Structured Imitation and Reinforcement Learning

Yisong Yue

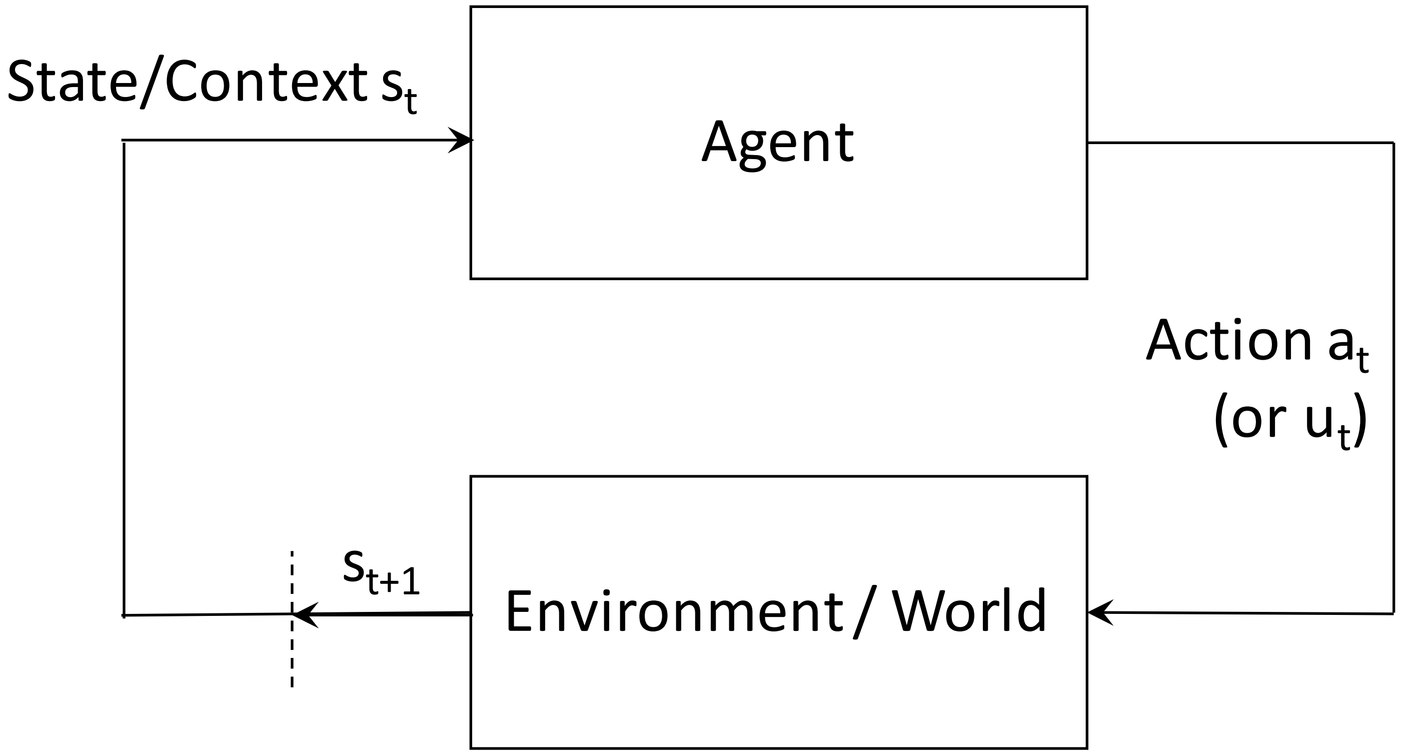
Imitation & Reinforcement Learning

Goal: Find "Optimal" Policy

Imitation Learning:
Optimize imitation loss

Reinforcement Learning:
Optimize environmental reward

(Known Dynamics => Optimal Control)



Imitation Learning Tutorial

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

Yisong Yue	Hoang M. Le
	
yyue@caltech.edu	hmle@caltech.edu
@YisongYue	@HoangMinhLe
yisongyue.com	hoangle.info

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

Starting Point

Standard IL/RL Objective

$$\operatorname{argmin}_h L(h)$$

$$s.t.$$

$$R(h) < \delta$$

Side Constraint

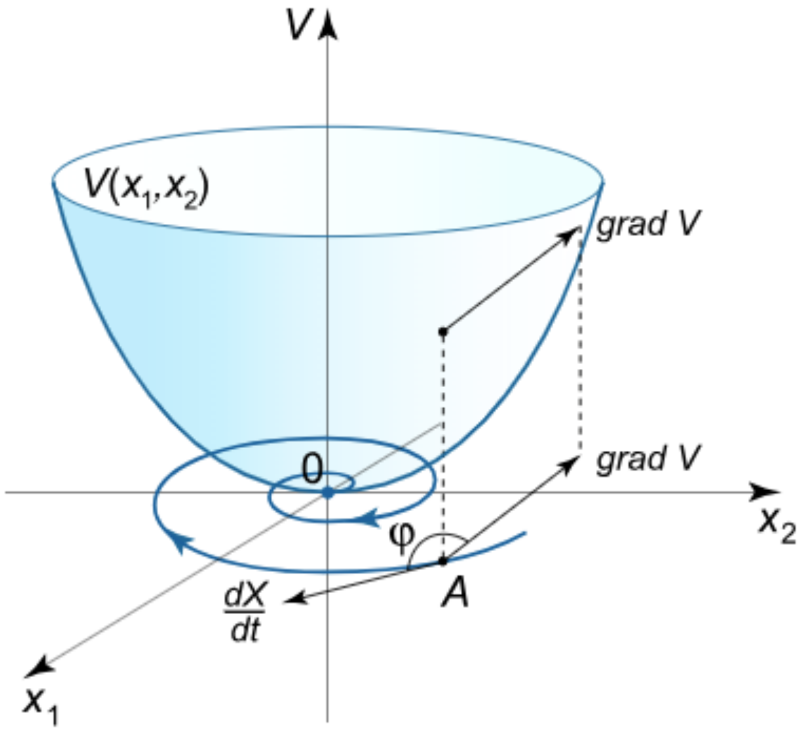
- Model-Based/Free
- On/Off Policy
- Imitation/Reinforcement
- Optimal Control

What can R encode?

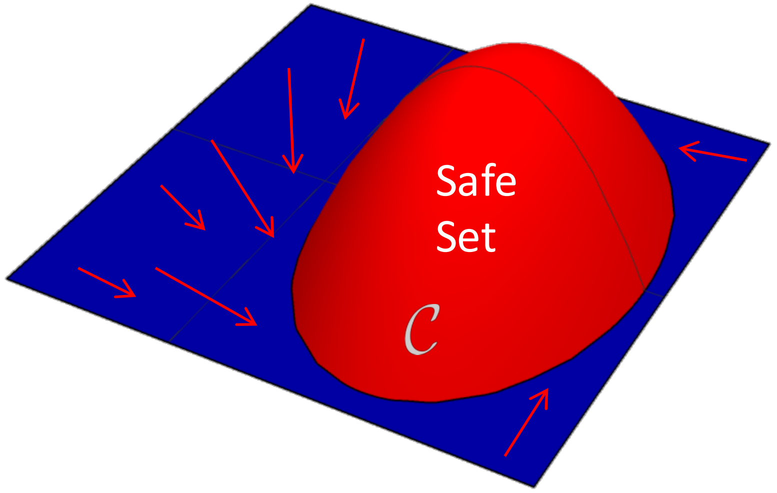
Side Guarantees

And Possibly Others:

