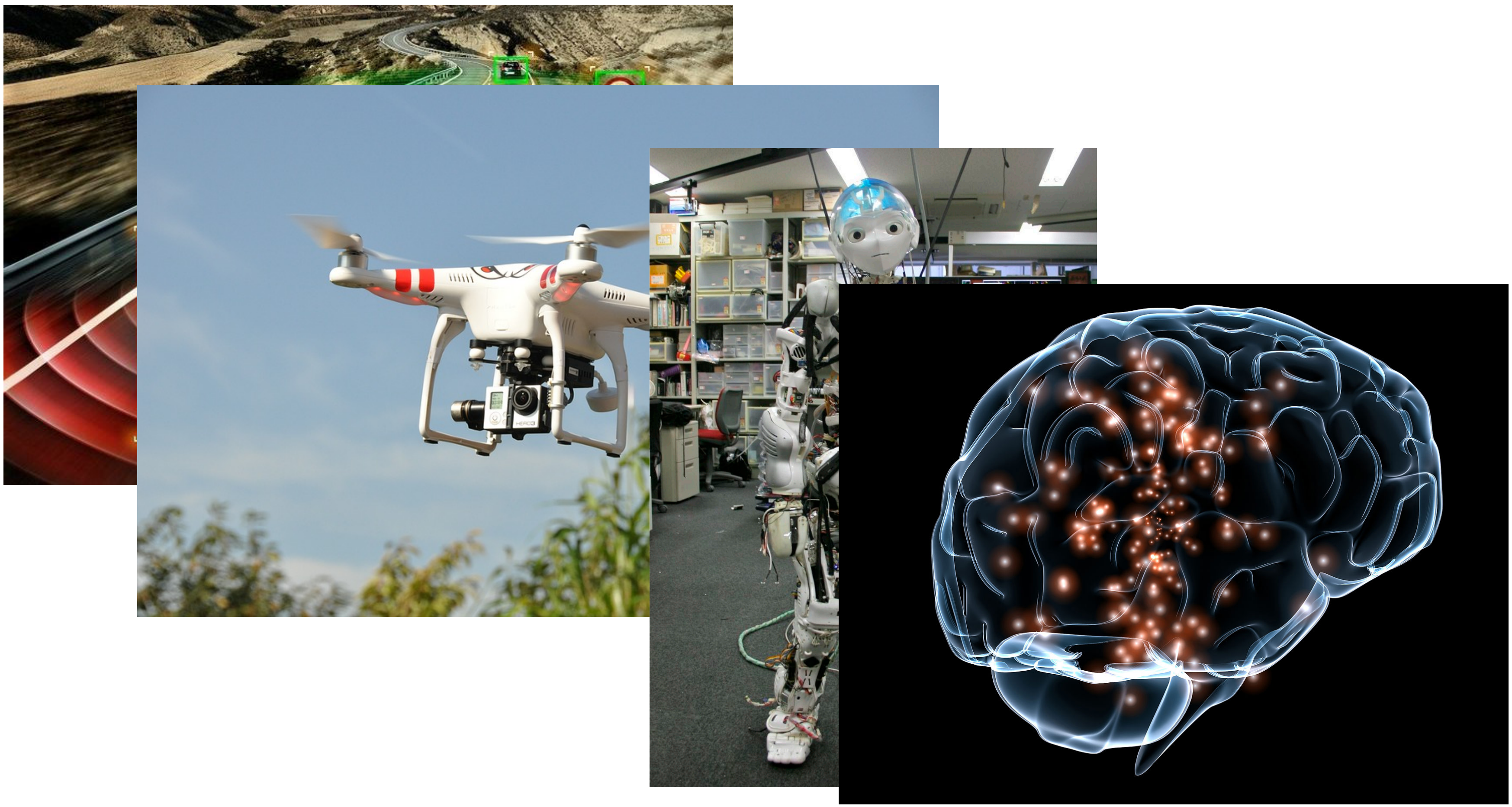
Learning to act by predicting the future

