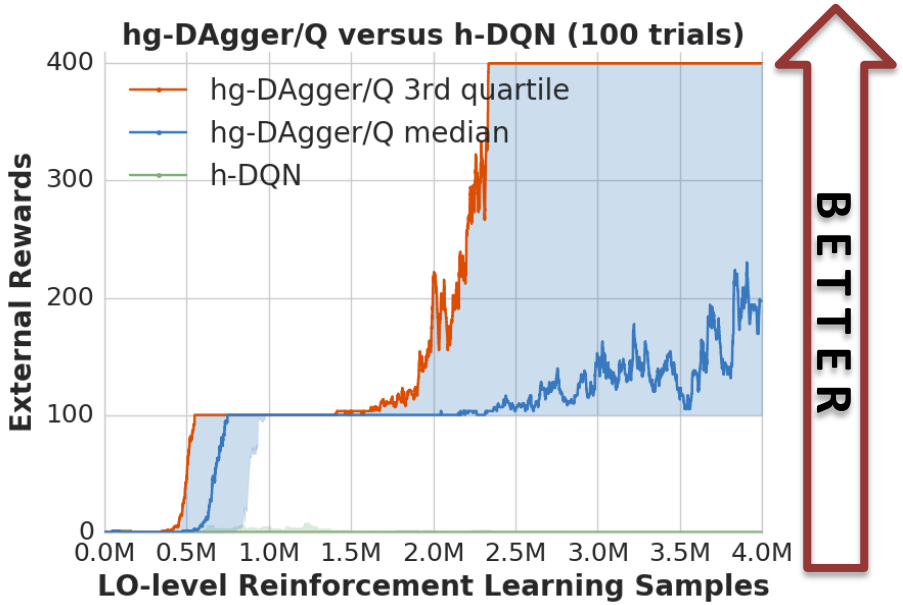


Learning recurrent representations for hierarchical behavior modeling
Eyrun Eyolfsdottir, Kristin Branson, Yisong Yue, Pietro Perona, ICLR 2017

Aside: Hierarchically Composing IL & RL

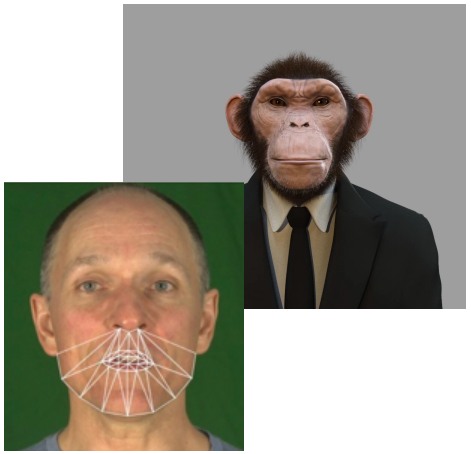
- IL for meta-controller (plan sub-goals)
- RL/IL for low-level controllers (individual sub-goals)



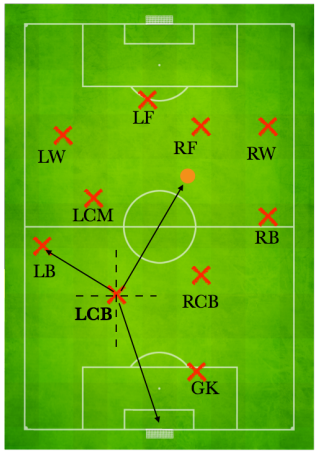


- More label efficient than flat IL
- Converge much faster than conventional hierarchical RL

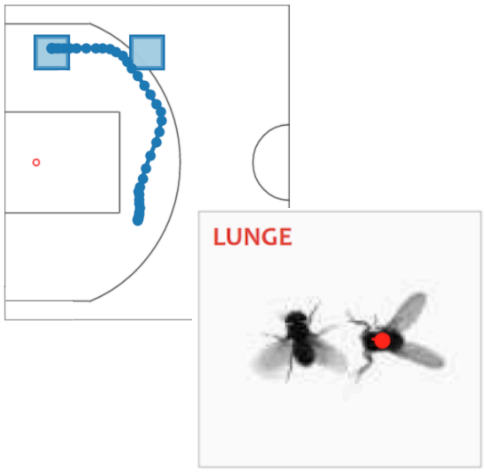
Hierarchical Imitation and Reinforcement Learning
Hoang Le, Nan Jiang, Alekh Agarwal, Miro Dudik, Yisong Yue, Hal Daume. ICML 2018



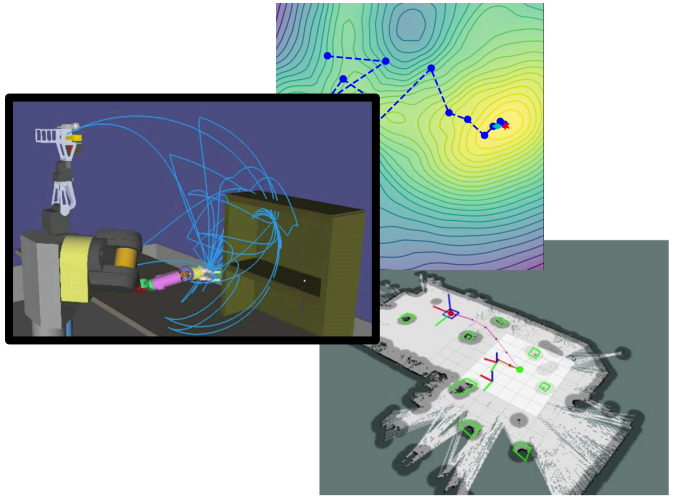
Speech Animation



Coordinated Learning



Hierarchical Behaviors (Generative)



Learning to Optimize

Side Constraints (Smoothness)

Optimization as Sequential Decision Making

- Many solvers are sequential:
  - Greedy
  - Search heuristics
  - Gradient descent

- Can view solver as “agent”
  - State = intermediate solution
  - Find a state with high reward (solution)

Optimization as Sequential Decision Making

Contextual Submodular Maximization

- Training set: (x, F_x)
- Greedily maximize F_x using only x
- **Learning Policies for Contextual Submodular Prediction** [ICML 2013]


Stephane Ross

Learning to Search

- Training set: $(x{=}\text{MILP}, y{=}\text{solution/search}{-}\text{trace})$
- Find y (or better solution)
- **Learning to Search via Retrospective Imitation** [arXiv]
- **Co-training for Policy Learning** [UAI 2019]


Jialin Song

Learning to Infer

- Training set: $(x{=}\text{data/model}, L{=}\text{likelihood})$
- Iteratively optimize L (generalizes VAEs)
- **Iterative Amortized Inference** [ICML 2018]
- **A General Method for Amortizing Variational Filtering** [NeurIPS 2018]