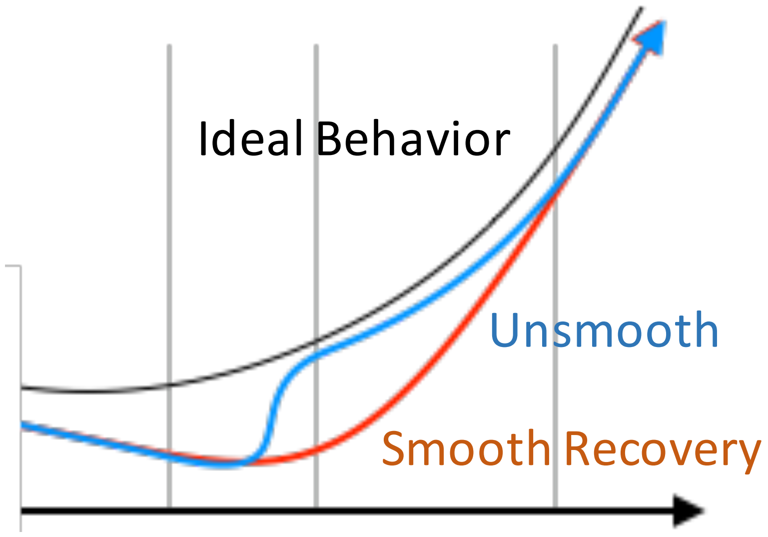
- Fairness
- Low-risk
- Temporal logic
- Etc...



Stability



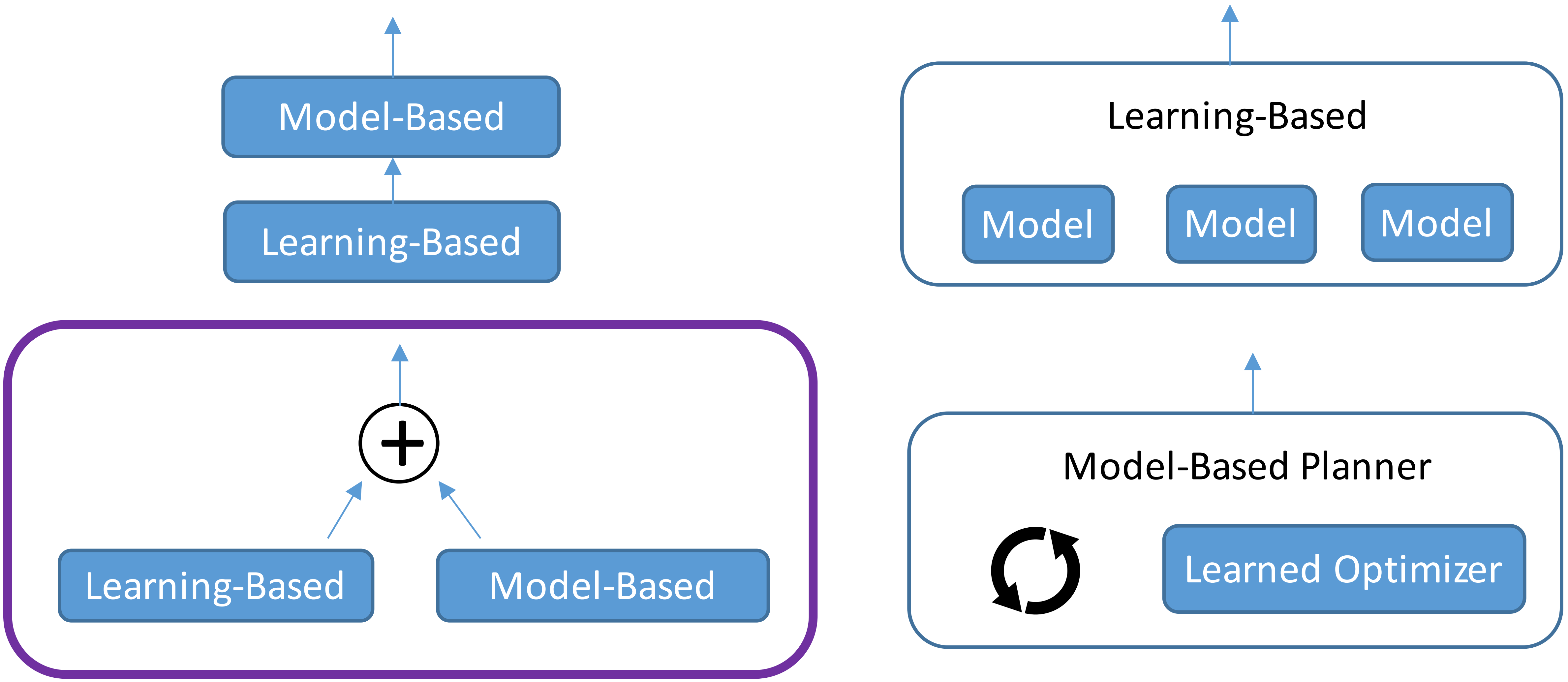
Safety

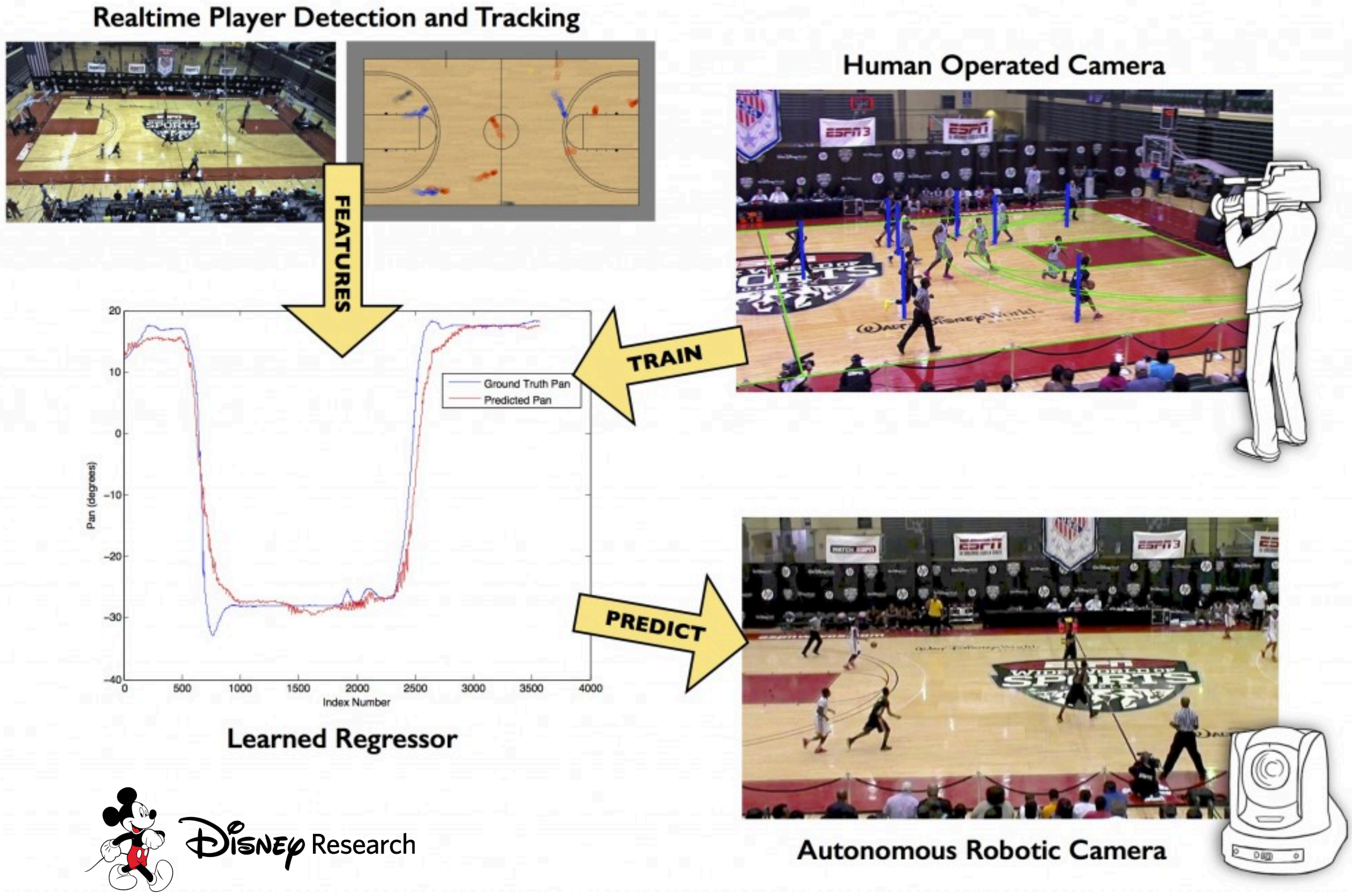


Smoothness

(Robust Control => Robust to Model Uncertainty)