Alexey Dosovitskiy and Vladlen Koltun
Intel Labs, Santa Clara

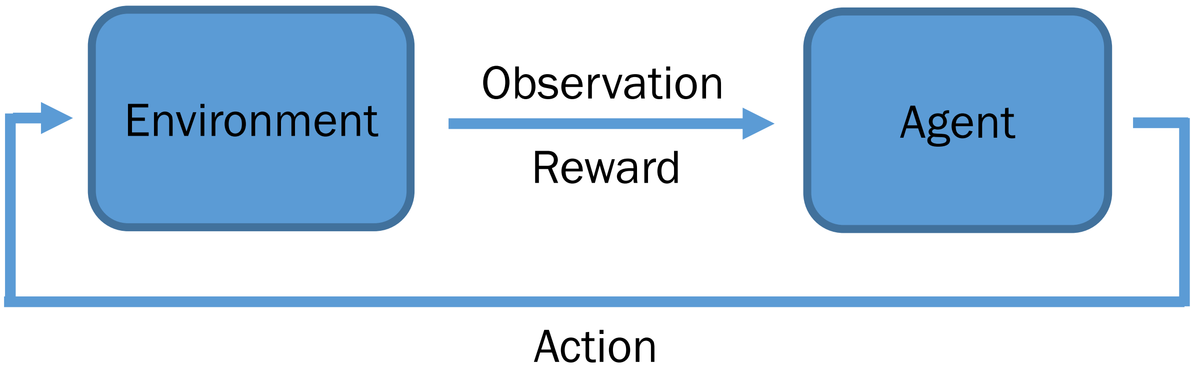
ICLR 2017, Toulon, France

Sensorimotor control

Aim: produce useful motor actions based on sensory inputs



Reinforcement Learning Formalism



- Aim: maximize the (discounted) sum of future rewards

Training signal

- Standard RL: scalar reward

- Reality: rich sensory stream

Goals

- Standard RL: a single goal given by the reward

- Reality: a wide variety of potential goals

Training procedure

- Standard RL: maximize future discounted reward

- Reality: learn about the world

Image from https://www.flickr.com/photos/lupuca/8720602322

Recent related work

- Training signal
  - Jaderberg et al., ICLR 2017
- Goals
  - Schaul et al., ICML 2015
- Training procedure
  - Exploration & intrinsic motivation
  - Osband et al. 2016, Osband et al. 2017, Houthooft et al. 2016, Fu et al. 2017, Bellemare et al. 2016, Usunier et al. 2016, Sukhbaatar et al. 2017, Gregor et al. 2016, …

Direct future prediction

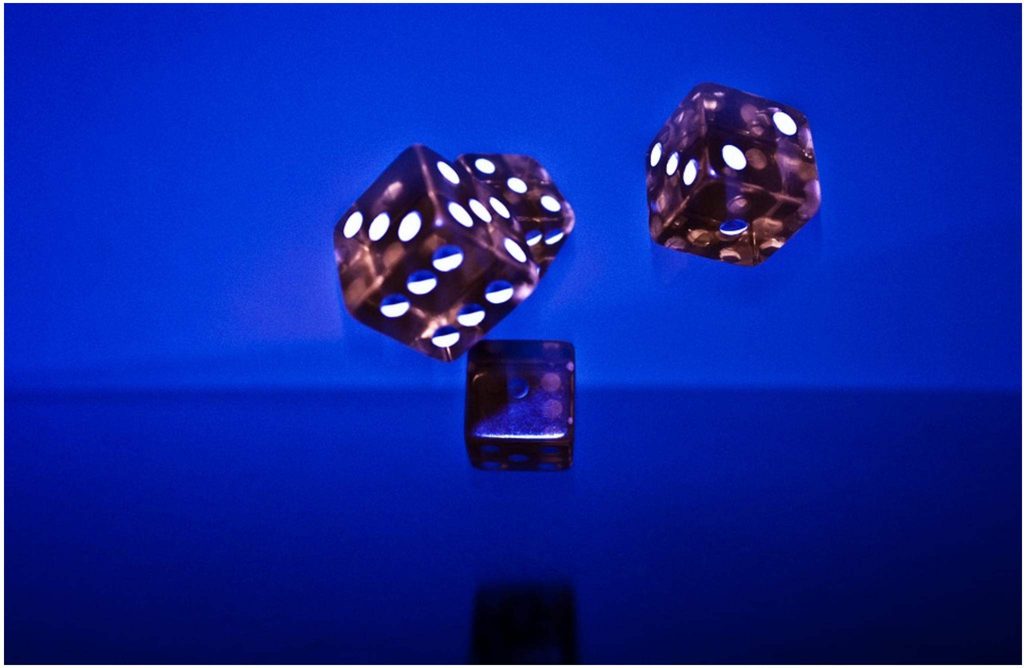
- Control as “future-supervised” learning

- Instead of learning to maximize returns, learn to predict the future

How to represent the future?

