Learning to Act by Predicting the Future

Alexey Dosovitskiy and Vladlen Koltun

Presented by Daniel Brown

Motivation

- Deep RL successes

- Key Challenges for Reinforcement Learning
  1. Learning to act in realistic domains from raw sensors
  2. Acquiring general skills for dynamic goals

- Proposal: Learn to act by predicting what will happen in the future.

Sensory input stream

Sensory input

s_t

Measurements

m_t

Predict effect of actions on future measurements

s_t

m_t

a_t

Shoot

m_{t+1}

Health = 87%

Enemies shot = 7

Ammo = 19

Predict effect of actions on future measurements

- Reduces sensorimotor control to supervised learning.

- Replaces sparse rewards with rich measurement streams.

- No need for a fixed goal during training or testing.

Specifics

Temporal offsets: $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5 = 1, 2, 4, 8, 16$

Future measurements: $f = \langle m_{t+1} - m_t, \ldots, m_{t+16} - m_t \rangle$

Goal weights: $g \in \mathbb{R}^n$

Agent Objective: $\max g^T f$

Methodology

Training is just regression:

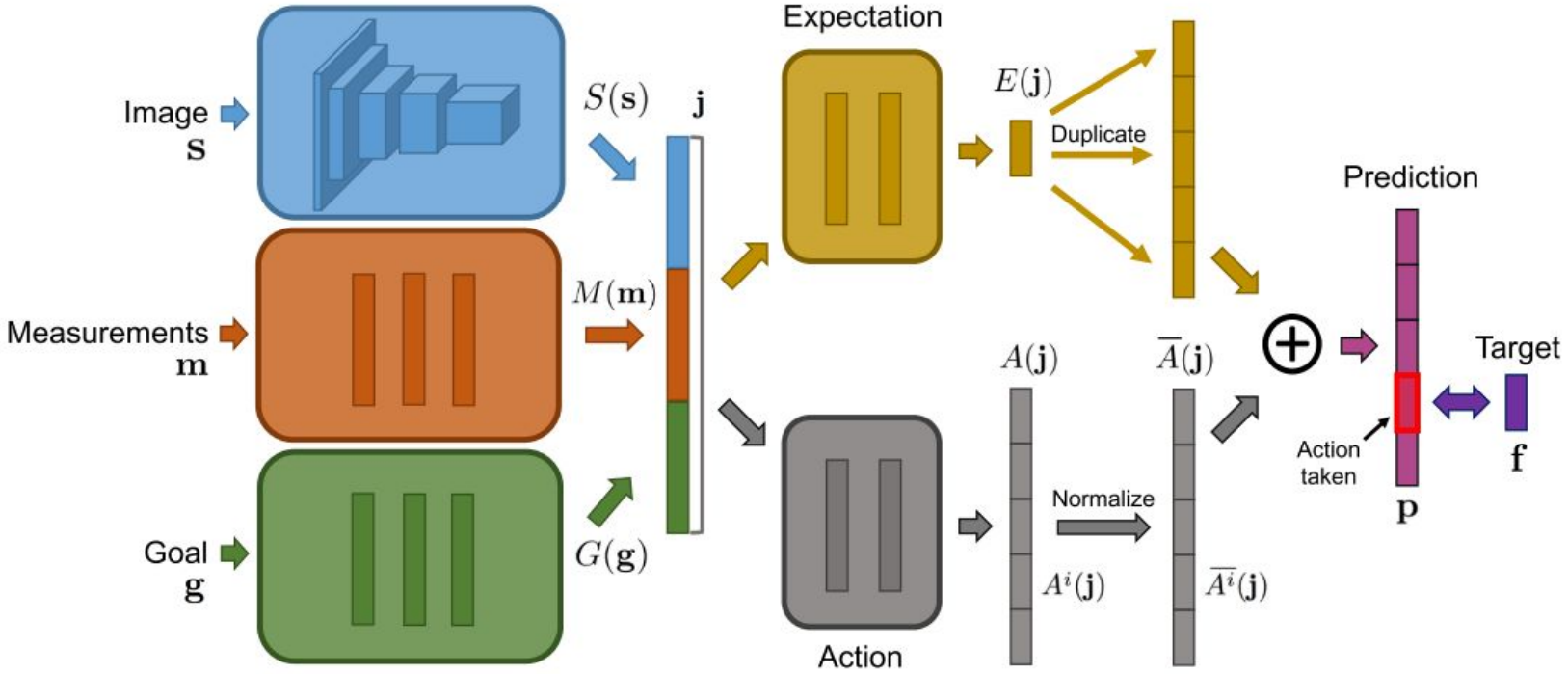
$$\mathcal{D} = \{\langle \mathbf{o}_i, a_i, \mathbf{g}_i, \mathbf{f}_i \rangle\}_{i=1}^{N}$$

$$\mathbf{p}_t^a = F(\mathbf{o}_t, a, \mathbf{g}; \boldsymbol{\theta})$$

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \| F(\mathbf{o}_i, a_i, \mathbf{g}_i; \boldsymbol{\theta}) - \mathbf{f}_i \|^2$$