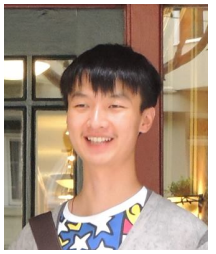

Joe Marino

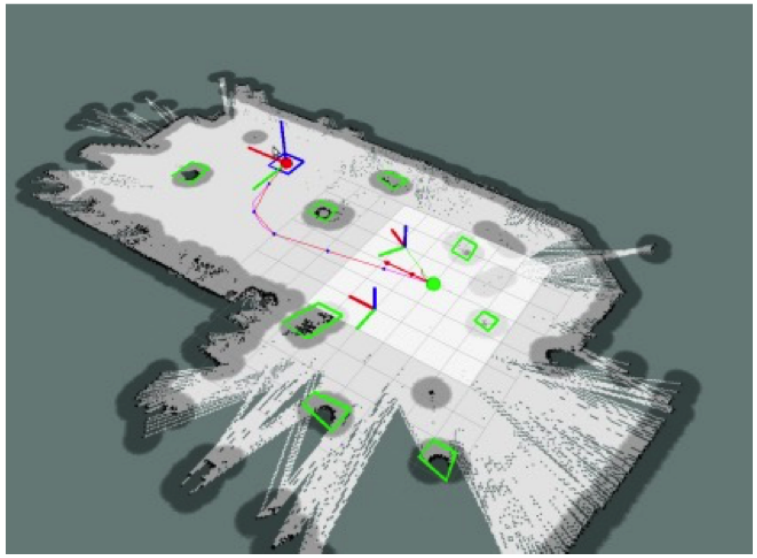
Ravi Lanka

Motivating Application

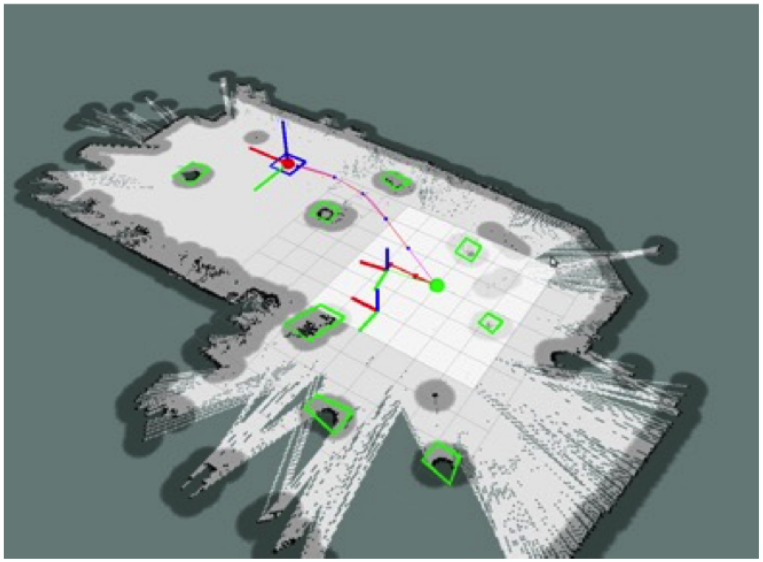
Risk-Aware Planning

Jialin Song





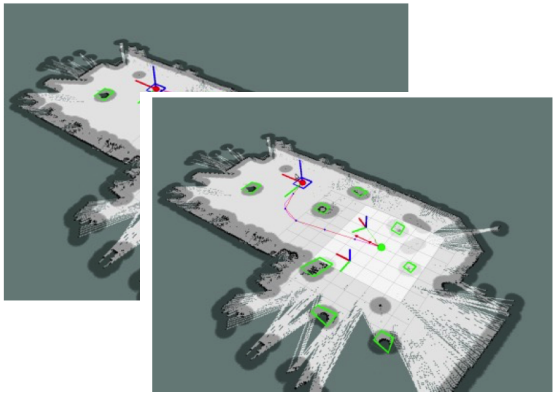
Low Risk



High Risk

- Compiled as mixed integer program
- Challenging optimization problem

NASA JPL



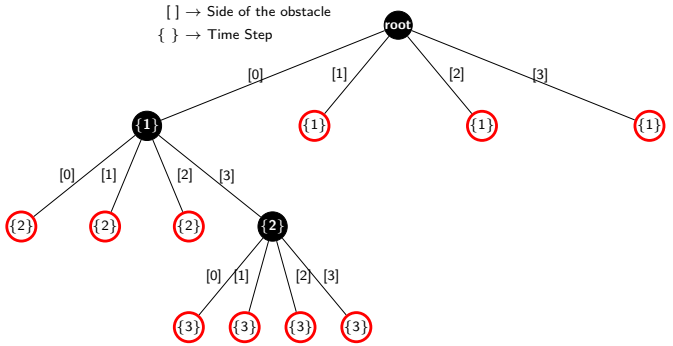
Distribution of Planning Problems

$$\min_{\mathbf{U}} \quad J(\mathbf{U}, \mathbf{X})$$

subject to,

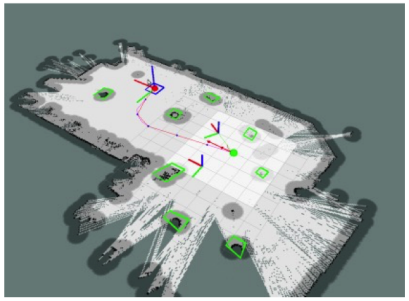
(Dynamic Constraint) $x_{t+1} = Ax_t + Bu_t,$

(Safety Constraints) $h_t^{iT} x_t \leq g_t^i$

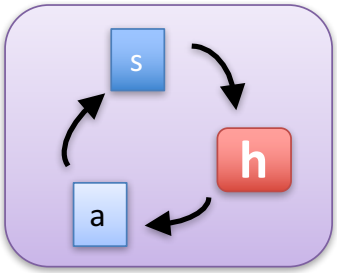


Compiled as Combinatorial Search Problems

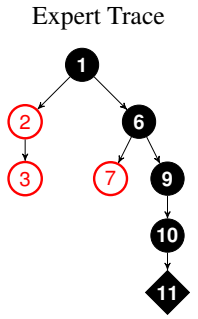
Test Instances



Imitation Learning

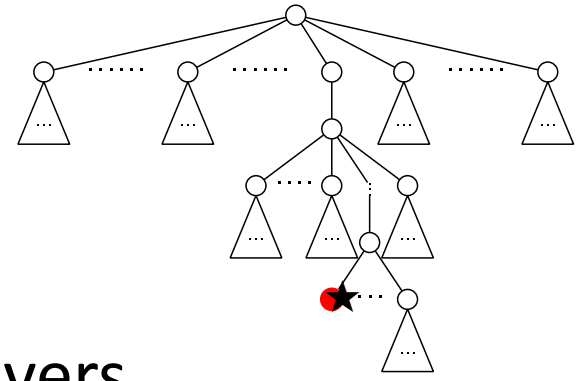


Collect Demonstrations



Our Approach Philosophy

- Leverage off-the-shelf solvers
  - Imitation learning
- Learn better solvers
  - Avoid mistakes of off-the-shelf solvers
- Scale up to larger problems
  - Not tractable with off-the-shelf solvers



Jialin Song

Retrospective Imitation

Ravi Lanka

- Given:
  - Family of Distributions of Search problems
    - Family is parameterized by size/difficulty

Difficulty levels: k=1,…,K

  - Solved Instances on the Smallest/Easiest Instances
    - “Demonstrations”
- Goal:
  - Learn to minimize mistakes
  - Scale up from Smallest/Easiest Instances
  - Formal Guarantees (see paper)