Naïve approach: predict pixels

- Simply predict the future observation
  - Oh et al. 2015, Finn et al. 2016, Chiappa et al. 2017, …
- Problem: uncertainty!



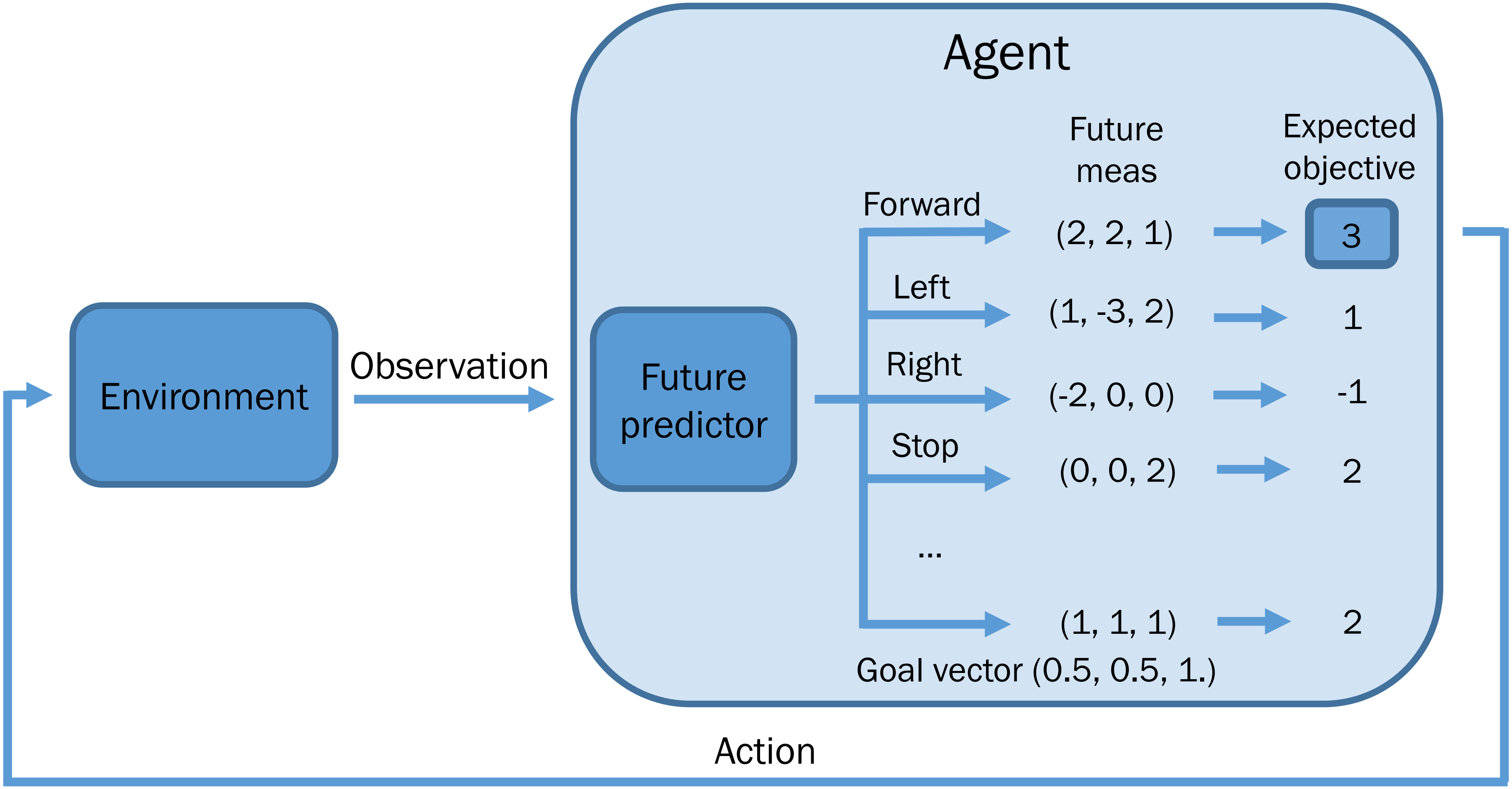
- We only need to predict relevant values

Our approach: predict measurements

- Predict the future values of measurements available to the agent

- Assumption: goals (objective functions) can be expressed as functions of measurements

