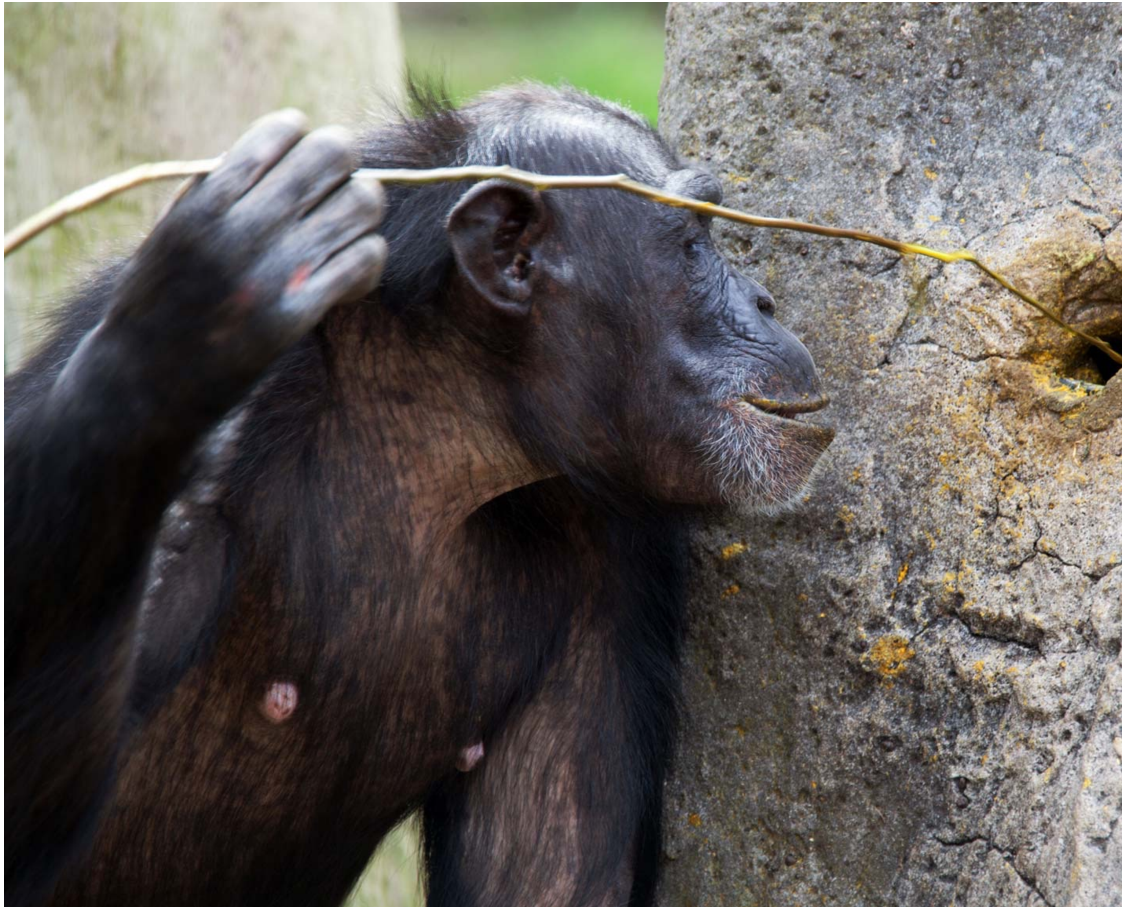


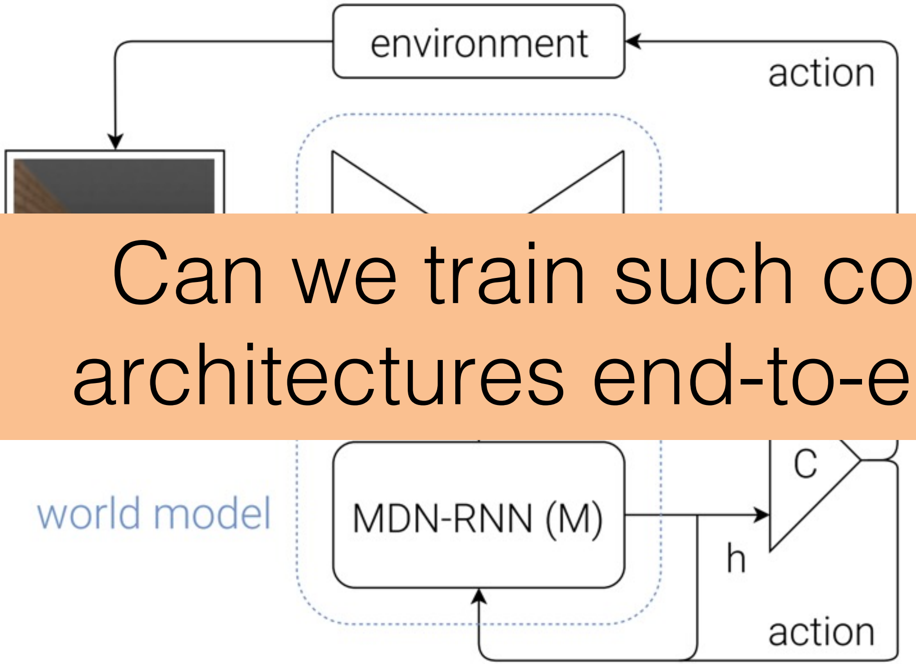

Evolving Agents that Learn More Like Animals

Sebastian Risi. Professor @ IT University of Copenhagen. Co-founder @ modl.ai

World Models

(Ha and Schmidhuber, NeurIPS 2018)



1. Collect 10,000 rollouts from a random policy.
2. Train VAE (V) to encode frames into $z \in \mathcal{R}^{32}$.

Can we train such complex heterogenous architectures end-to-end through evolution?