Connections to Curriculum Learning & Transfer Learning

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

Retrospective Imitation

- Two-Stage Algorithm

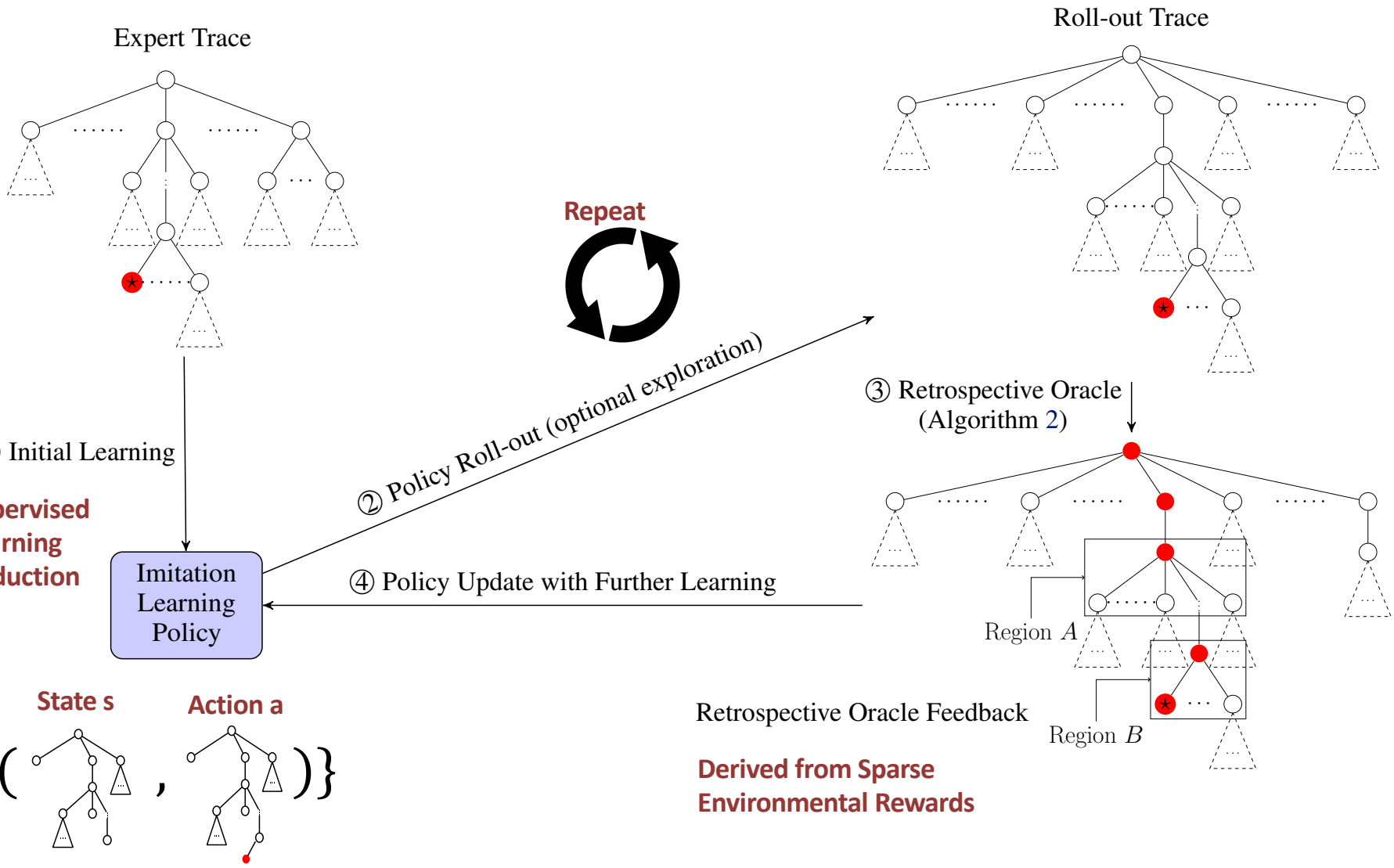
- Core Algorithm
  - Fixed problem difficulty
  - Reductions to Supervised Learning

Interactive IL w/ Sparse Environmental Rewards

- Full Algorithm w/ Scaling Up
  - Uses Core Algorithm as Subroutine

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

Retrospective Imitation (Core Algorithm)



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

Retrospective Imitation (Full Algorithm)