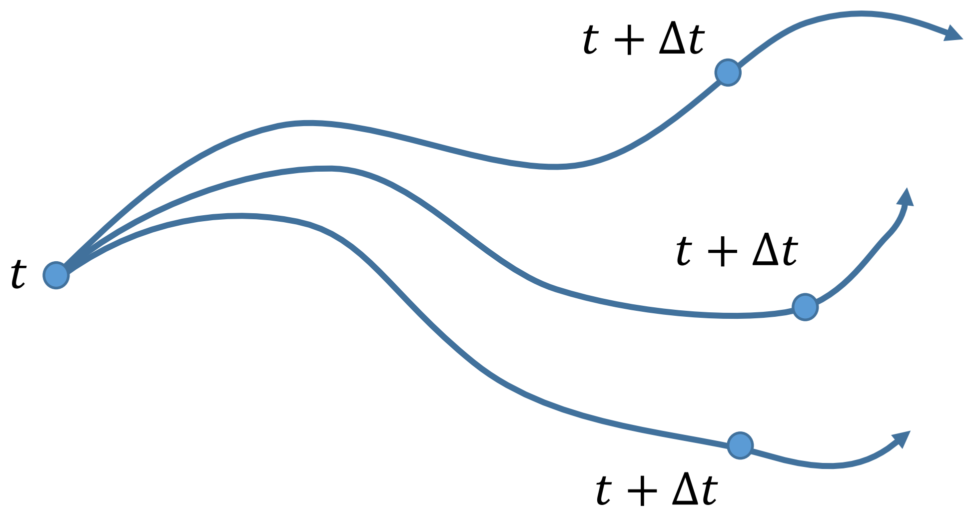
Using future predictions to act



Objective is linear in future measurements

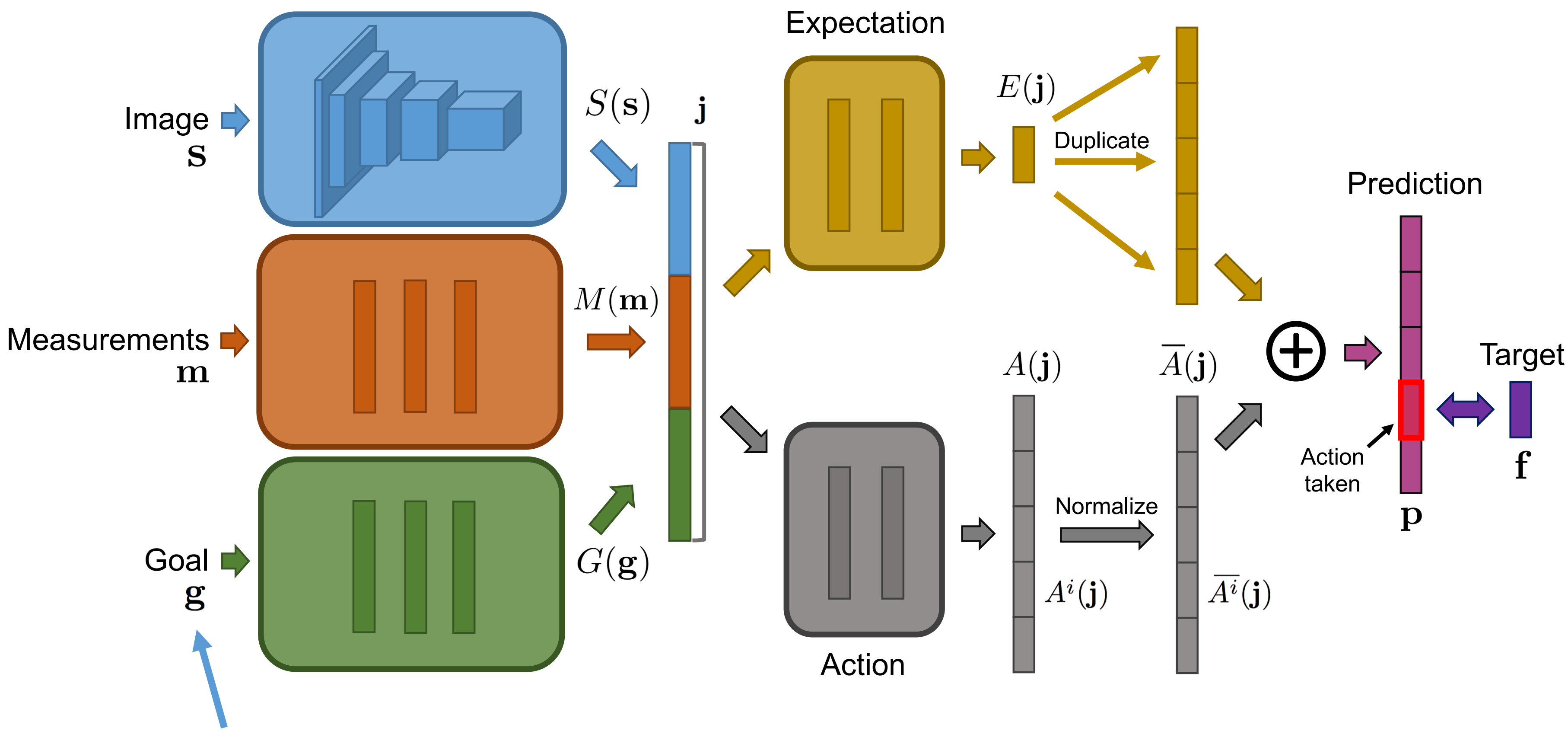
Direct future prediction

- Predict the future measurements for each action
- Simple supervised learning



- The future is stochastic — Predict expectations
- … and depends on the future actions — On-policy

Network architecture