Blending Models & Black-Box Learning

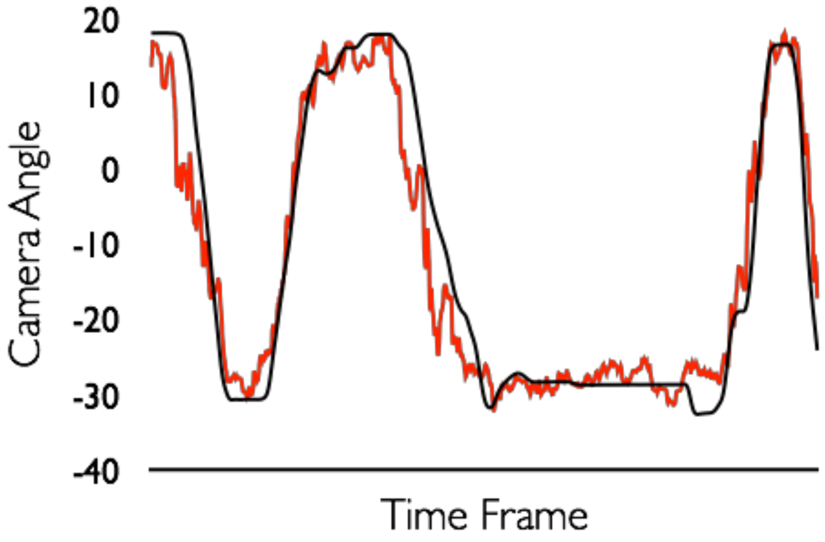




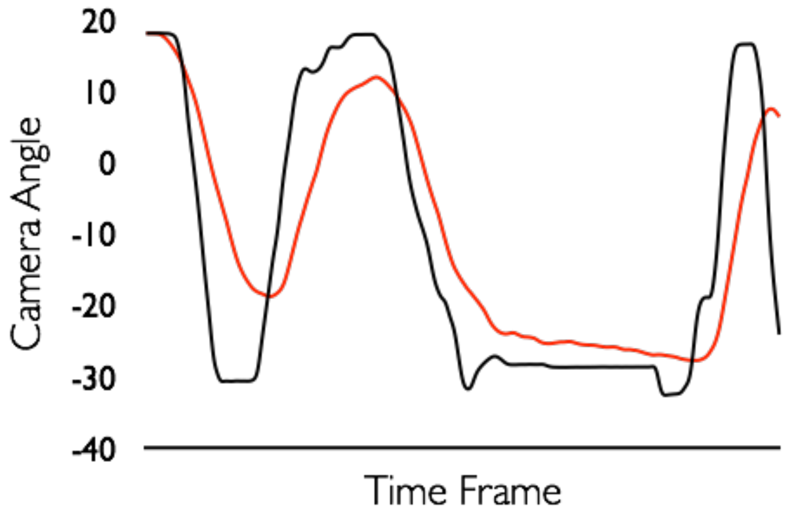
Realtime Player Detection and Tracking
FEATURES
Human Operated Camera
TRAIN
PREDICT
Ground Truth Pan
Predicted Pan
Pan (degrees)
Index Number
20
10
0
−10
−20
−30
−40
0
500
1000
1500
2000
2500
3000
3500
4000
Learned Regressor
Autonomous Robotic Camera
Disney Research

Naïve Approach

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



In practice, post-hoc smoothing:



Starting Point

IL/RL Objective

$$\begin{aligned} &argmin_h L(h) \\ &s.t. \\ &\quad R(h) < \delta \end{aligned}$$

$\Leftrightarrow$

$$argmin_h L(h) + \lambda R(h)$$

Smoothness Constraint

What is R?

Regularize to Function Class

(h is "close to" some g)

$$\begin{aligned} &\quad argmin_h L(h) \\ &s.t. \\ &\exists g \in G: \|h - g\|^2 < \delta \end{aligned}$$

$\Leftrightarrow$

$$argmin_{h,g} L(h) + \lambda \|h - g\|^2$$

Model-Based Controllers

(provably smooth)

Intractable?

Alternative Formulation

$$argmin_h L(h)$$

$$s.t.$$

$$h \; is \; close \; to \; f \in F$$

$$h \; is \; close \; to \; g \in G$$

Deep Neural Nets
Random Forests
Etc…

Model-Based Controllers
(provably smooth)

Smooth Policy Class


Hoang Le

Policy

Black Box Predictor

Smooth Model

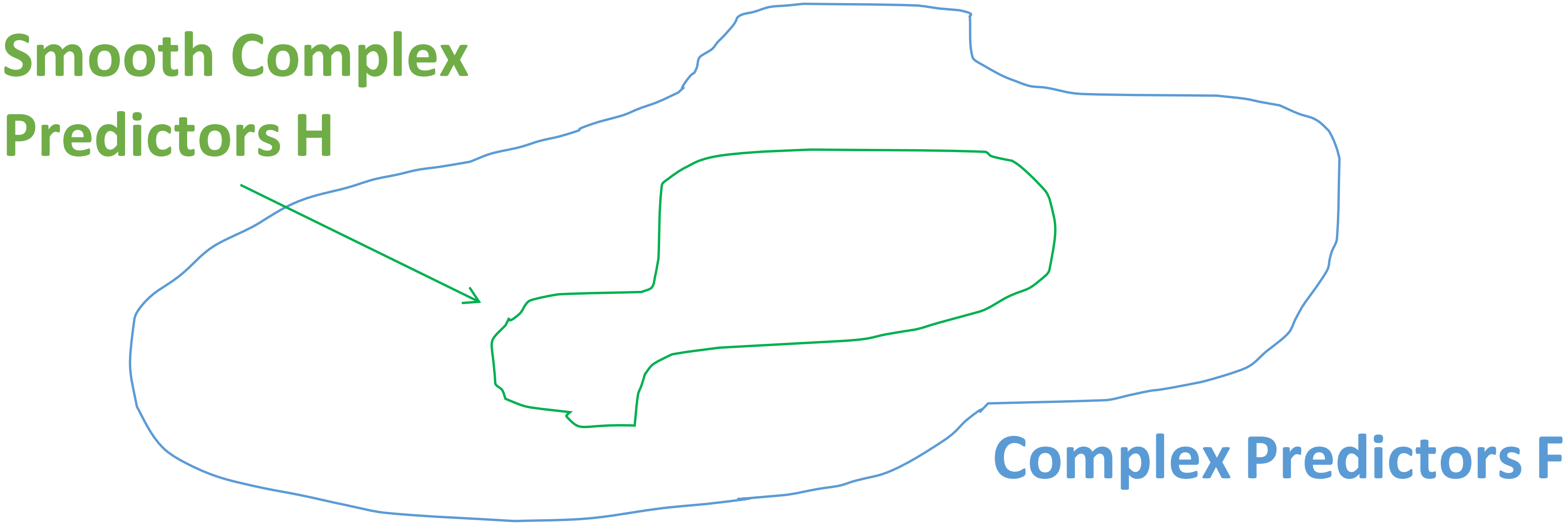
$$argmin_{h=(f,g)} L(h) \quad \text{s.t.} \quad h(s) = argmin_{a'} (f(s) - a')^2 + \gamma(g(s) - a')^2$$