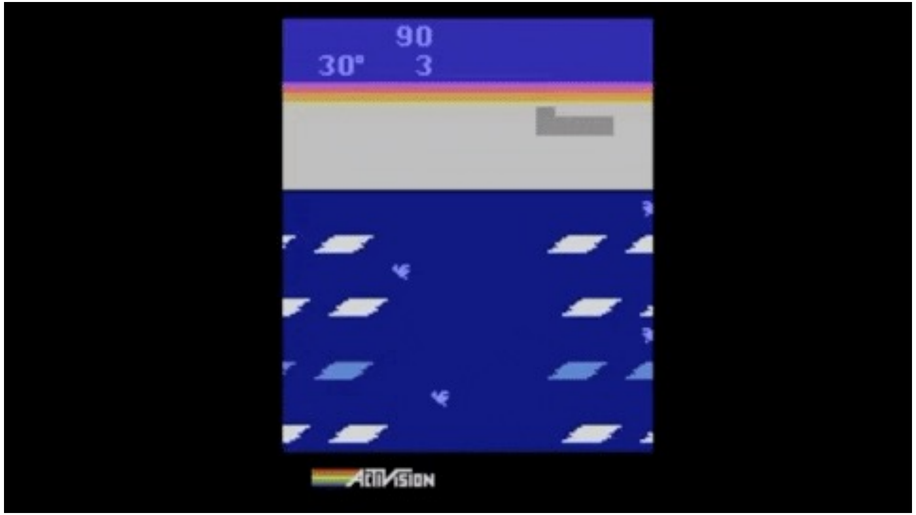
- Relies on different training methods
- Random rollouts might not be enough in complex emvironments

Evolution for the Win

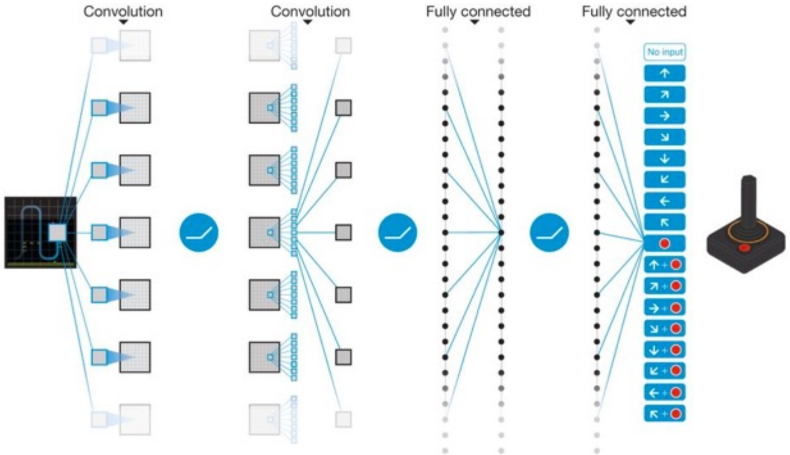
- Salimans et al. ”Evolution Strategies as a Scalable Alternative to Reinforcement Learning”, 2017.

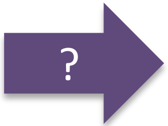
- Such et al. ”Deep Neuroevolution: Genetic Algorithms Are a Competitive Alternative for Training Deep Neural Networks for Reinforcement Learning”, 2018.

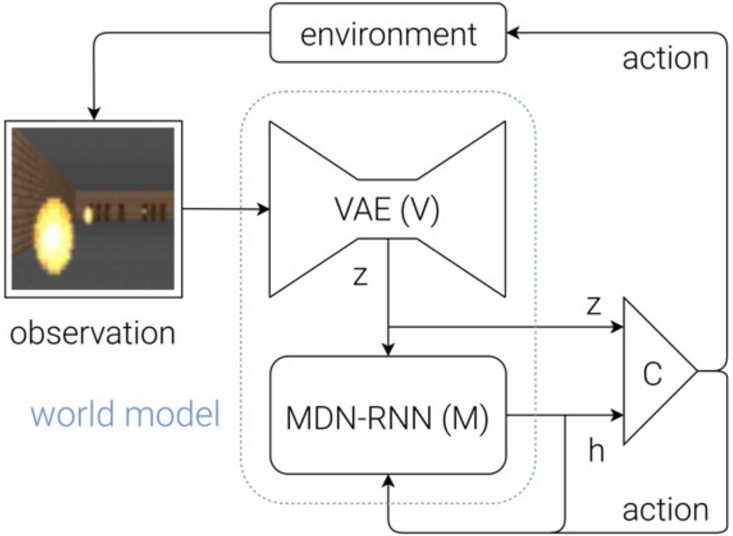
	DQN	ES	A3C	RS	GA	GA
Frames	200M	1B	1B	1B	1B	6B
Time	~7-10d	~ 1h	~ 4d	~ 1h or 4h	~ 1h or 4h	~ 6h or 24h
Forward Passes	450M	250M	250M	250M	250M	1.5B
Backward Passes	400M	0	250M	0	0	0
Operations	1.25B U	250M U	1B U	250M U	250M U	1.5B U
amidar	**978**	112	264	143	263	377
assault	4,280	1,674	**5,475**	649	714	814
asterix	4,359	1,440	**22,140**	1,197	1,850	2,255
asteroids	1,365	1,562	**4,475**	1,307	1,661	2,700
atlantis	279,987	**1,267,410**	911,091	26,371	76,273	129,167
enduro	**729**	95	-82	36	60	80
frostbite	797	370	191	1,164	**4,536**	**6,220**
gravitar	473	**805**	304	431	476	764
kangaroo	7,259	**11,200**	94	1,099	3,790	**11,254**
seaquest	**5,861**	1,390	2,355	503	798	850
skiing	-13,062	-15,443	-10,911	-7,679	†**-6,502**	†**-5,541**
venture	163	760	23	488	**969**	†**1,422**
zaxxon	5,363	6,380	**24,622**	2,538	6,180	7,864



Can GAs scale to complex architectures with multiple different and interacting components?