Can change dynamically!

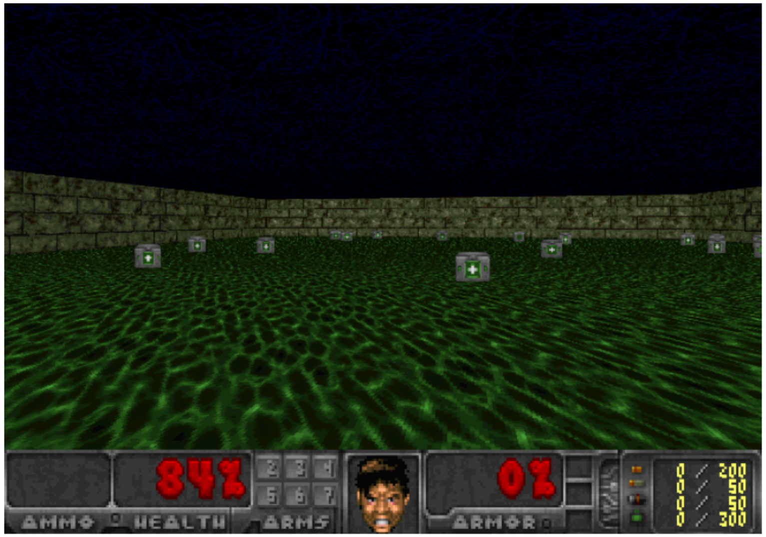
Technical details

- One frame as input, no memory
- Predict multiple future steps: 1,2,4,8,16,32
- Epsilon-greedy policy
- Small experience replay
- Parallel exploration – 8 copies of the agent

Experiments: ViZDoom environment

- Based on Doom
- Natural measurements: ammo, health, frags
- Wide variety of tasks and scenarios