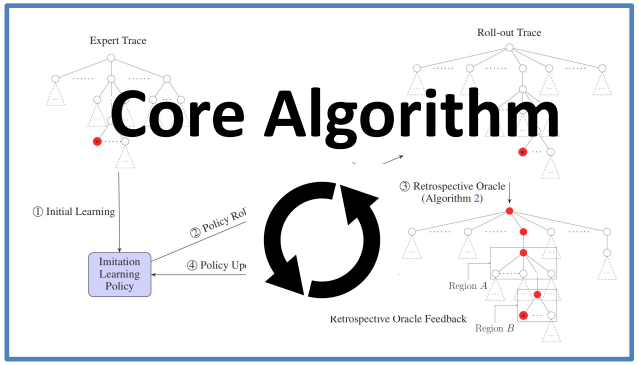
Initialize k=1

Initialize Gurobi/SCIP/CPlex

Problem Difficulty k

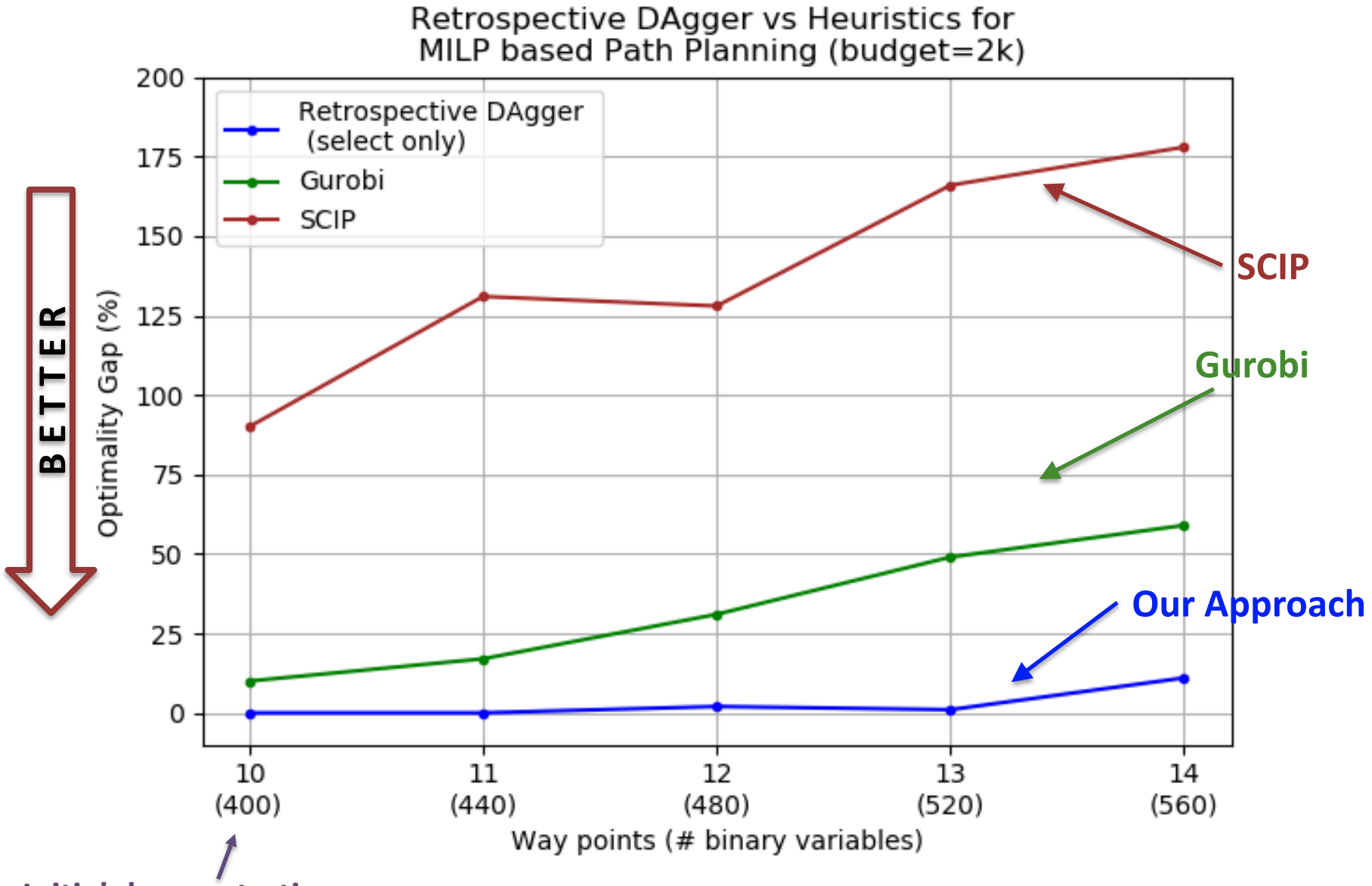
Base Solver

Instances & Demonstrations



k=k+1
Use trained h

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



Learning to Search via Retrospective Imitation

R. Lanka, J. Song, A. Zhao, A. Bhatnagar, Y. Yue, M. Ono. arXiv

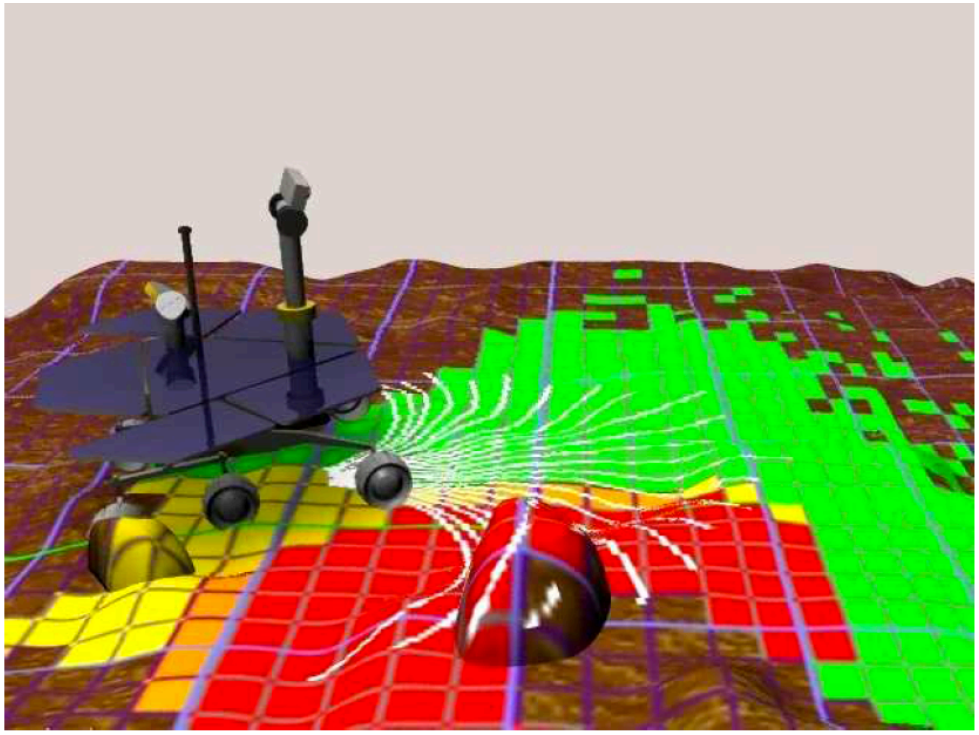
Ongoing: Integration with ENav

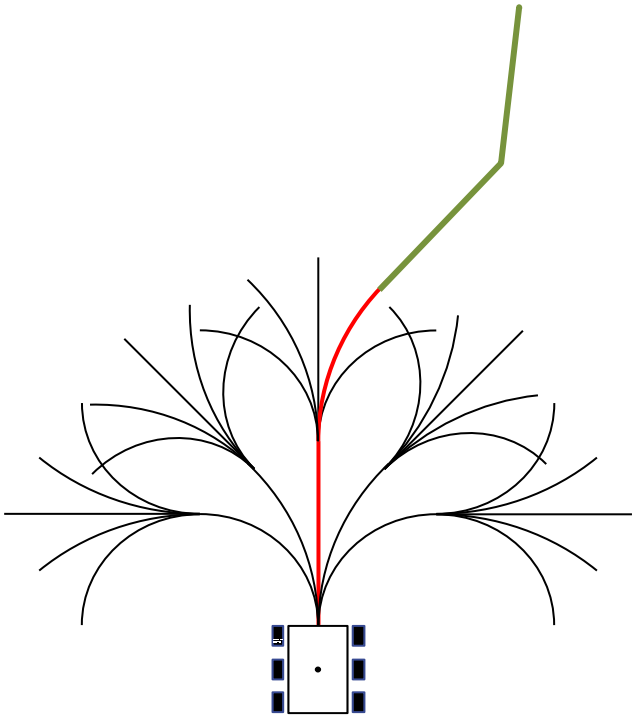

Ravi Lanka


Hiro Ono

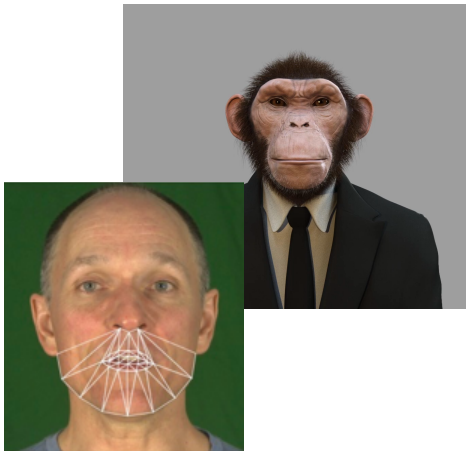

Olivier Toupet


Neil Abcouwer

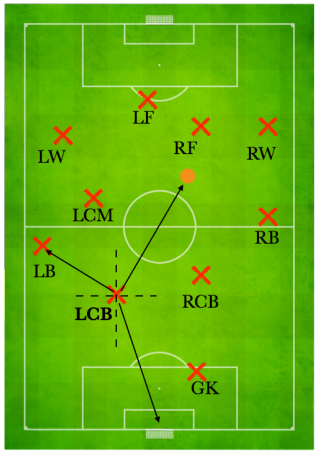




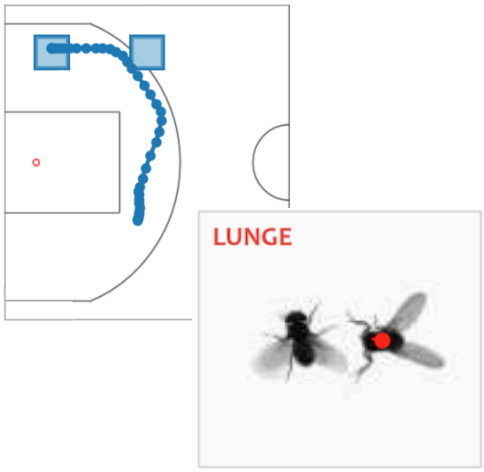
NASA JPL



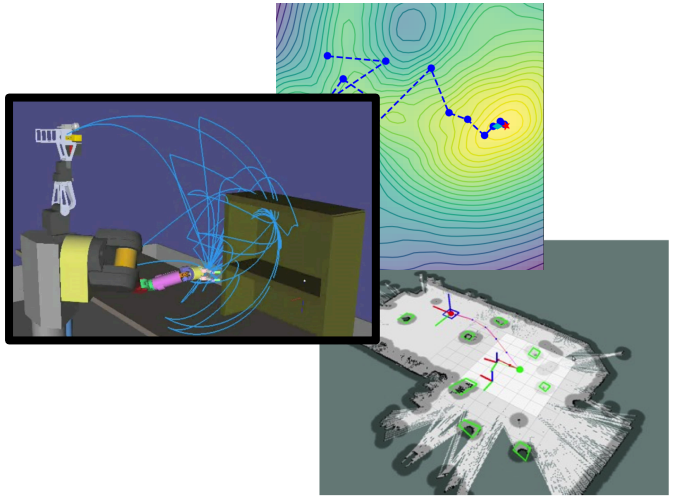
Speech Animation



Coordinated Learning



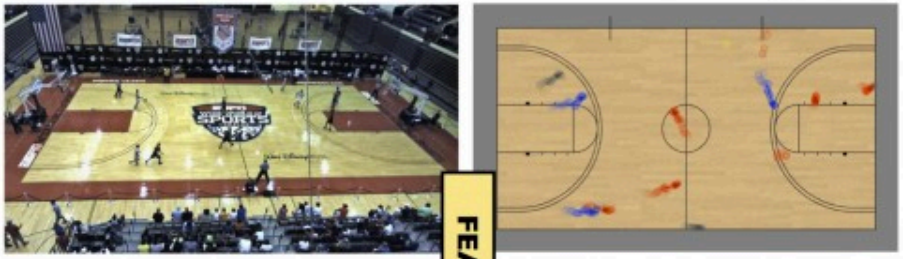
Hierarchical Behaviors (Generative)