At test time pick best action for current goal:

$$a_t = \arg\max_{a \in \mathcal{A}} \mathbf{g}^\top F(\mathbf{o}_t, a, \mathbf{g}; \boldsymbol{\theta})$$



$$\overline{A^i}(\mathbf{j}) = A^i(\mathbf{j}) - \frac{1}{w}\sum_{k=1}^{w} A^k(\mathbf{j})$$

Experiments

D1: Basic

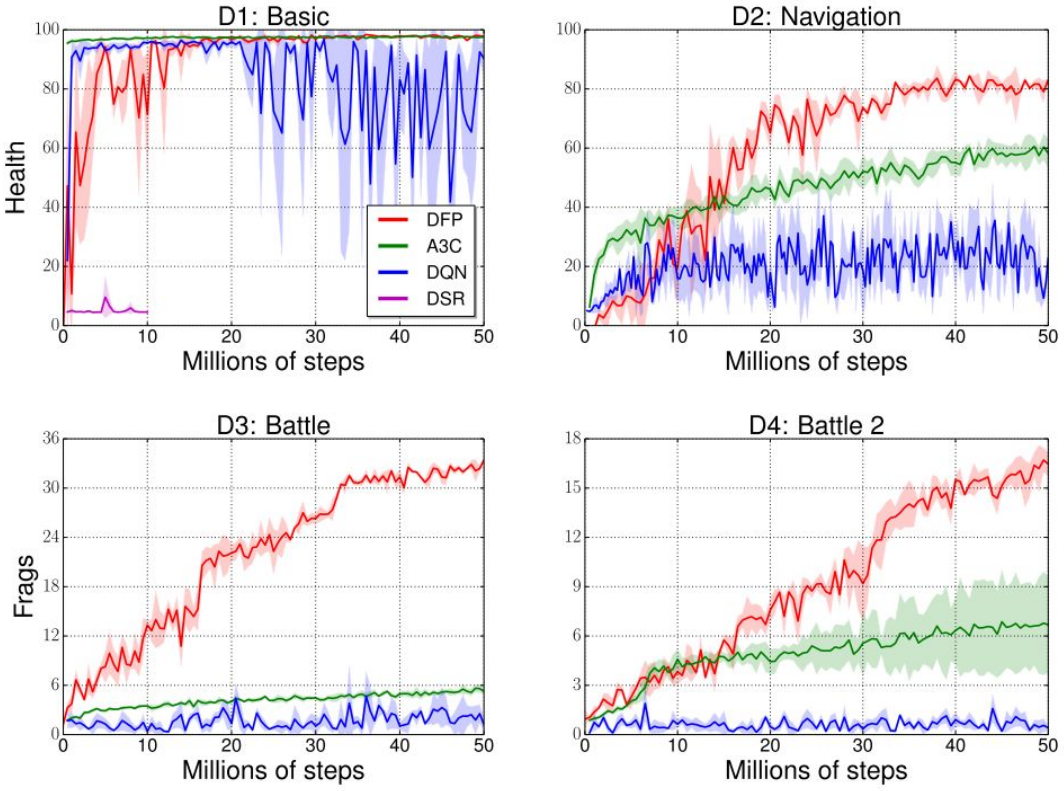
D2: Navigation

D3: Battle

D4: Battle 2

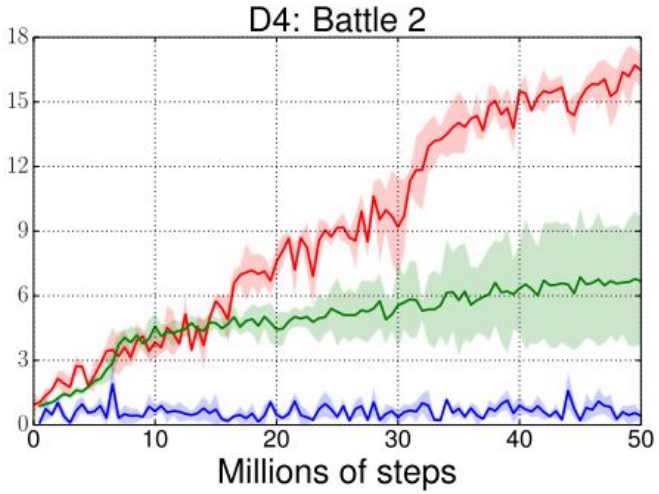
Results

- Beats standard RL
- Generalizes to new goals
- Generalizes to new envs
- Winner of 2016 VizDoom