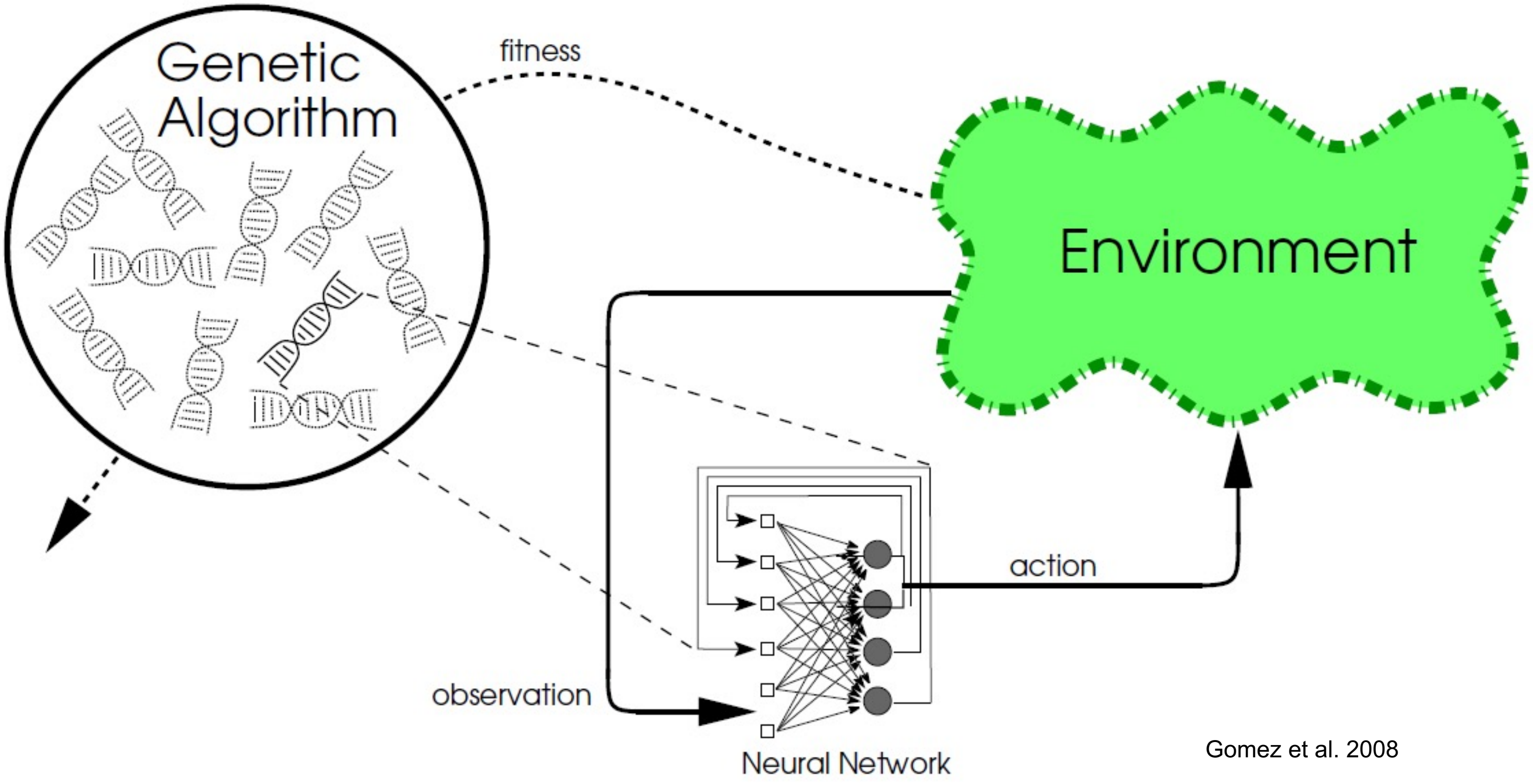




Feed-forward architecture

Complex heterogenous architecture

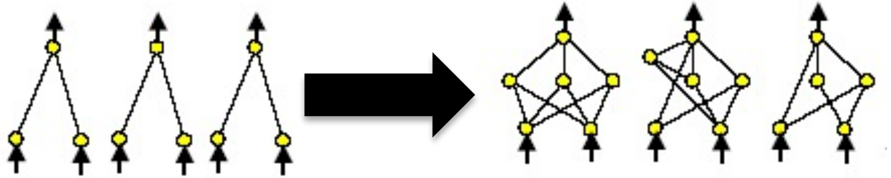
Neuroevolution



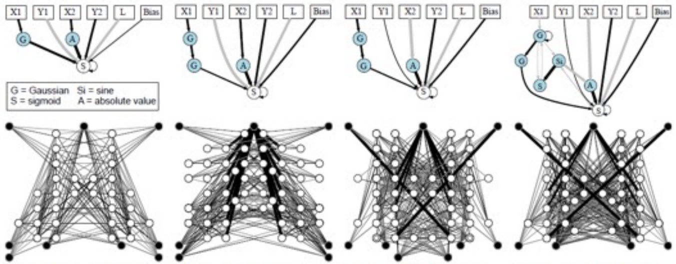
Gomez et al. 2008

Neuroevolution

Direct encodings
(e.g. NEAT)

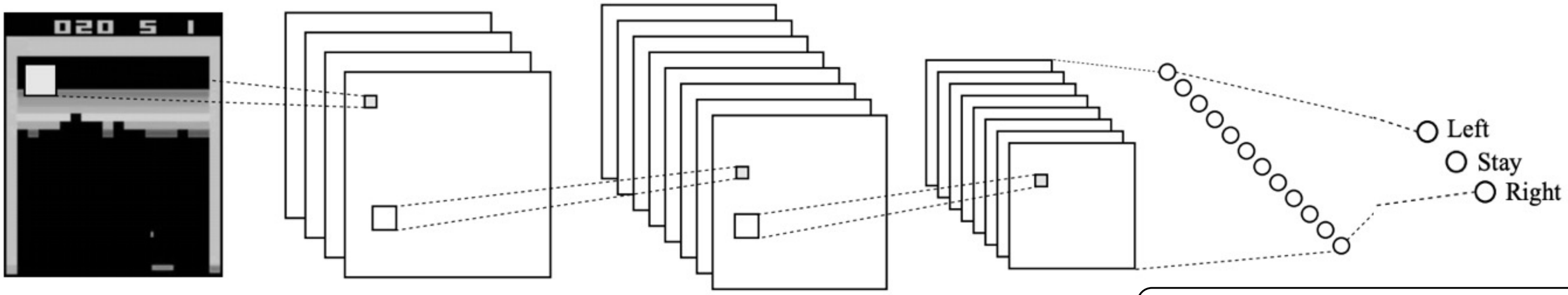


Indirect encodings
(e.g. ES-HyperNEAT)