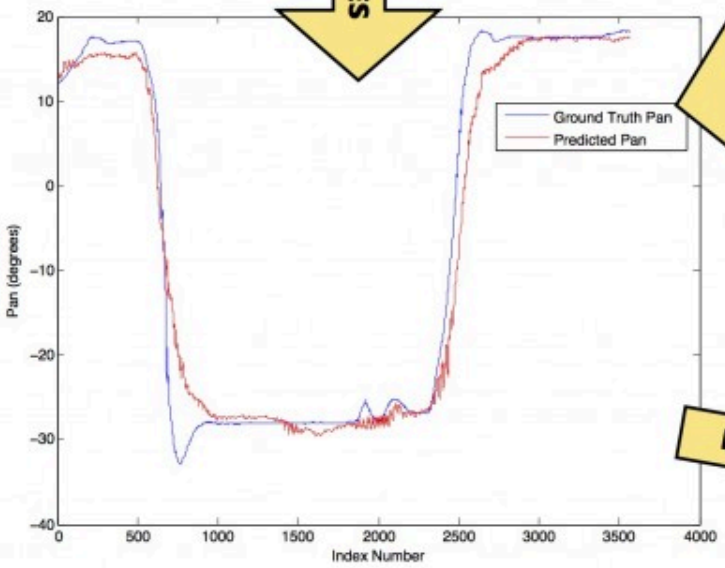
Learning to Optimize

Side Constraints (Smoothness)

Realtime Player Detection and Tracking



FEATURES



Learned Regressor

TRAIN

Human Operated Camera

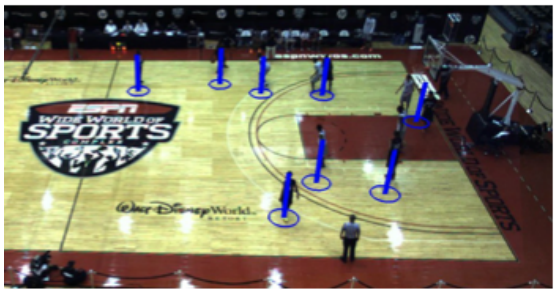
PREDICT

Autonomous Robotic Camera

Disney Research

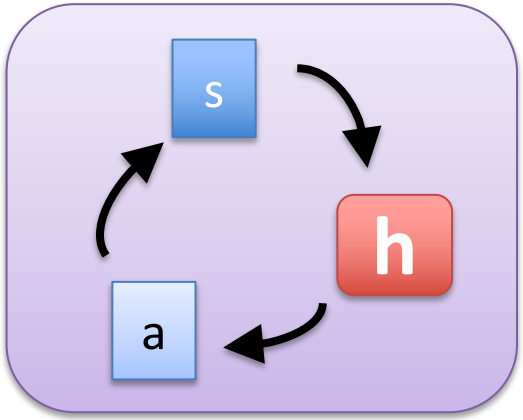
Problem Formulation

- Input: stream of x_t
  - E.g., noisy player detections



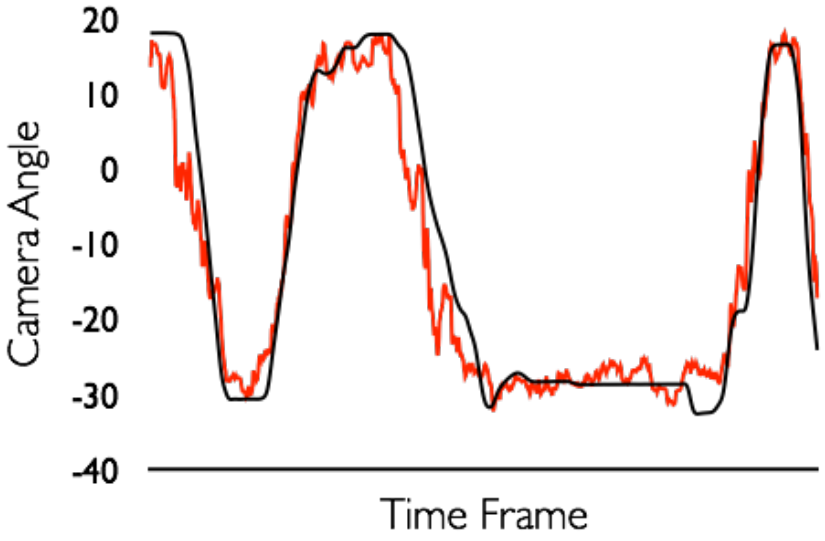
- State $s_t = (x_{t:t-K}, a_{t-1:t-K})$
  - Recent detections and actions

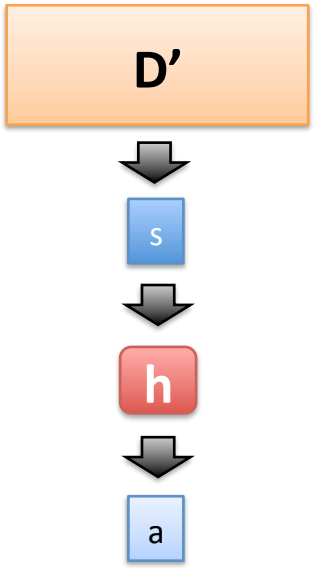
- Goal: learn $h(s_t) \rightarrow a_t$
  - Imitate expert



Naïve Approach

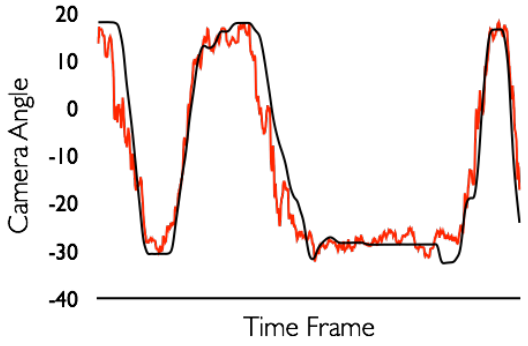
- Supervised learning of demonstration data
  - Train predictor per frame
  - Predict per frame



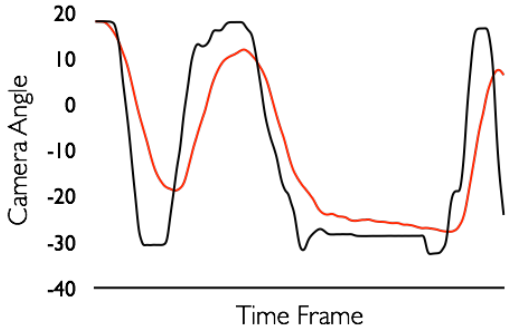


What is the Problem?