Pros

Much better than standard RL



Visual Doom AI Competition

Full Deatchmatch

Place	Team	Bot	Total frags
1	**IntelAct**	**IntelAct**	256
2	**The Terminators**	**Arnold**	164
3	**TUHO**	**TUHO**	51
4	ColbyCS	ColbyMules	18
5	5vision	5vision	12
6	Ivomi	Ivomi	-2
7	PotatoesArePrettyOk	WallDestroyerXxx	-9

56% improvement

Didn’t test on the training data

- 100 randomly textured environments (90 train / 10 test)
- Training on harder domains generalizes to easier domains

		Train				
		D3	D4	D3-tx	D4-tx	D4-tx-L
Test	D3	33.6	17.8	29.8	20.9	22.0
	D4	1.6	17.1	5.4	10.8	12.4
	D3-tx	3.9	8.1	22.6	15.6	19.4
	D4-tx	1.7	5.1	6.2	10.2	12.7

Table 2: Generalization across environments.

Ablation study

- D3-tx environment.

D3: Battle

		frags
all measurements	all offsets	22.6
all measurements	one offset	17.2
frags only	all offsets	10.3
frags only	one offset	5.0

Table 4: Ablation study. Predicting all measurements at all temporal offsets yields the best results.

Goal generalization

	(a) fixed goal $(0.5, 0.5, 1)$			(b) random goals $[0, 1]$			(c) random goals $[-1, 1]$		
test goal	ammo	health	frags	ammo	health	frags	ammo	health	frags
$(0.5, 0.5, 1)$	83.4	97.0	33.6	92.3	96.9	31.5	49.3	94.3	28.9
$(0, 0, 1)$	0.3	-3.7	11.5	4.3	30.0	20.6	21.8	70.9	24.6
$(1, 1, -1)$	28.6	-2.0	0.0	22.1	4.4	0.2	89.4	83.6	0.0
$(-1, 0, 0)$	1.0	-8.3	1.7	1.9	-7.5	1.2	0.9	-8.6	1.7
$(0, 1, 0)$	0.7	2.7	2.6	9.0	77.8	6.6	3.0	69.6	7.9

Goal generalization

	(a) fixed goal $(0.5, 0.5, 1)$			(b) random goals $[0, 1]$			(c) random goals $[-1, 1]$		
test goal	ammo	health	frags	ammo	health	frags	ammo	health	frags
$(0.5, 0.5, 1)$	83.4	97.0	33.6	92.3	96.9	31.5	49.3	94.3	28.9
$(0, 0, 1)$	0.3	-3.7	11.5	4.3	30.0	20.6	21.8	70.9	24.6
$(1, 1, -1)$	28.6	-2.0	0.0	22.1	4.4	0.2	89.4	83.6	0.0
$(-1, 0, 0)$	1.0	-8.3	1.7	1.9	-7.5	1.2	0.9	-8.6	1.7
$(0, 1, 0)$	0.7	2.7	2.6	9.0	77.8	6.6	3.0	69.6	7.9

Goal generalization

Training Method

test goal	(a) fixed goal $(0.5, 0.5, 1)$			(b) random goals $[0, 1]$			(c) random goals $[-1, 1]$		
	ammo	health	frags	ammo	health	frags	ammo	health	frags
$(0.5, 0.5, 1)$	83.4	97.0	33.6	92.3	96.9	31.5	49.3	94.3	28.9
$(0, 0, 1)$	0.3	-3.7	11.5	4.3	30.0	20.6	21.8	70.9	24.6
$(1, 1, -1)$	28.6	-2.0	0.0	22.1	4.4	0.2	89.4	83.6	0.0
$(-1, 0, 0)$	1.0	-8.3	1.7	1.9	-7.5	1.2	0.9	-8.6	1.7
$(0, 1, 0)$	0.7	2.7	2.6	9.0	77.8	6.6	3.0	69.6	7.9

Goal generalization

	(a) fixed goal $(0.5, 0.5, 1)$			(b) random goals $[0, 1]$			(c) random goals $[-1, 1]$		
test goal	ammo	health	frags	ammo	health	frags	ammo	health	frags
$(0.5, 0.5, 1)$	83.4	97.0	33.6	92.3	96.9	31.5	49.3	94.3	28.9
$(0, 0, 1)$	0.3	-3.7	11.5	4.3	30.0	20.6	21.8	70.9	24.6
$(1, 1, -1)$	28.6	-2.0	0.0	22.1	4.4	0.2	89.4	83.6	0.0
$(-1, 0, 0)$	1.0	-8.3	1.7	1.9	-7.5	1.2	0.9	-8.6	1.7
$(0, 1, 0)$	0.7	2.7	2.6	9.0	77.8	6.6	3.0	69.6	7.9