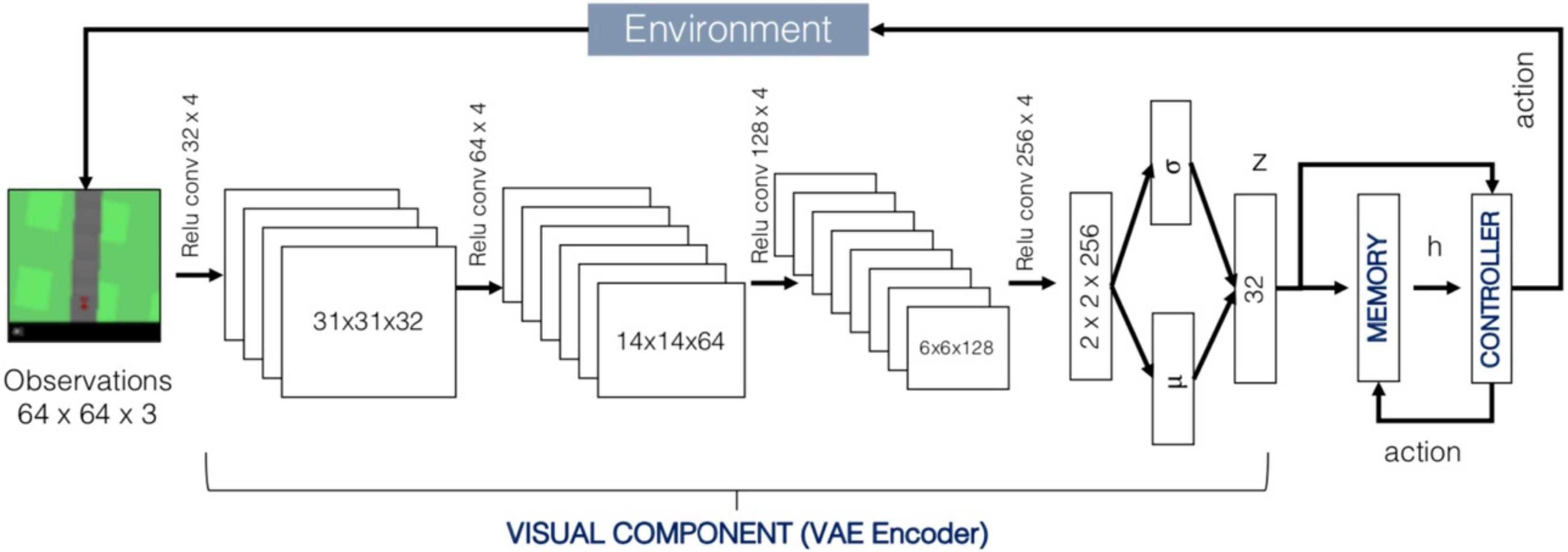
(a) Gen 24. ANN: 30 n, 184 c, CPPN: 2 n, 9 c, f=0.85) (b) Gen 30 (ANN: 52 n, 280 c, CPPN: 3 n, 10 c, f=0.93) (c) Gen 106 (ANN: 42 n, 310 c, CPPN: 3 n, 10 c, f=5.96) (d) Gen 237 (ANN: 40 n, 356 c, CPPN: 5 n, 18 c, f=10.00)

Deep Neuroevolution



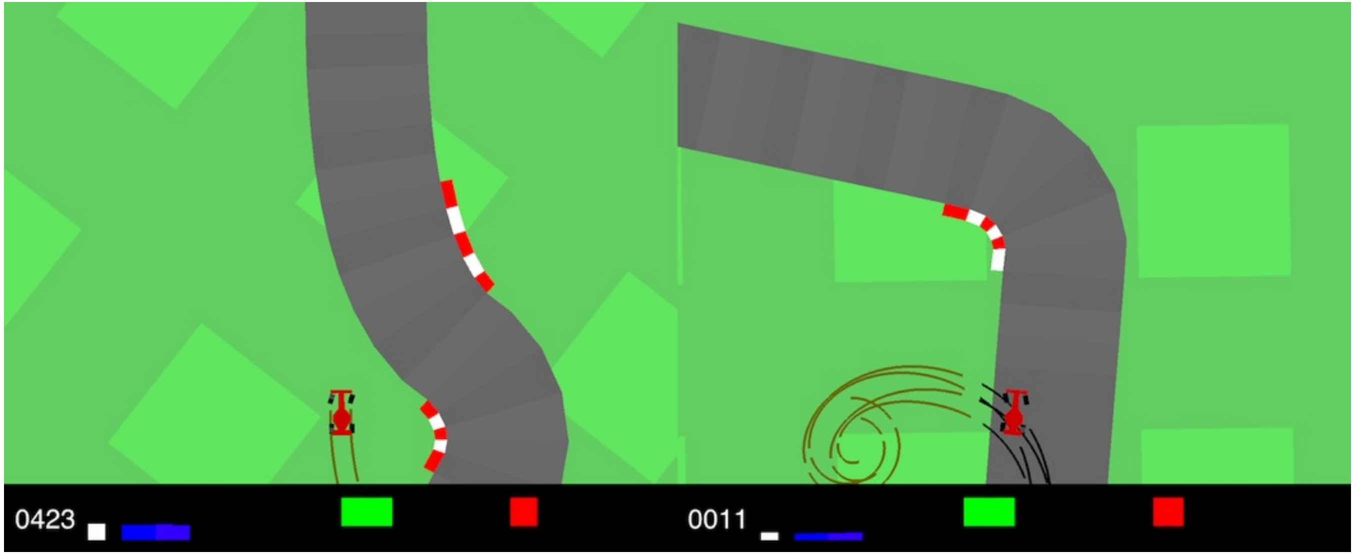
Mutations: $\theta' = \theta + \sigma\epsilon$

Deep Neuroevolution of Recurrent and Discrete World Models (Risi & Stanley, GECCO 2019)



Mutations: $\theta' = \theta + \sigma\epsilon$

Model	#Params	WM Training [13]	GA Training
VAE	4,348,547	SGD - 1 epoch	Pop size 200 Rollouts 1 Solved: 1,200
MD-RNN	384,071	SGD - 20 epochs	
Controller	867	CMA-ES - Pop 64 Rollouts 16 Solved: 1,800 Gen.	

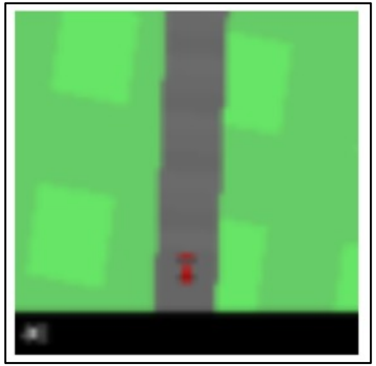


METHOD	AVG. SCORE
DQN (PRIEUR, 2017)	343 ± 18
A3C (CONTINUOUS) (JANG ET AL., 2017)	591 ± 45
A3C (DISCRETE) (KHAN & ELIBOL, 2016)	652 ± 10
CEOBILLIONAIRE (GYM LEADERBOARD)	838 ± 11