- Basically takes “infinite” training data to train smooth model.
  - Via input/output examples



- In practice, people do post-hoc smoothing



Cannot Rely 100% on Learning!

- People have models of smoothness!
  - Kalman Filters
  - Linear Autoregressors
  - Etc...
- Pure ML approach throws them away!
  - ”black box”

Hybrid Model-Based + Black-Box

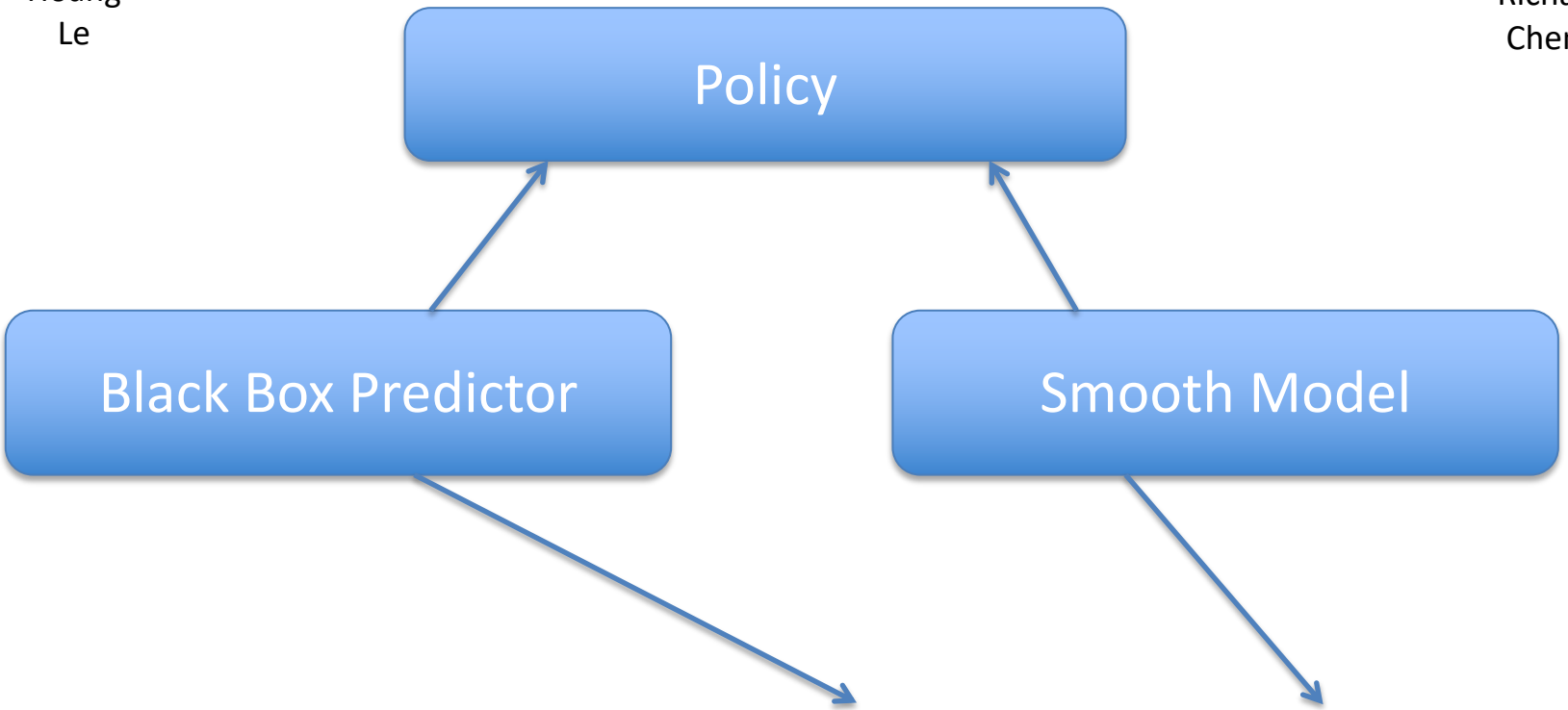
- Model-based approaches
  - Strong assumptions, well specified
  - Lacks flexibility
  - E.g., Kalman Filter, Linear Autoregressor
- Black-box approaches
  - Assumption free, underspecified
  - Requires a lot of training data
  - E.g., random forest, deep neural network
- **Best of both worlds?**

Conventional Models

Hoang Le

New Policy Class

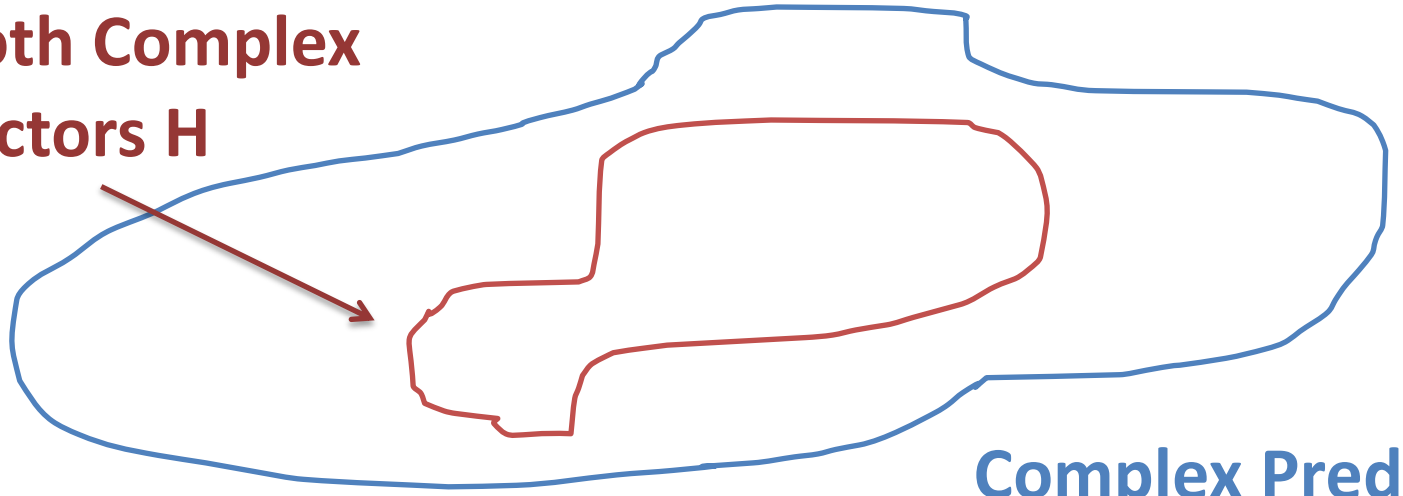
Richard Cheng



$$h\left(s_t \equiv (x_{t:t-K}, a_{t-1:t-K})\right) = argmin_{a'}(f(s_t) - a')^2 + \lambda(g(a_{t-1:t-K}) - a')^2$$

$$= \frac{f(s_t) + \lambda g(a_{t-1:t-K})}{1+\lambda}$$

Functional Regularization



$$h\left(s_t \equiv (x_{t:t-K}, a_{t-1:t-K})\right) = argmin_{a'}(f(s_t) - a')^2 + \lambda(g(a_{t-1:t-K}) - a')^2$$