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

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

Test-Time Functional Regularization


Hoang Le



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

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

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

Test-Time Functional Regularization

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

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

- By construction: h "close" to g

Run-time regularization

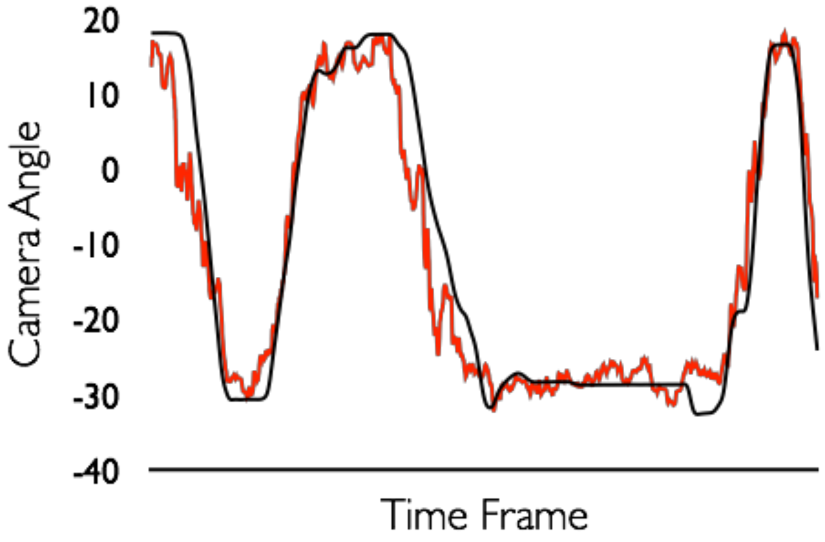
- Certifications on g => (relaxed) certifications on h

E.g., "smoothness"

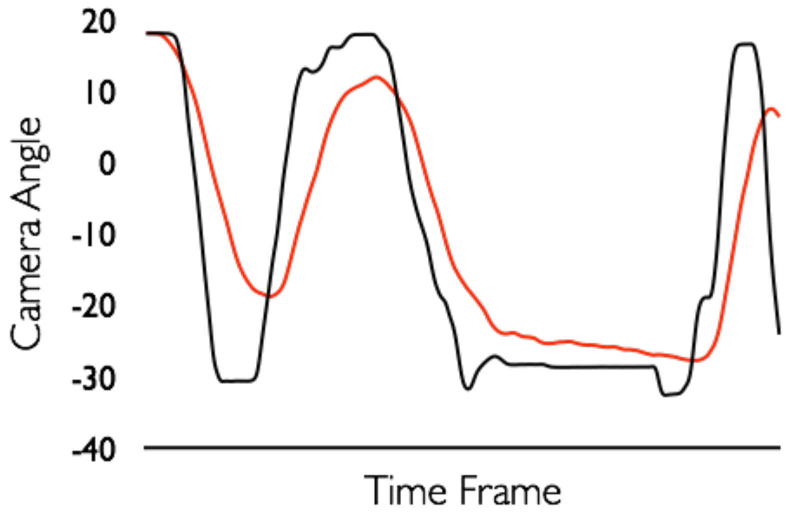
- Compatible with many forms of IL/RL

Reminder: Naïve Approach

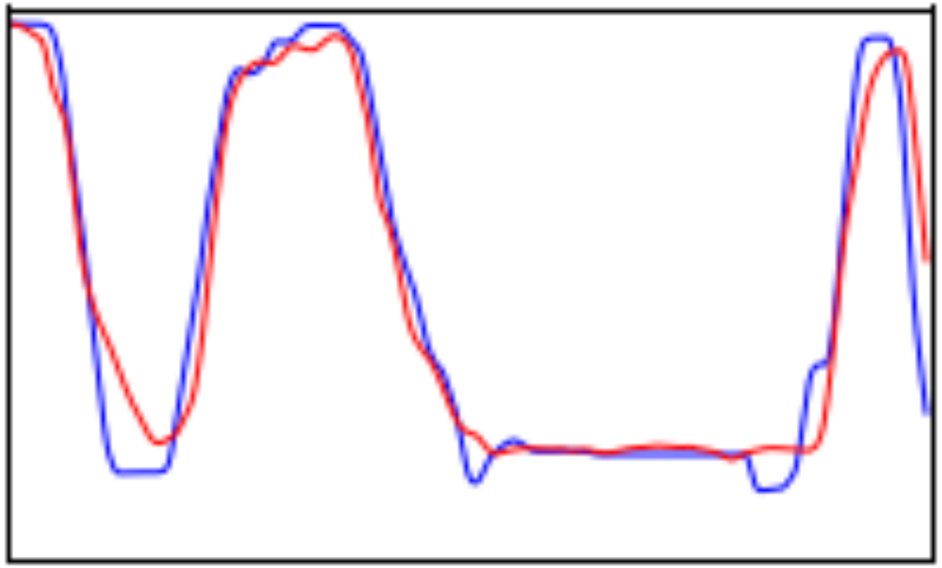
- Supervised learning of demonstration data
  - Train predictor per frame
  - Predict per frame

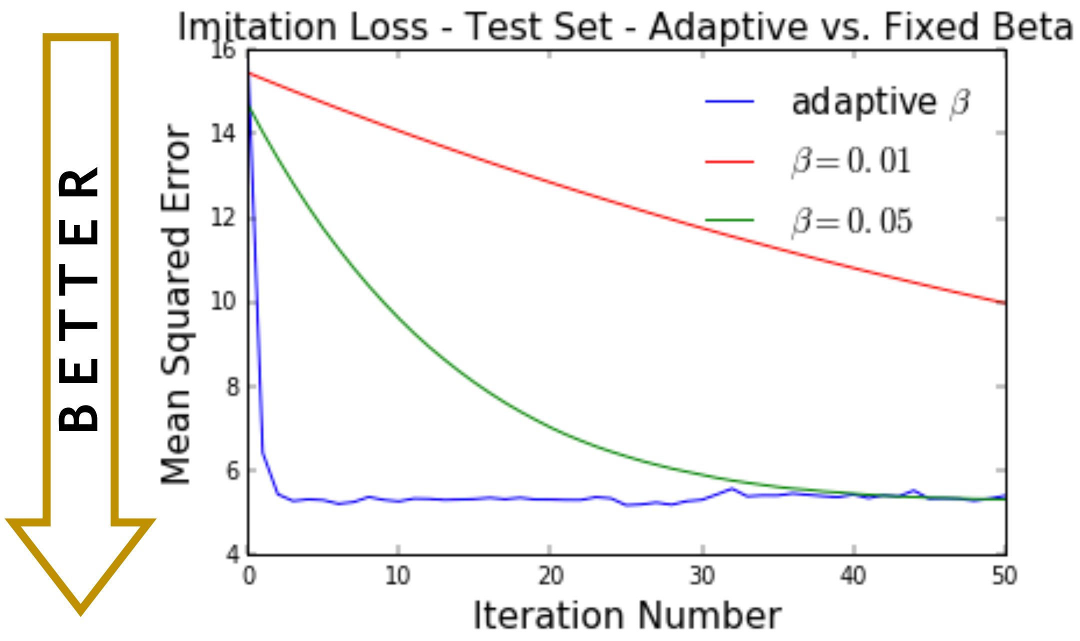


In practice, post-hoc smoothing:



Our Results



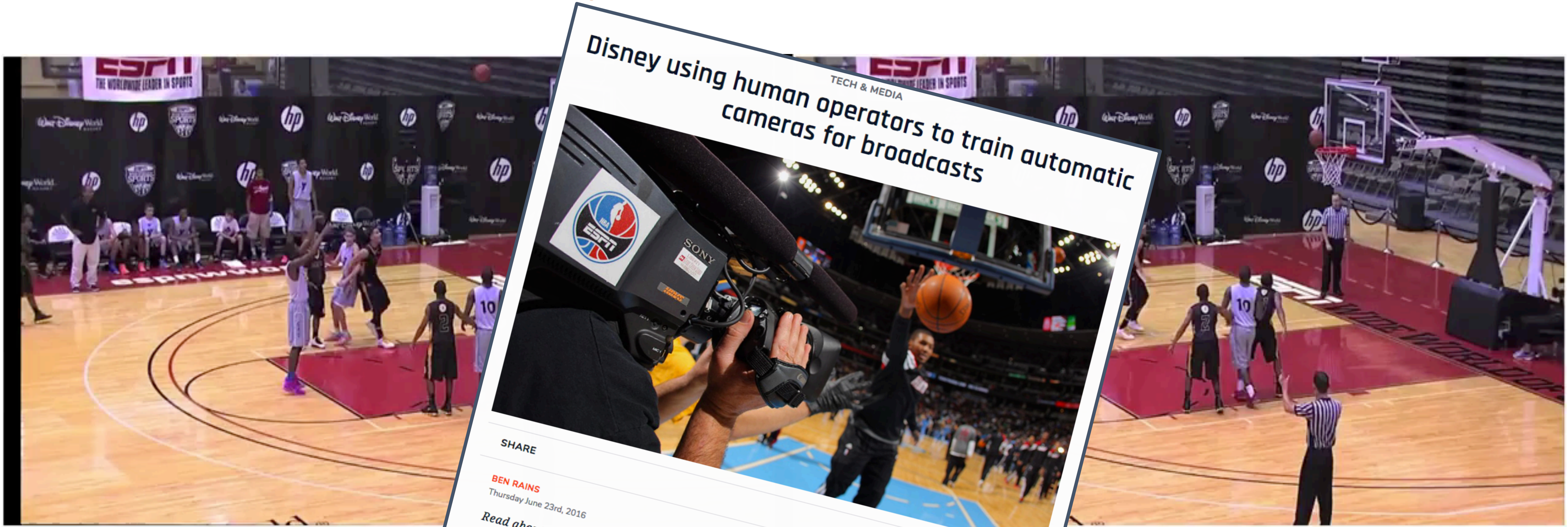


Provably Smooth Predictions

Provably Faster Learning
(Natural Policy Updates)

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

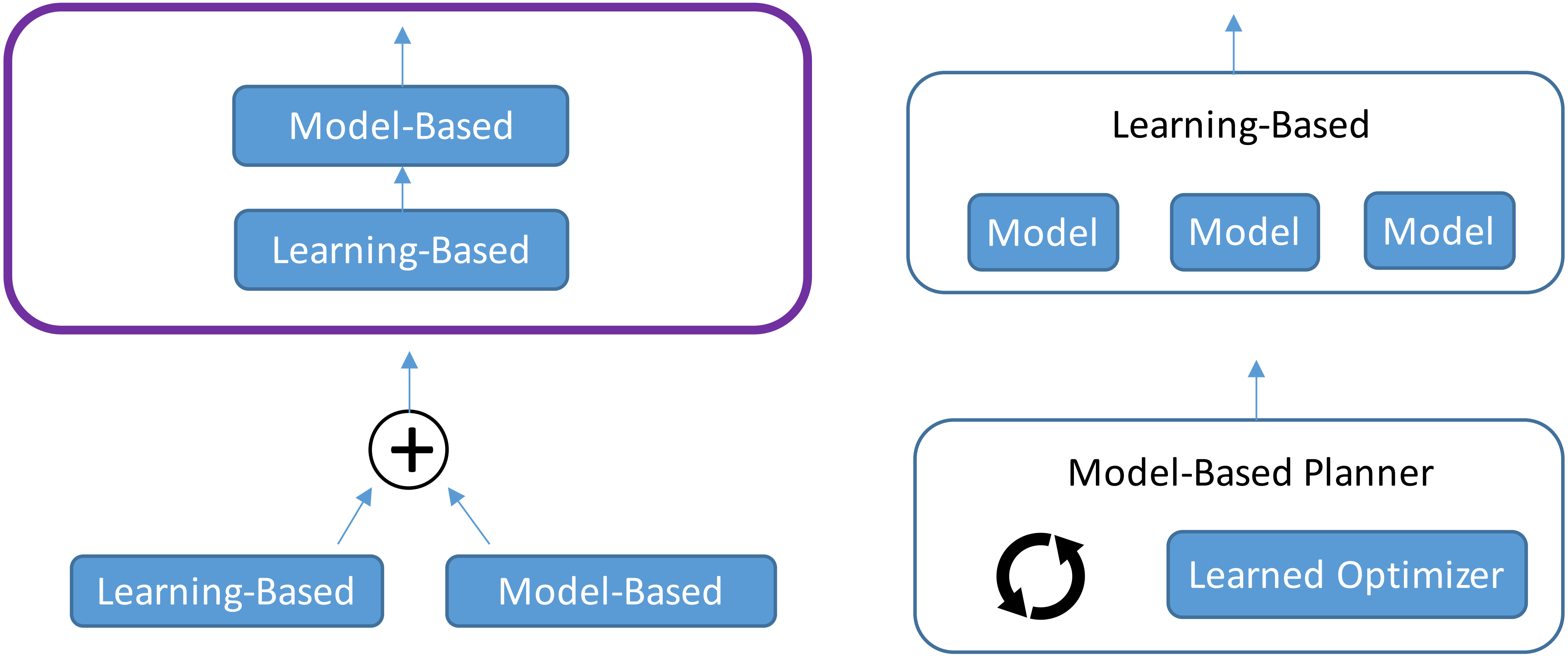
Read about the latest sports tech news, innovations, ideas and products that impact players, fans and the sports industry overall at SportTechie.com.

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 camera operators to help create a smoother and cleaner broadcast.

Learning Online Smooth P... g Recurrent Decision Trees

Jianhui Chen, Hoang Le, Peter Carr, ...