Methods require pre-processing such as edge detection / stacking recent frames

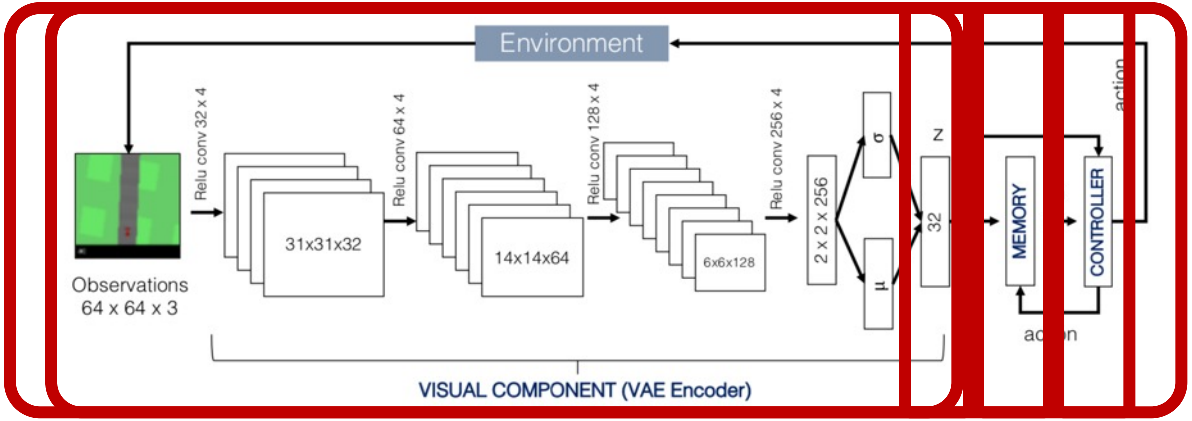
Experimental Setup

- `CarRacing-v0` Solution: Average score > 900 on 100 randomly generated tracks
- 3 network outputs: left/right steering, acceleration and braking
- Reward of -0.1 every frame, reward of +100/N for each visited track tile
- GA with truncation selection (no crossover), population size 200
- Top 3 elites are evaluated 20 times

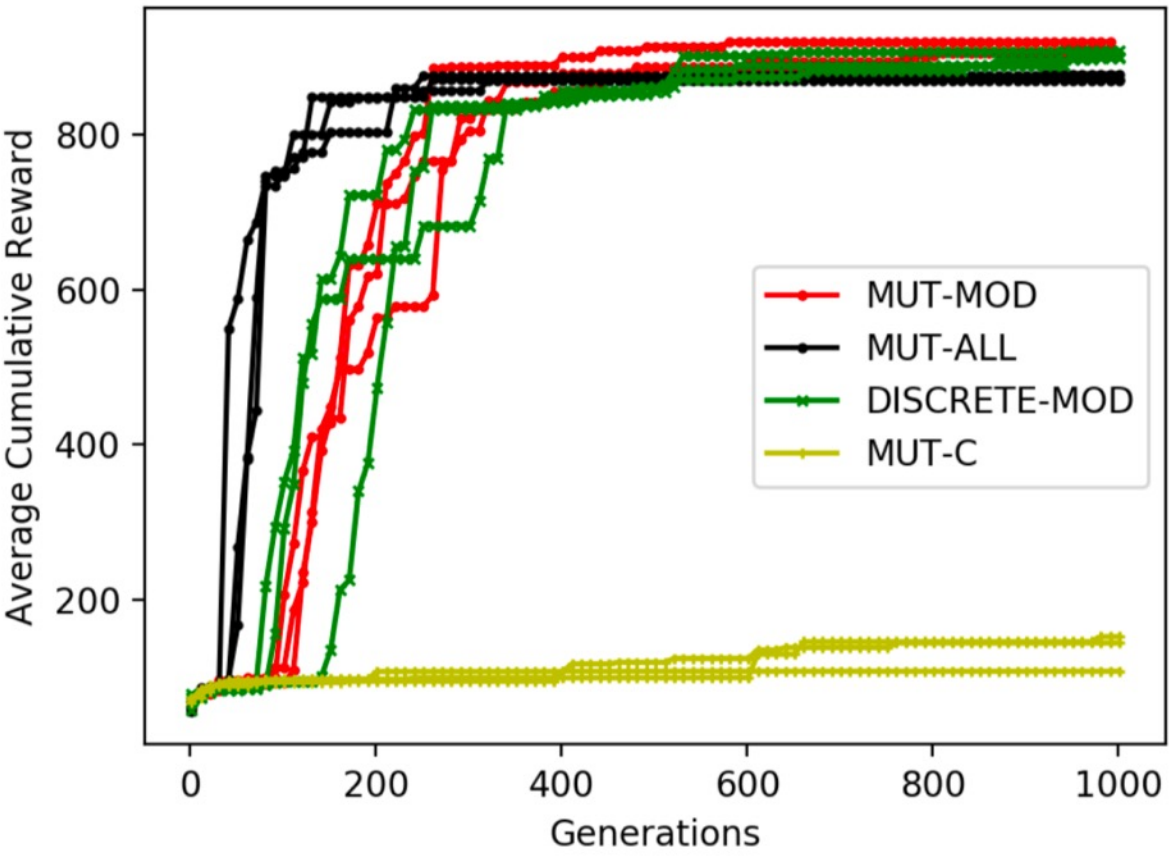


Four experimental setups:

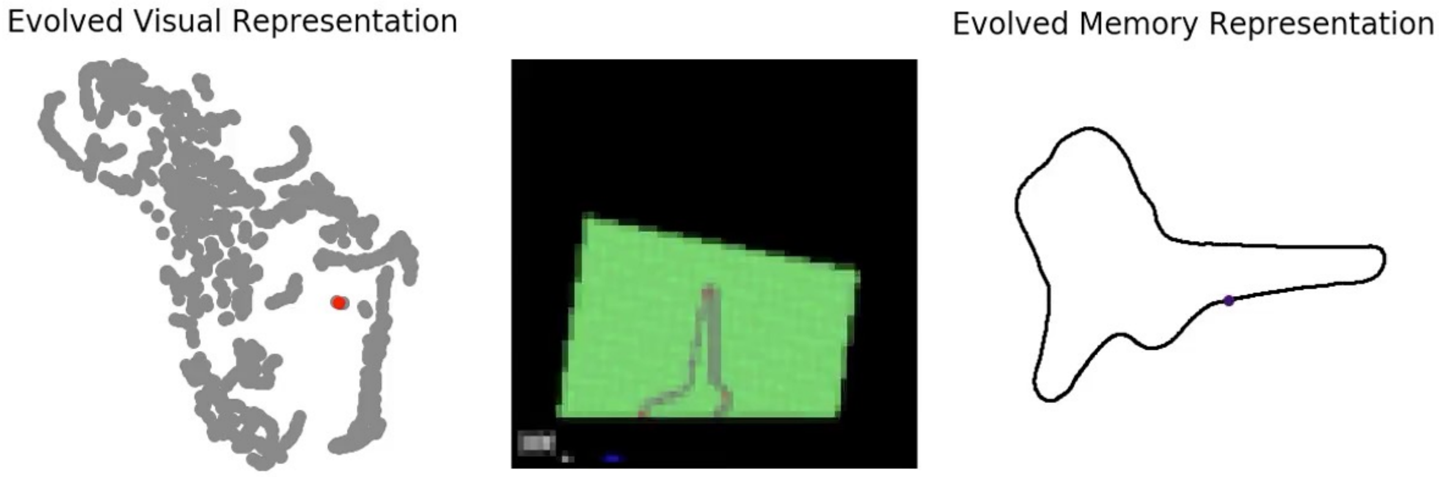
- MUT-ALL
- MUT-MOD
- MUT-C
- DISCRETE-MOD