ViZDoom – tasks



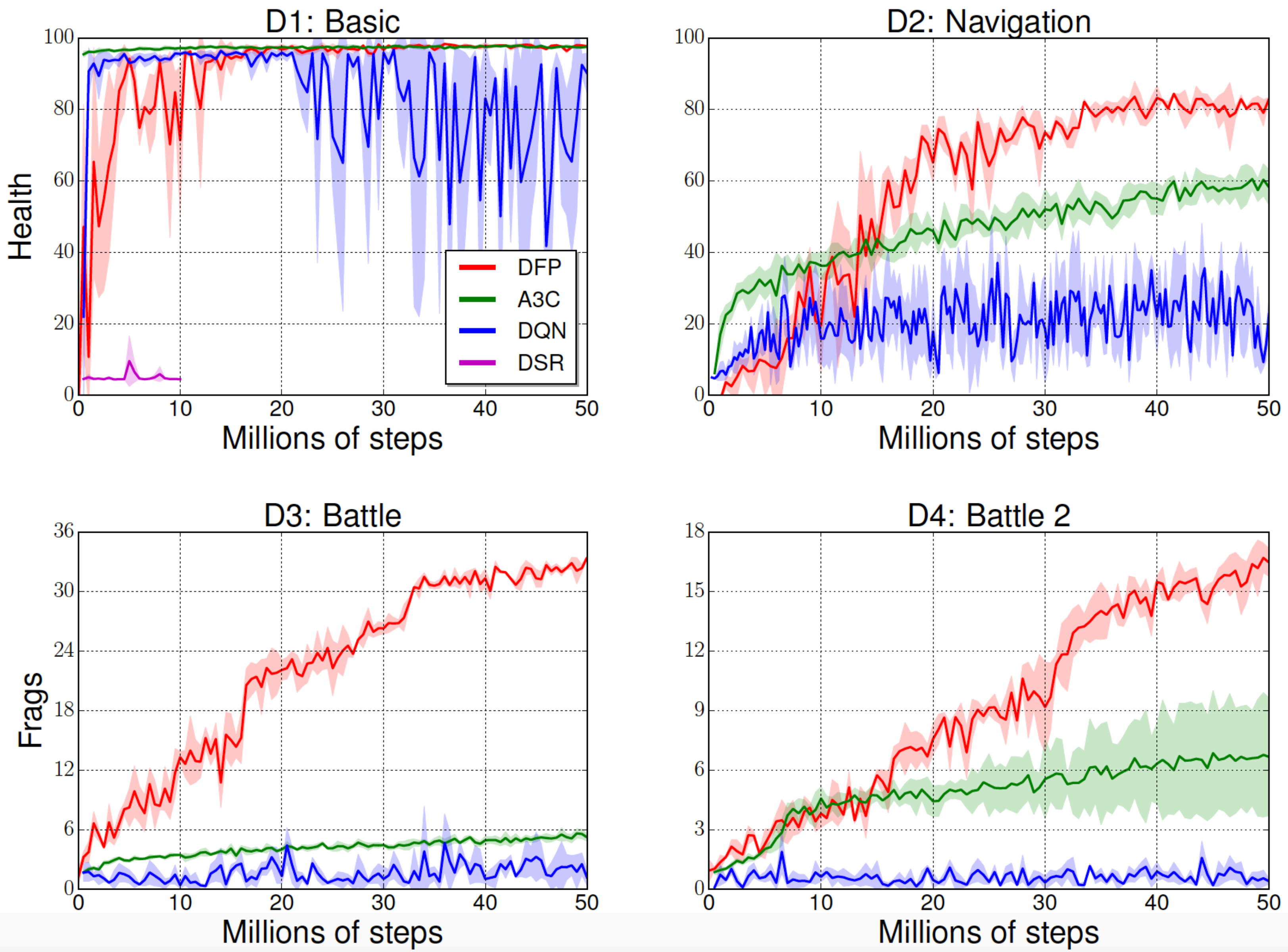
D1: Basic

D2: Navigation

D3: Battle

D4: Battle 2

Comparison to existing methods



Learning to Act by Predicting the Future

Alexey Dosovitskiy Vladlen Koltun

Generalization

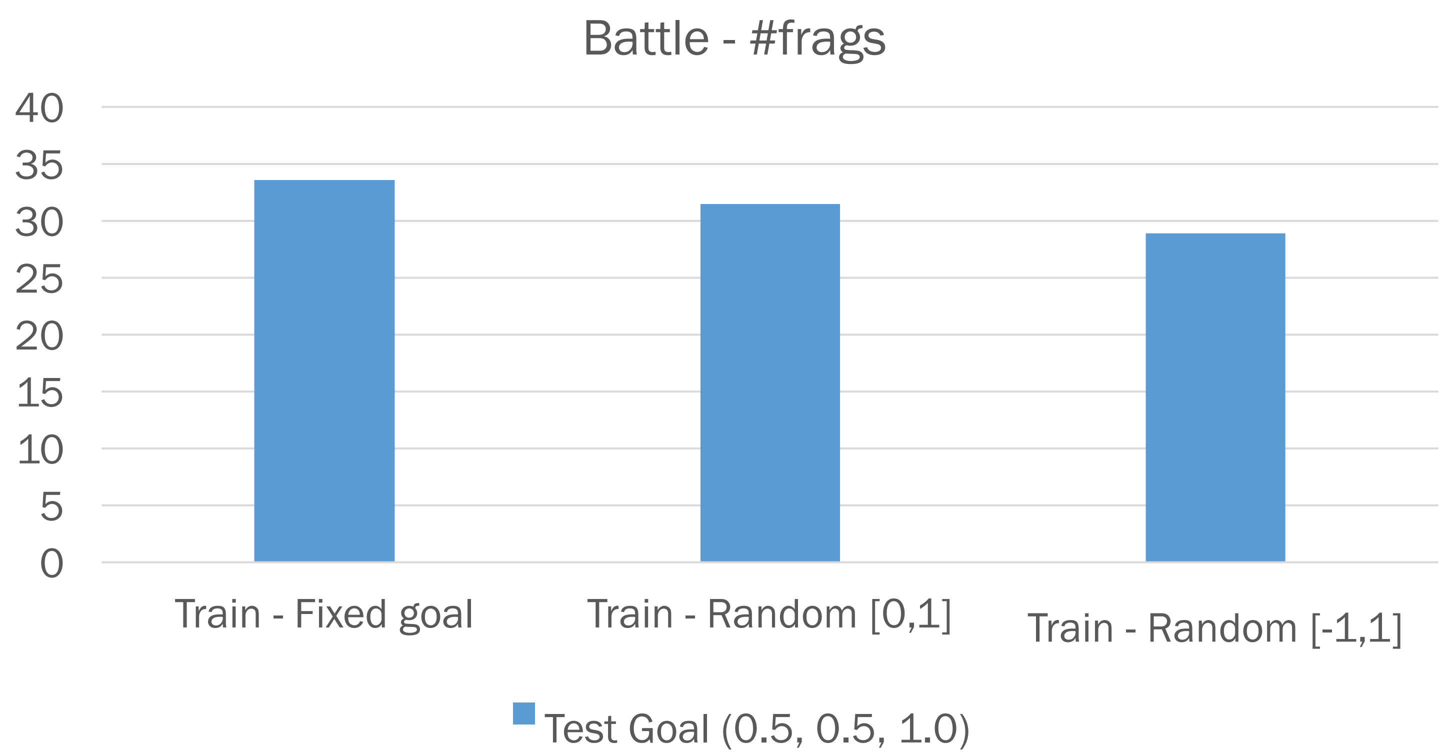
This was testing on the training set

Can we generalize?

Generalization across goals