Blending Models & Black-Box Learning



Model-Based Control

New State

Current Action (aka control input)

$$s_{t+1} = F(s_t, u_t) + \epsilon$$

Current State

Unmodeled Disturbance / Error

(Value Iteration is also contraction mapping)

Robust Control (fancy contraction mappings)

- Stability guarantees (e.g., Lyapunov)
- Precision/optimality depends on error

Learning Residual Dynamics

F = nominal dynamics
$\tilde{F}$ = learned dynamics

New State

Current Action (aka control input)

$$s_{t+1} = F(s_t, u_t) + \tilde{F}(s_t, u_t) + \epsilon$$

Current State

Unmodeled Disturbance / Error

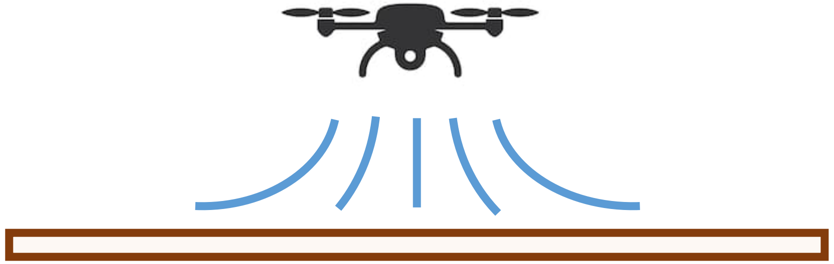
Leverage robust control (fancy contraction mappings)

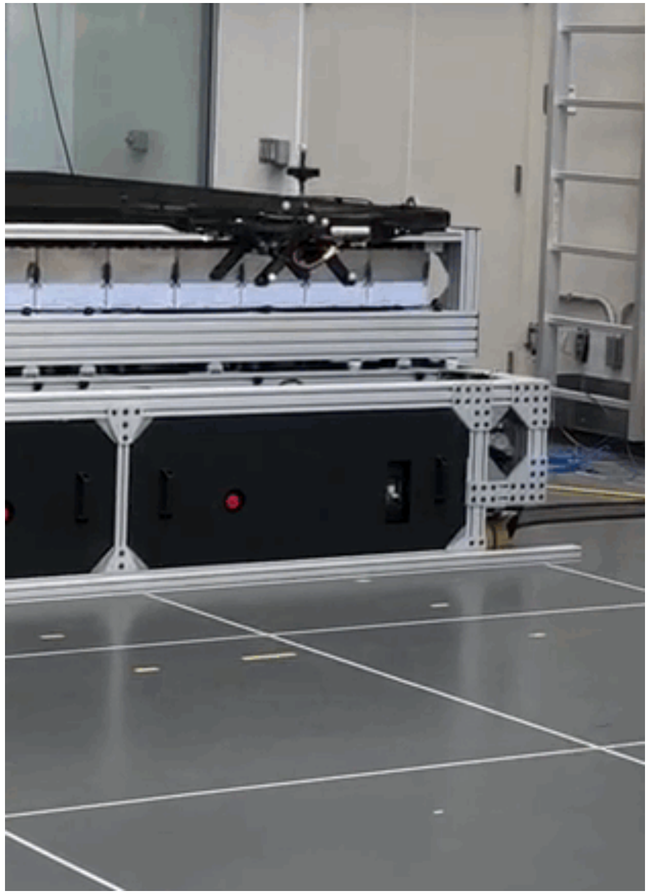
- Preserve stability (even using deep learning)
- Requires $\tilde{F}$ Lipschitz & bounded error

Stable Drone Landing

Guanya Shi

Ground effect





Neural Lander: Stable Drone Landing Control using Learned Dynamics
Guanya Shi, Xichen Shi, Michael O'Connell, Rose Yu, Kamyar Azizzadenesheli, Anima Anandkumar, Yisong Yue, Soon-Jo Chung. arXiv

Control System Formulation

Learn the Residual

- Dynamics:

$$\begin{cases} \dot{\mathbf{p}} = \mathbf{v}, & m\dot{\mathbf{v}} = m\mathbf{g} + R\mathbf{f}_u + \mathbf{f}_a \\ \dot{R} = RS(\boldsymbol{\omega}), & J\dot{\boldsymbol{\omega}} = J\boldsymbol{\omega} \times \boldsymbol{\omega} + \boldsymbol{\tau}_u + \boldsymbol{\tau}_a \end{cases}$$

- Control:

$$\begin{cases} \mathbf{f}_u = [0, 0, T]^\top \\ \boldsymbol{\tau}_u = [\tau_x, \tau_y, \tau_z]^\top \\ \begin{bmatrix} T \\ \tau_x \\ \tau_y \\ \tau_z \end{bmatrix} = \begin{bmatrix} c_T & c_T & c_T & c_T \\ 0 & c_T l_{\text{arm}} & 0 & -c_T l_{\text{arm}} \\ -c_T l_{\text{arm}} & 0 & c_T l_{\text{arm}} & 0 \\ -c_Q & c_Q & -c_Q & c_Q \end{bmatrix} \begin{bmatrix} n_1^2 \\ n_2^2 \\ n_3^2 \\ n_4^2 \end{bmatrix} \end{cases}$$

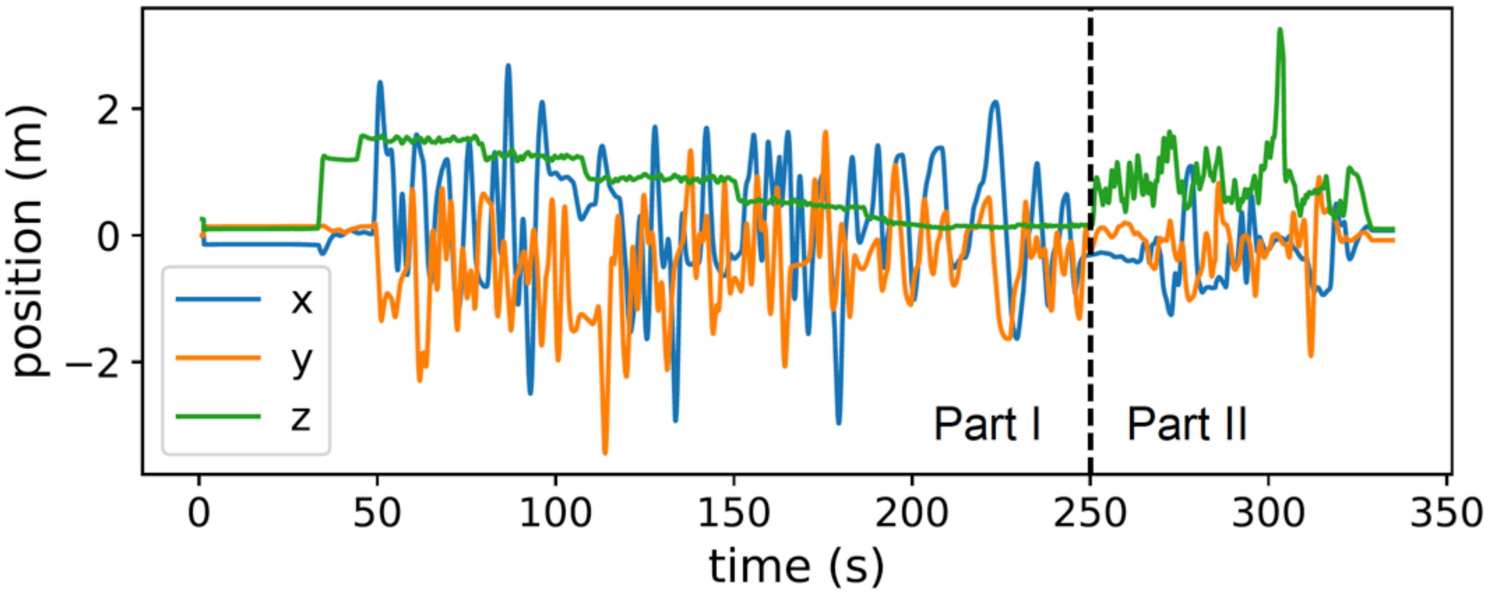
- Unknown forces & moments:

$$\begin{cases} \mathbf{f}_a = [f_{a,x}, f_{a,y}, f_{a,z}]^\top \\ \boldsymbol{\tau}_a = [\tau_{a,x}, \tau_{a,y}, \tau_{a,z}]^\top \end{cases}$$

Learn the Residual

Data Collection (Manual Exploration)