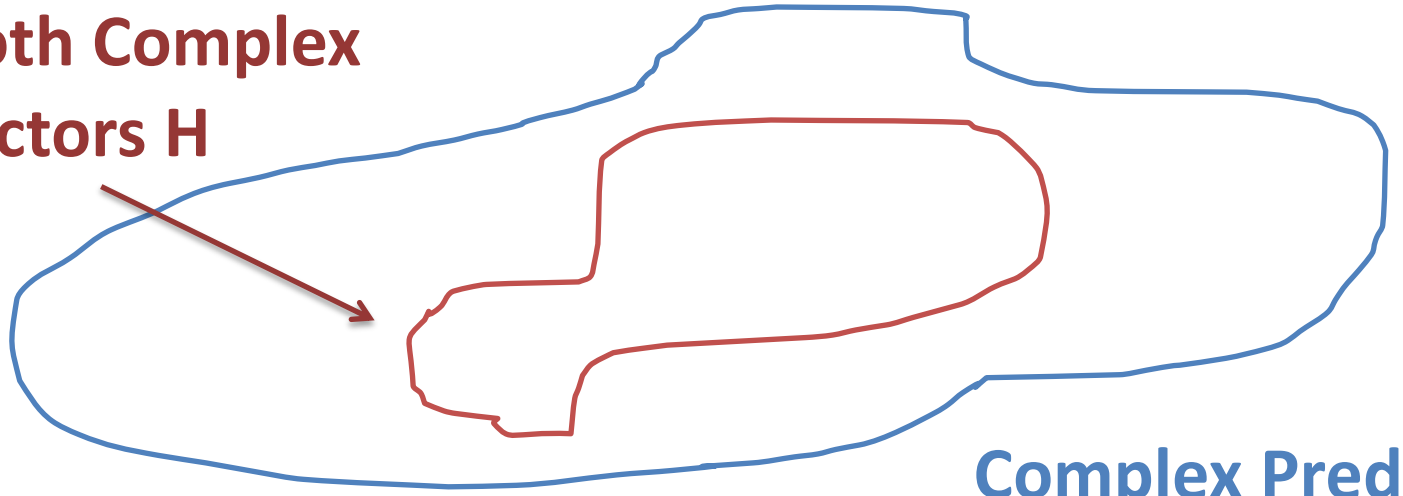
$$= \frac{f(s_t) + \lambda g(a_{t-1:t-K})}{1+\lambda}$$

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, Gabor Orosz, Swarat Chaudhuri, Yisong Yue, Joel Burdick. ICML 2019

Batch Policy Learning under Constraints
Hoang Le, Cameron Voloshin, Yisong Yue. ICML 2019

Stability of g → (relaxed) stability of h!



$$h\left(s_t \equiv (x_{t:t-K}, a_{t-1:t-K})\right) = argmin_{a'}(f(s_t) - a')^2 + \lambda(g(a_{t-1:t-K}) - a')^2$$

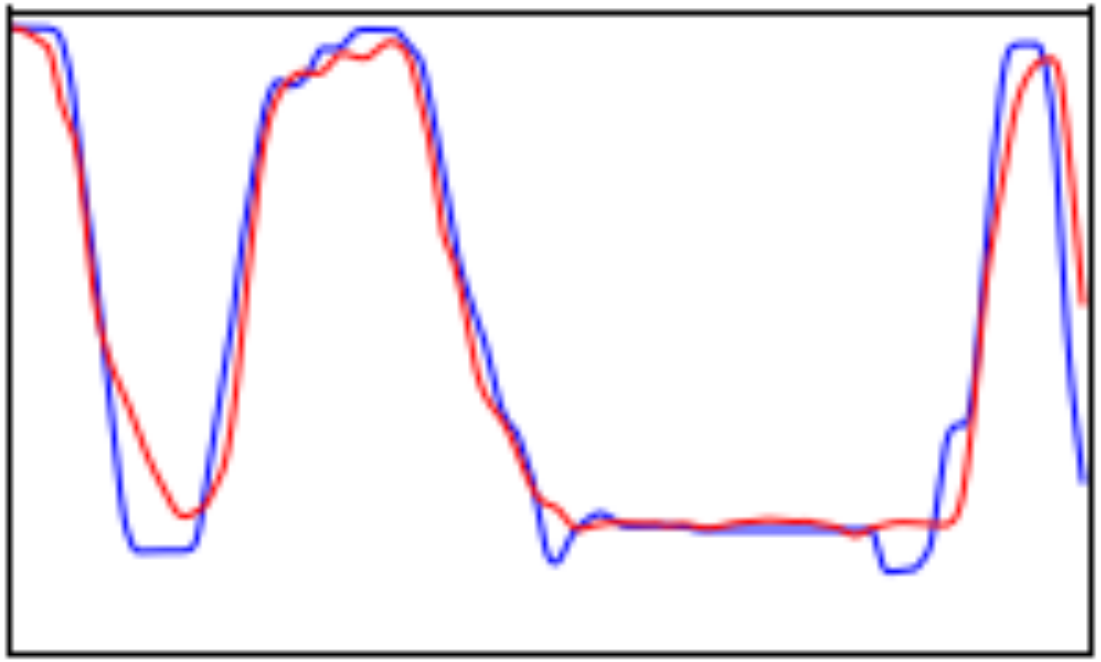
$$= \frac{f(s_t) + \lambda g(a_{t-1:t-K})}{1+\lambda}$$

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, Gabor Orosz, Swarat Chaudhuri, Yisong Yue, Joel Burdick. ICML 2019

Batch Policy Learning under Constraints
Hoang Le, Cameron Voloshin, Yisong Yue. ICML 2019

Our Result



$$h\left(s_t \equiv (x_{t:t-K}, a_{t-1:t-K})\right) = \frac{f(s_t) + \lambda g(a_{t-1:t-K})}{1+\lambda}$$

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

Qualitative Comparison

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 operators to help create a smoother and cleaner broadcast.

B pproach

Learning Online **Recurrent Decision Trees**
Jianhui Chen, Hoang

Lessons Learned

- **Intuition:** Let model do most of work
  - Certifiable guarantees (e.g., smoothness, stability)
  - Black box (deep neural net) adds flexibility
  - "Regularization" improves learning
    - Exponentially faster convergence w.r.t. SEARN

Lipschitz from smooth temporal dynamics

- Other settings?
  - Deep learning + robust control / formal methods?