- Train with a random goal vector in every episode
  - Uniform [0,1]
  - Uniform [-1,1]
- Change the goal vector at test time
  - The end goal does not have to be known at training time!

Generalization across goals

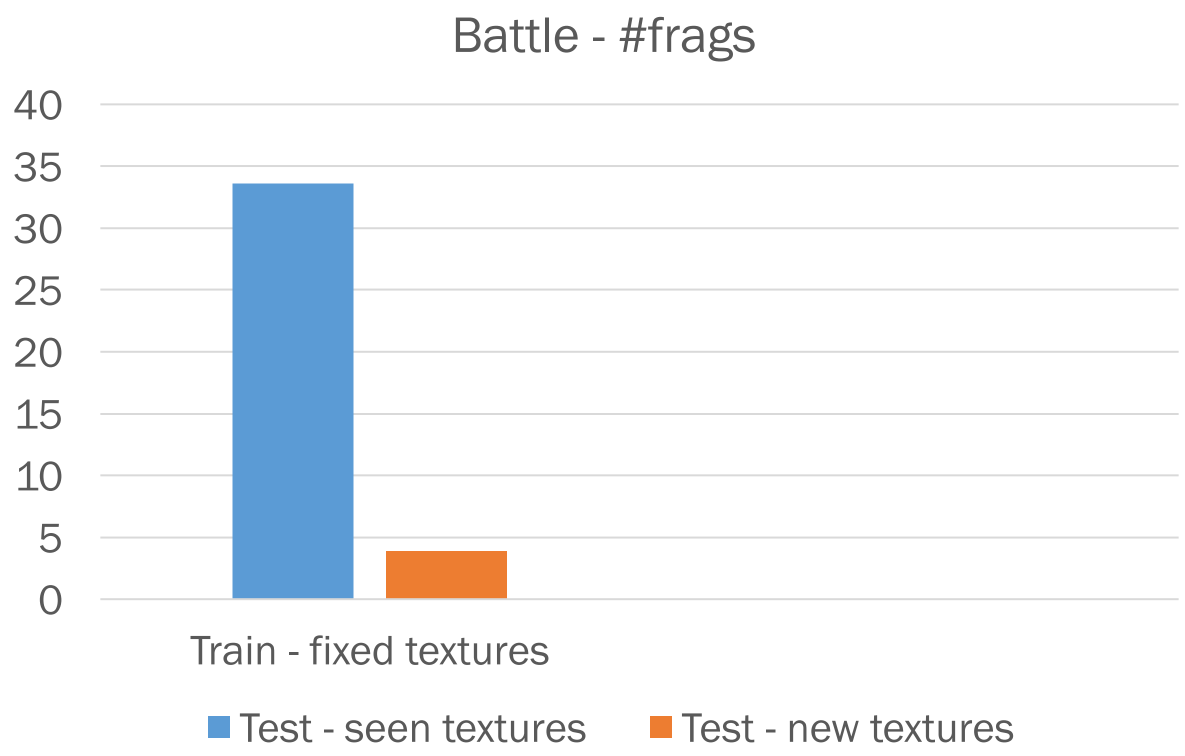


- Goal-agnostic training performs very close to training with a fixed goal
- Generalizes to different goals much better