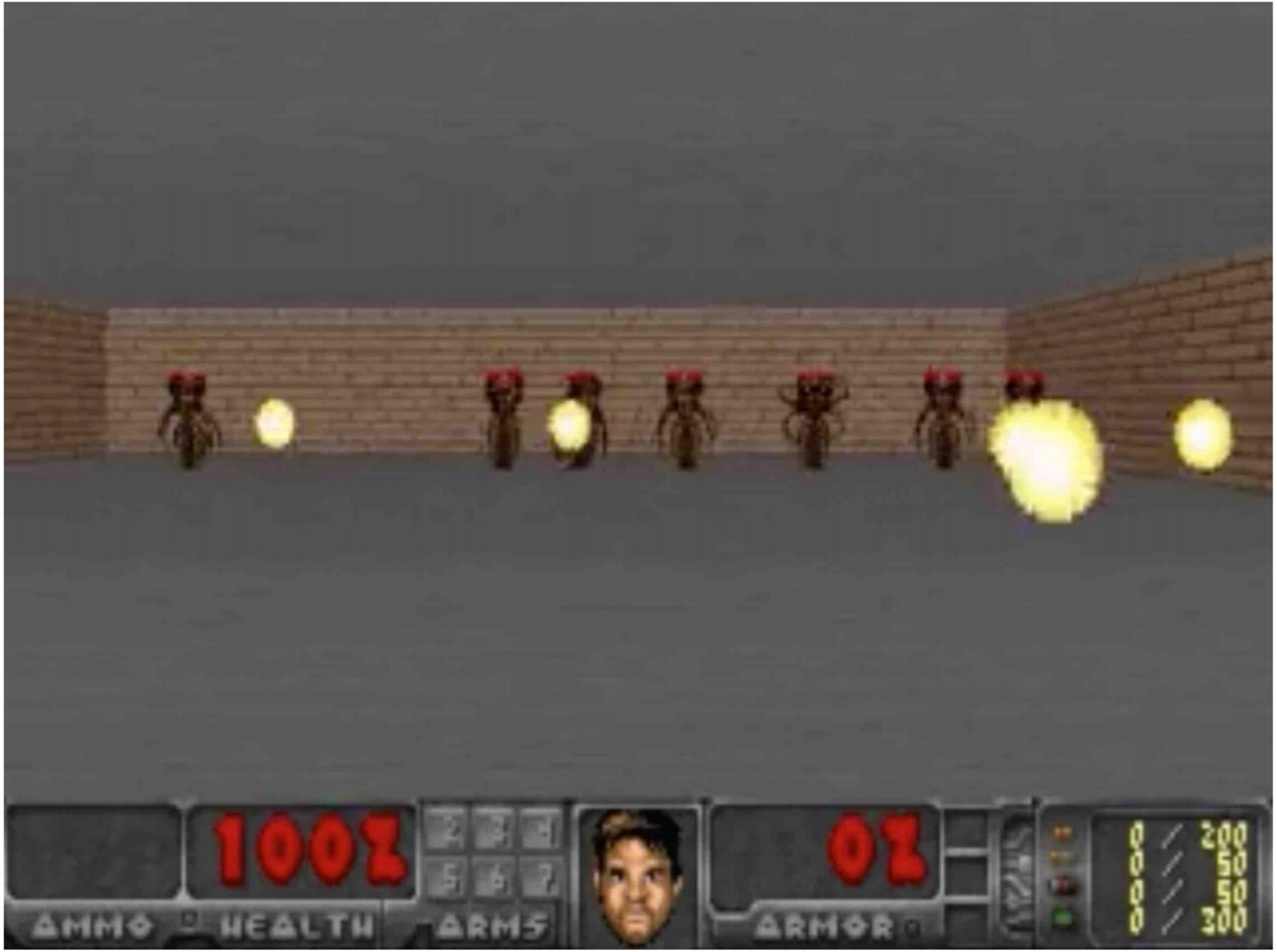


Training Results



Method	Average Score
DQN [29]	343 ± 18
DQN + Dropout [8]	893 ± 41
A3C (Continious) [16]	591± 45
A3C (Discrete) [19]	652 ± 10
CEOBILLIONAIRE (Gym leaderboard)	838 ± 11
World model [13]	**906** ± 21
World model with random MDN-RNN [43]	870 ± 120
GA (ours)	**903** ± 72