Current Research:
Safe Exploration

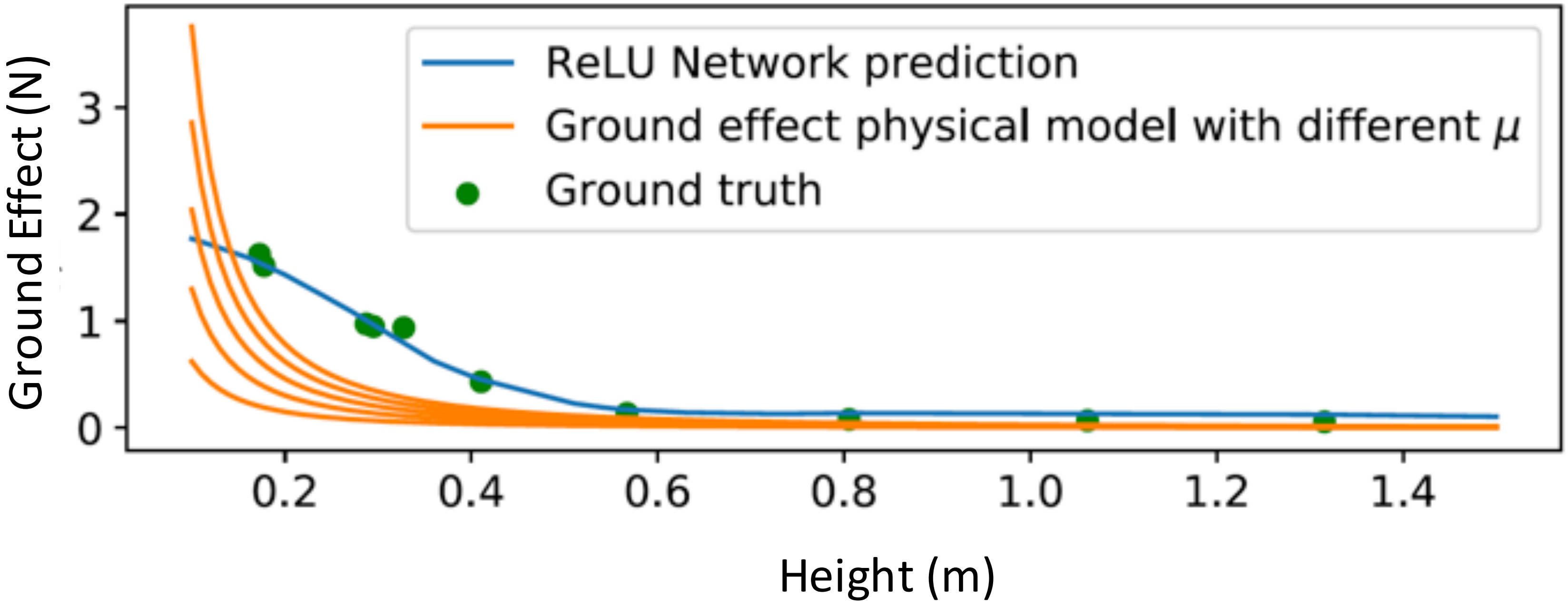


Ensures $\widetilde{F}$ is Lipshitz
[Bartlett et al., NeurIPS 2017]
[Miyato et al., ICLR 2018]

Spectral-Normalized 4-Layer Feed-Forward

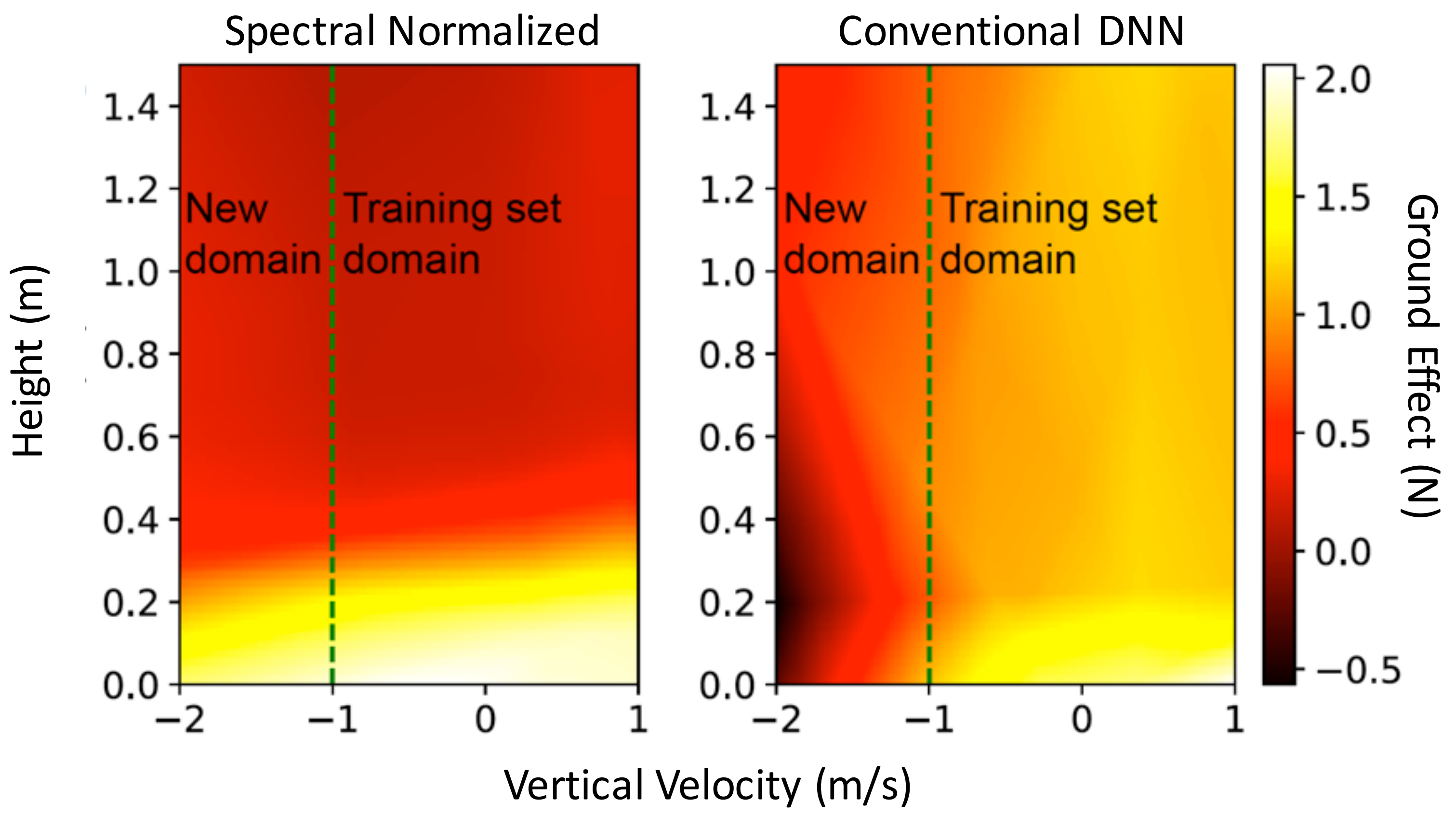
- Learn ground effect: $\tilde{F}(s,u) \rightarrow \mathbf{f}_a = [f_{a,x}, f_{a,y}, f_{a,z}]^\top$
- (s,u): height, velocity, attitude and four control inputs

Prediction Results



Neural Lander: Stable Drone Landing Control using Learned Dynamics
Guanya Shi, Xichen Shi, Michael O'Connell, Rose Yu, Kamyar Azizzadenesheli, Anima Anandkumar, Yisong Yue, Soon-Jo Chung. arXiv