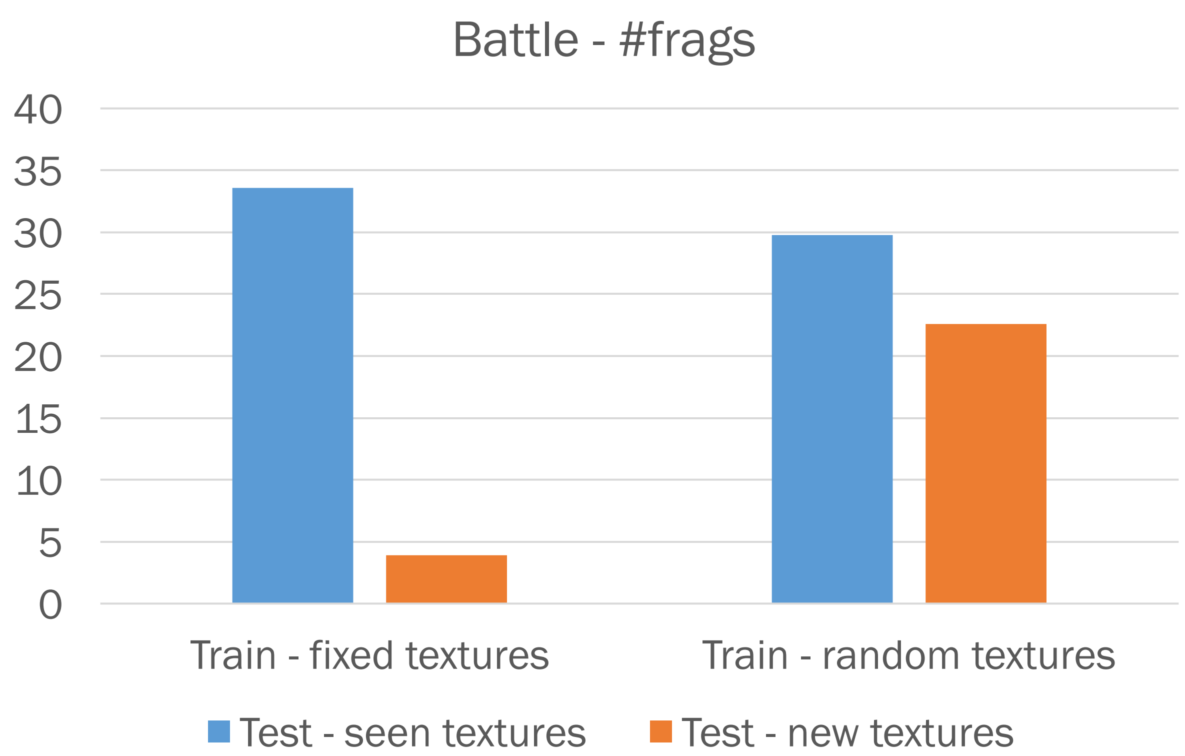
Generalization across environments

- Train with randomized textures

Generalization across environments



Generalization across environments



- Good generalization to previously unseen textures and labyrinth layouts

ViZDoom Competition: Full Deathmatch

Place	Team	1	2	3	4	5	6	7	8	9	10	11	12	Total
1	**IntelAct**	29	21	23	21	6	11	9	6	30	32	33	35	256
2	**The Terminators**	22	17	21	15	13	12	6	5	14	13	13	13	164
3	**TUHO**	8	11	13	12	0	-1	-1	-4	2	2	6	3	51
4	ColbyCS	2	4	0	1	-1	0	-1	0	3	3	4	3	18
5	5vision	3	0	4	2	1	0	1	0	0	-1	1	1	12
6	Ivomi	3	0	1	0	1	-1	-4	-4	1	1	0	0	-2
7	PotatoesArePrettyOk	0	0	2	0	-1	-3	-1	0	-2	-1	-1	-2	-9

Summary

- Simple finite-horizon supervised training performs very well on visuomotor control tasks

- Predicting measurements instead of rewards:
  - Better training signal
  - Flexible goal setting, goal-agnostic learning

- Training with random textures leads to good generalization across environments

The end

More details in our poster yesterday!

Code: https://github.com/IntelVCL/DirectFuturePrediction
Includes our environments and pre-trained models

Videos: IntelVCL youtube channel