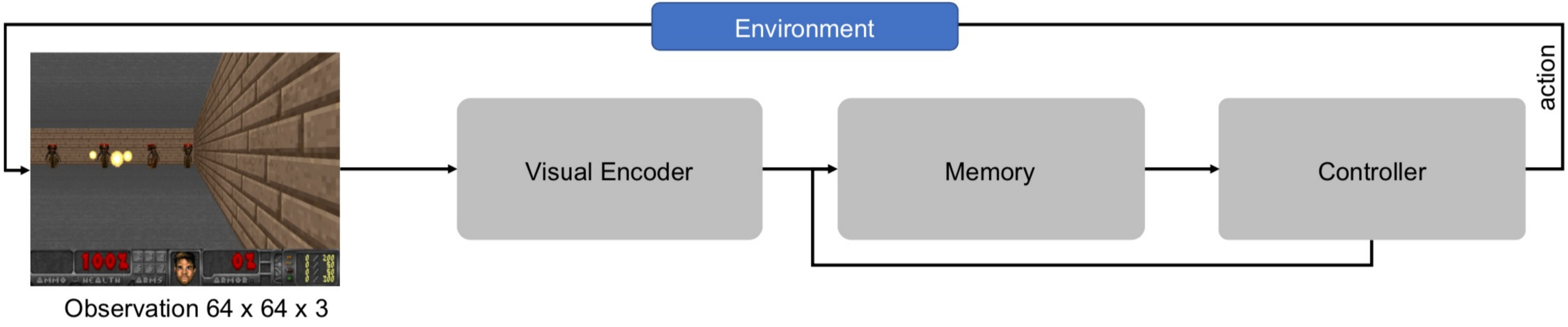



100%
AMMO
HEALTH
ARMS
0%
ARMOR
0 / 200
0 / 50
0 / 50
0 / 300

Novel approach: **Deep Innovation Protection**

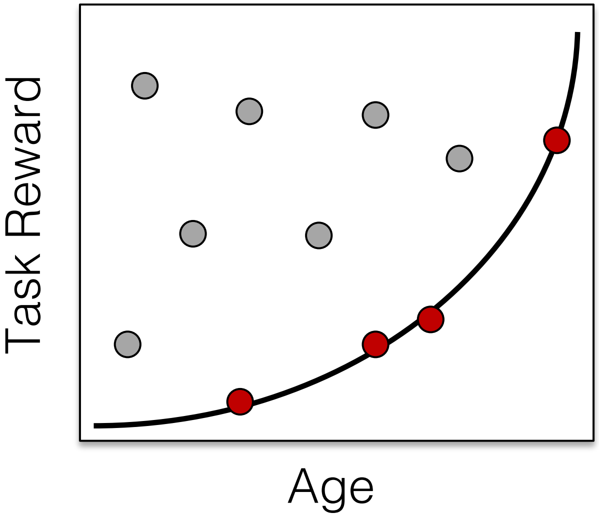
(Risi & Stanley, AAAI 2021)

- Use multi-objective optimization NSGA-II with additional ”age” objective
- Age is reset to zero when either the VAE or MDN-RNN is changed (lower age is better)