Prediction Results



Neural Lander: Stable Drone Landing Control using Learned Dynamics

Guanya Shi, Xichen Shi, Michael O'Connell, Rose Yu, Kamyar Azizzadenesheli, Anima Anandkumar, Yisong Yue, Soon-Jo Chung. arXiv

Controller Design (simplified)

Guanya Shi

- Nonlinear Feedback Linearization:

$$u_{nominal} = K_s \eta \qquad \eta = \begin{bmatrix} p - p^* \\ v - v^* \end{bmatrix}$$

Desired Trajectory (tracking error)

Feedback Linearization (PD control)

- Cancel out ground effect $\tilde{F}(s, u_{old})$: $u = u_{nominal} + u_{residual}$

Requires Lipschitz & small time delay

Controller Design (simplified)


Guanya Shi

- Nonlinear Feedback Linearization:

$$u_{nominal} = K_s \eta \qquad \eta = \begin{bmatrix} p - p^* \\ v - v^* \end{bmatrix}$$

Desired Trajectory
(tracking error)

Stability Guarantee:
(simplified)

Time delay

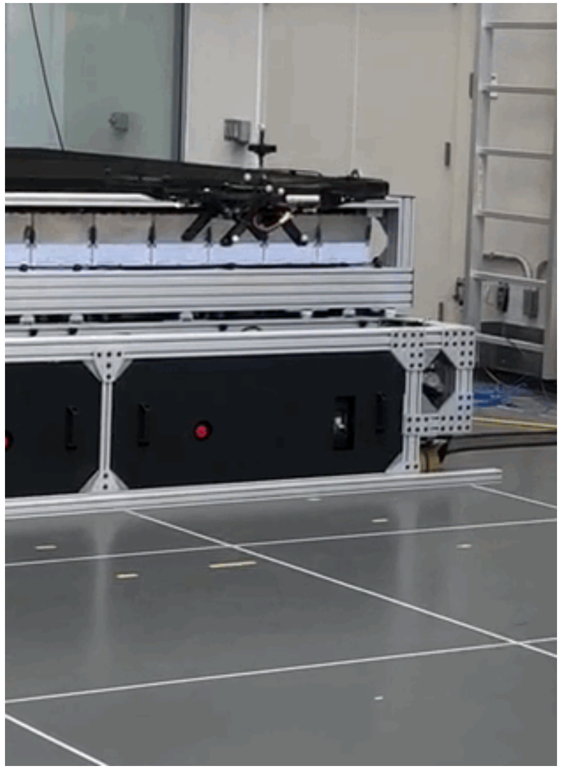
Unmodeled disturbance

$$\|\eta(\mathrm{t})\| \leq \|\eta(0)\| \exp\left\{\frac{\lambda_{min}(K) - \tilde{L}\rho}{C} t\right\} + \frac{\epsilon}{\lambda_{min}(K) - \tilde{L}\rho}$$

Lipschitz of NN

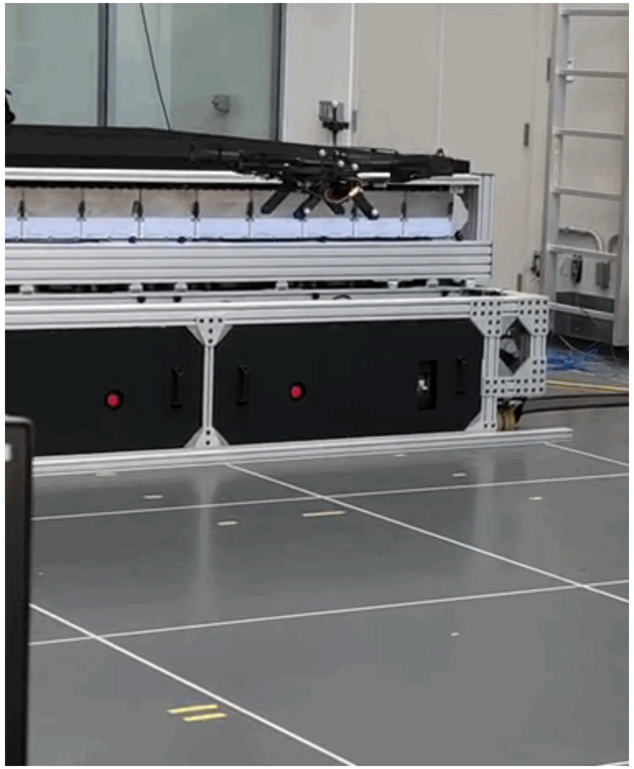
$$\Longrightarrow \|\eta(\mathrm{t})\| \rightarrow \frac{\epsilon}{\lambda_{min}(K) - \tilde{L}\rho}$$ Exponentially fast

Robust Landing Control



PD

PID



Neural-Lander (PD+Fa)

https://www.youtube.com/watch?v=C_K8MkC_SSQ

Takeaways

- Control methods => analytic guarantees

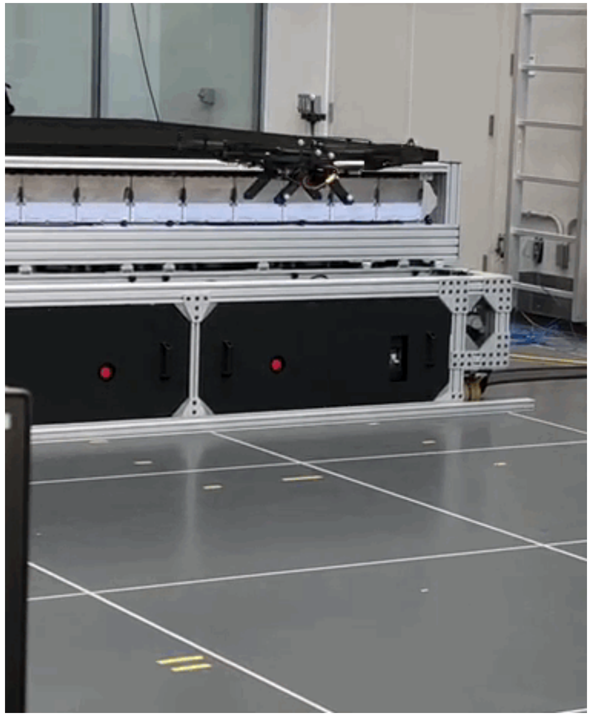
(side guarantees)

- Blend w/ learning => improve precision/flexibility

- Preserve side guarantees (sometimes relaxed)

- Sometimes interpret as functional regularization

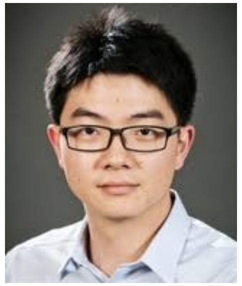
(speeds up learning)




Hoang Le


Guanya Shi

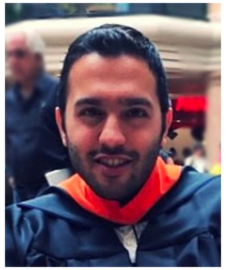

Jimmy Chen


Xichen Shi


Michael O'Connell


Andrew Kang


Rose Yu


Kamyar Azizzadenesheli


Peter Carr


Anima Anandkumar


Soon-Jo Chung


Jim Little

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

Learning Online Smooth Predictors for Real-time Camera Planning using Recurrent Decision Trees
Jianhui Chen, Hoang Le, Peter Carr, Yisong Yue, Jim Little. CVPR 2016

Neural Lander: Stable Drone Landing Control using Learned Dynamics
Guanya Shi, Xichen Shi, Michael O'Connell, Rose Yu, Kamyar Azizzadenesheli, Anima Anandkumar, Yisong Yue, Soon-Jo Chung. arXiv