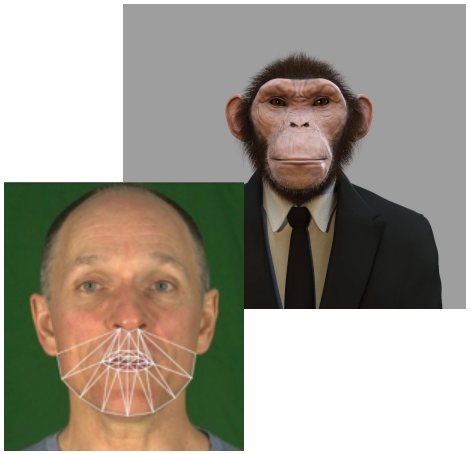

Aaron Ames


Soon-Jo Chung

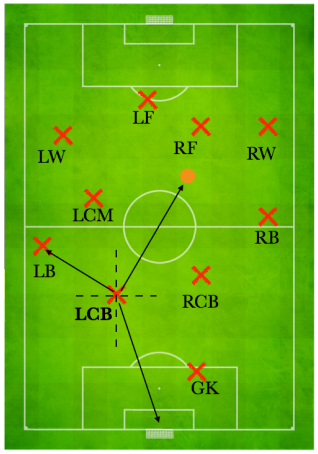

Joel Burdick


Swarat Chaudhuri

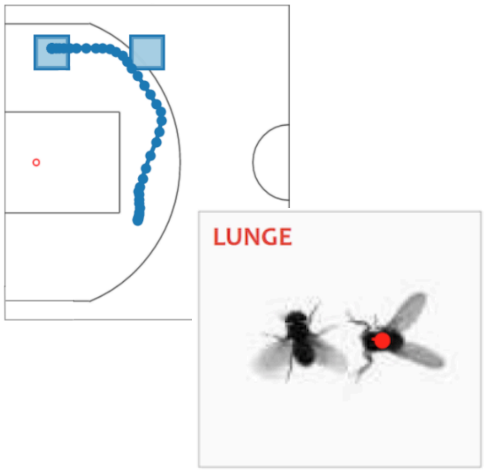
Episodic Learning with Control Lyapunov Functions for Uncertain Robotic Systems
Andrew Taylor, Victor Dorobantu, et al., IROS 2019
Imitation-Projected Programmatic Reinforcement Learning,
Abhinav Verma, Hoang Le, et al., NeurIPS 2019



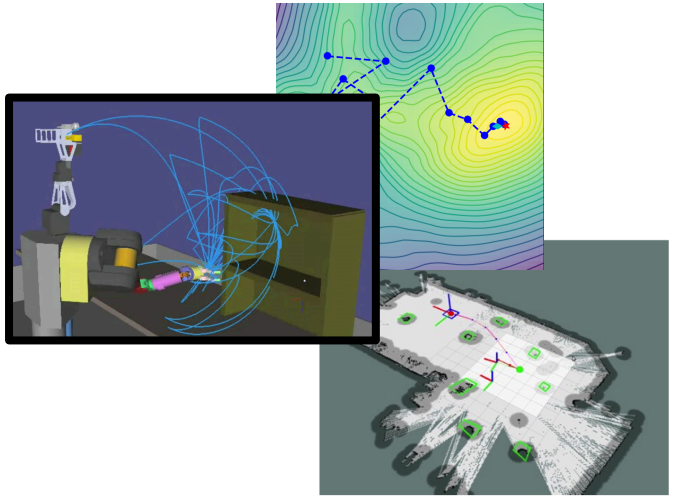
Speech Animation



Coordinated Learning



Hierarchical Behaviors (Generative)

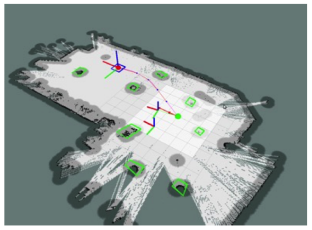


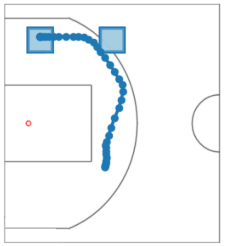
Learning to Optimize

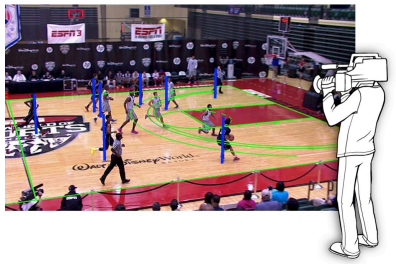
Side Constraints (Smoothness)

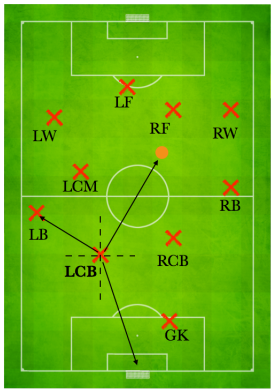
New Frontiers in Imitation learning

- **Incorporating Structure**
  - Dynamics of output space
  - Latent structure of input space
  - New feedback oracles
- **New Algorithmic Frameworks**
  - Black Box + Dynamics Model
  - Black Box + Latent Graphical Model
  - Retrospective Imitation
- **Cool Applications!**

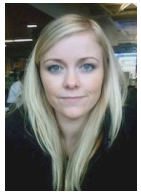






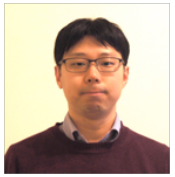



Eric Zhan


Eyrun Eyolfsdottir


Stephan Zheng


Hoang Le


Taehwan Kim


Sarah Taylor


Stephane Ross


Jialin Song


Joe Marino


Richard Cheng


Ravi Lanka

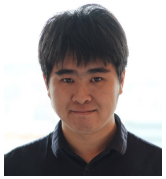

Andrew Taylor


Victor Dorobantu


Jimmy Chen

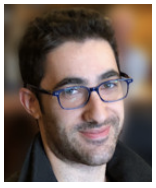

Andrew Kang


Meera Krishnamoorthy


Yukai Liu

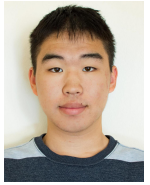

Aadyot Bhatnagar


Albert Zhao

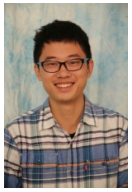

Cameron Voloshin


Abhinav Verma


Milan Cvitkovic


Albert Tseng


Stephan Mandt


Robin Zhou


Patrick Lucey


Gabor Orosz


Debadeepta Dey


Kristin Branson


Jim Little


Peter Carr


Rose Yu


Adith Swaminathan


Matthew Hausknecht


Aaron Ames


Iain Matthews


Pietro Perona


Drew Bagnell


Miro Dudik


Alekh Agarwal

Nan Jiang

Joel Burdick

Swarat Chaudhuri

Hal Daume

Masahiro Ono

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

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

Generating Long-term Trajectories using Deep Hierarchical Networks, Stephan Zheng et al., NeurIPS 2016

Generating Multi-agent Trajectories using Programmatic Weak Supervision, Eric Zhan et al., ICLR 2019

Learning Calibratable Policies using Programmatic Style-Consistency, Eric Zhan et al., arxiv

NAOMI: Non-autoregressive Multiresolution Sequence Imputation, Yukai Liu et al., NeurIPS 2019

Learning recurrent representations for hierarchical behavior modeling, Eyrun Eyolfsdottir et al., ICLR 2017

Data-Driven Ghosting using Deep Imitation Learning, Hoang Le et al., SSAC 2017 *(Best Paper Runner Up)*

Coordinated Multi-agent Imitation Learning, Hoang Le et al., ICML 2017

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

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

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

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

A General Method for Amortizing Variational Filtering, Joseph Marino et al., NeurIPS 2018

Hierarchical Imitation and Reinforcement Learning, Hoang Le et al., ICML 2018

Episodic Learning with Control Lyapunov Functions for Uncertain Robotic Systems, Andrew Taylor, Victor Dorobantu, et al., IROS 2019

A Control Lyapunov Perspective on Episodic Learning via Projection to State Stability, Andrew Taylor, Victor Dorobantu, et al., CDC 2019

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