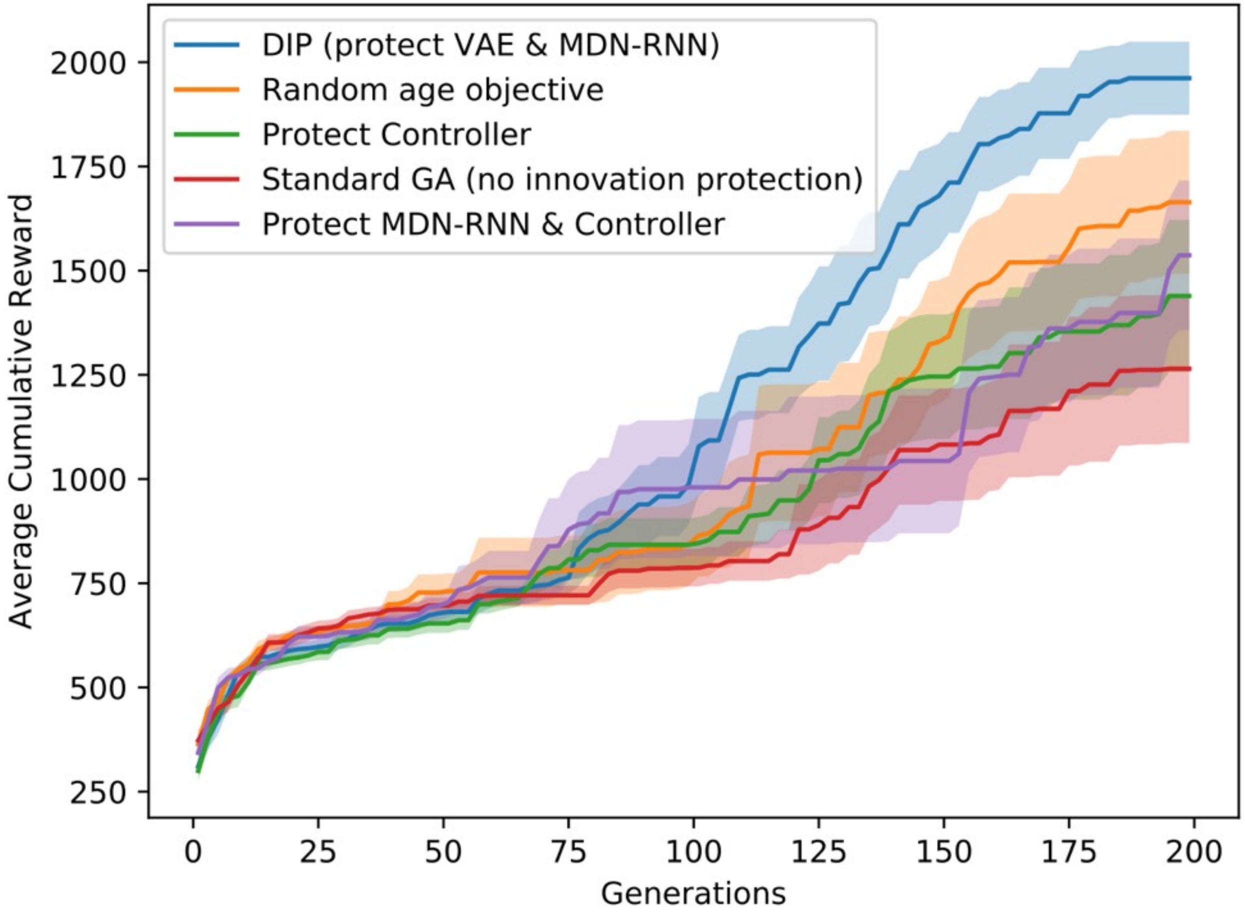
Optimization Details

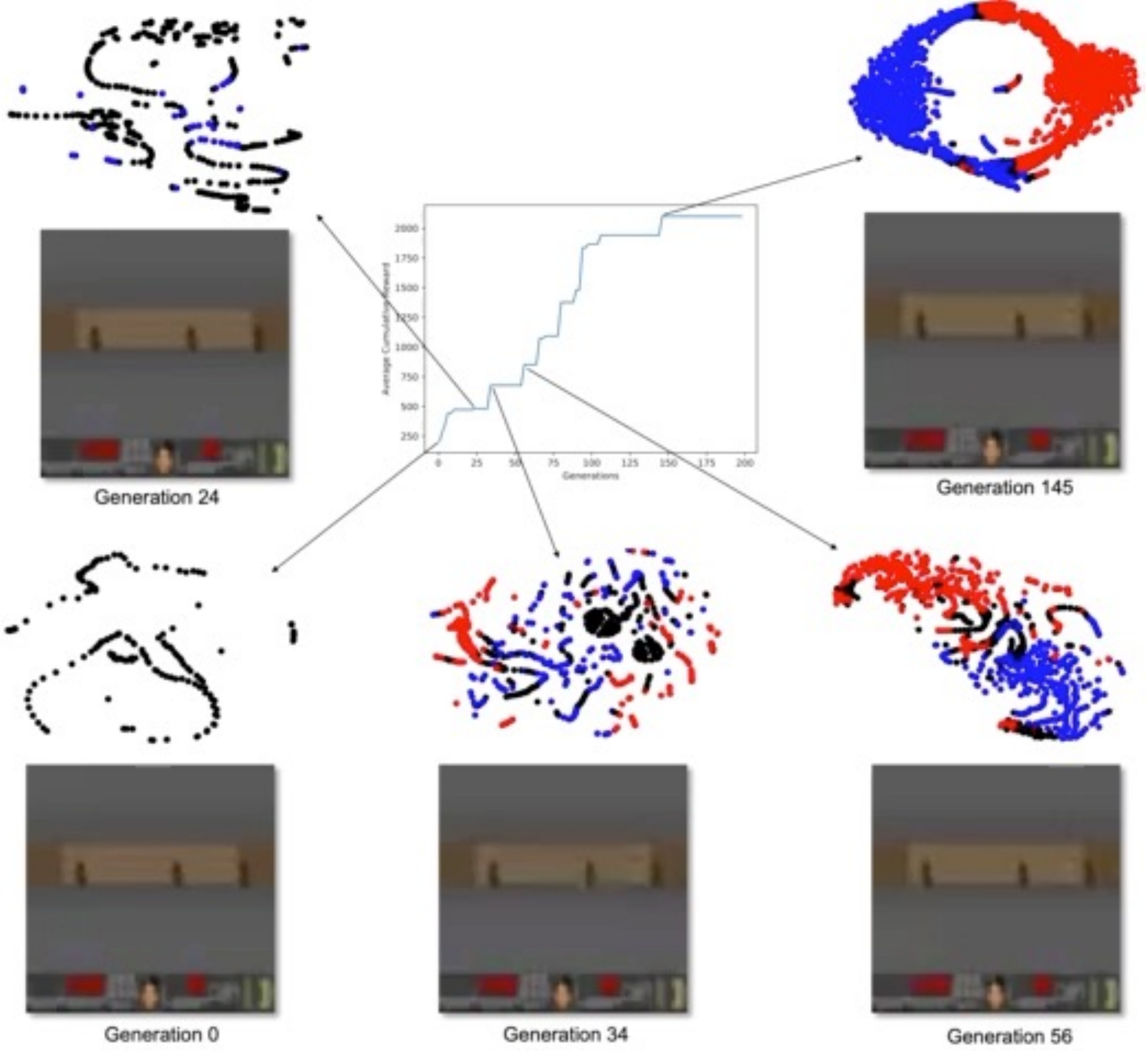
Pareto Front



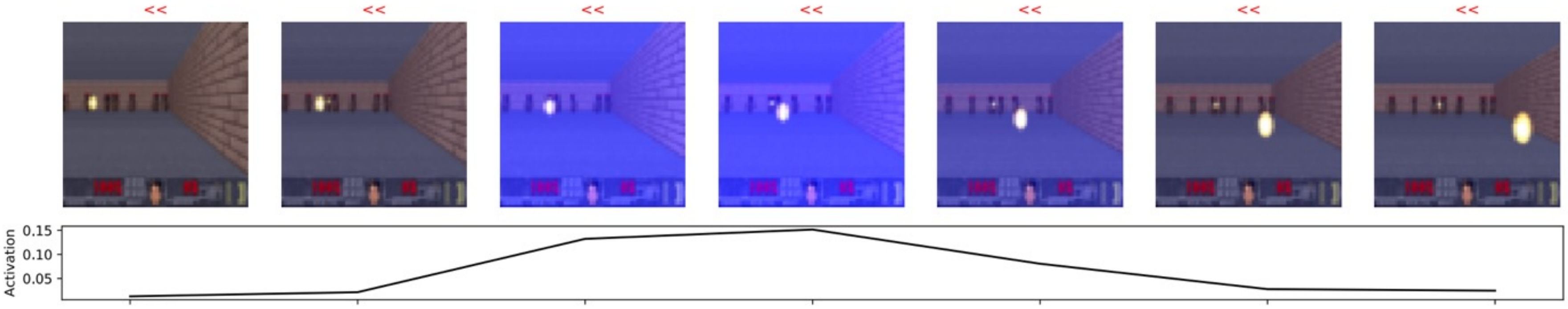
Algorithm 1 Deep Innovation Protection

Generate random population of size N with age objectives set to 0
for $generation = 1$ to i **do**
    for Individual in Population **do**
        Objective[1] = age
        Objective[2] = accumulated task reward
        Increase individual's age by 1
    end for
    Assign ranks based on Pareto fronts
    Generate set of non-dominated solutions
    Add solutions, starting from first front, until number solution = N
    Generate child population through binary tournament selection and mutations
    Reset age to 0 for all individuals whose VC or MDN-RNN was mutated
end for

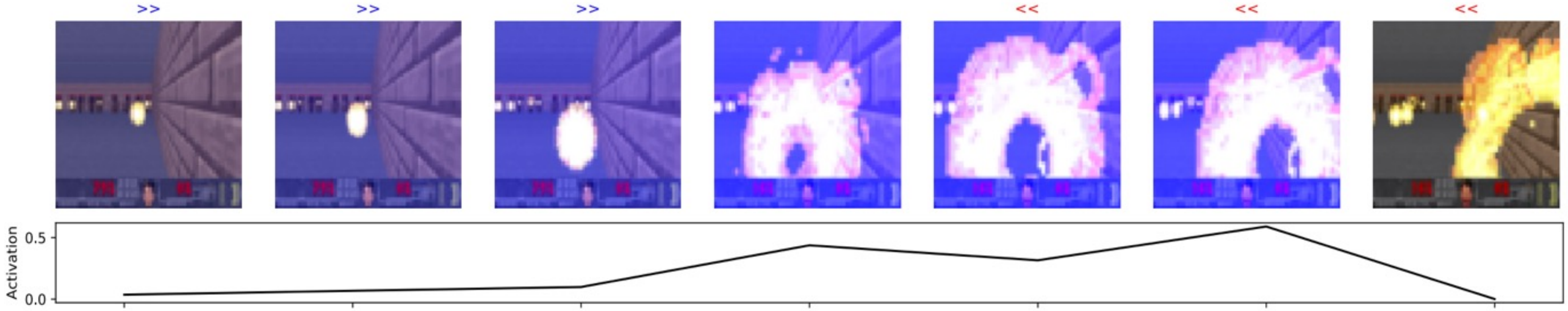

DIP (protect VAE & MDN-RNN)
Random age objective
Protect Controller
Standard GA (no innovation protection)
Protect MDN-RNN & Controller
Average Cumulative Reward
Generations
250
500
750
1000
1250
1500
1750
2000
0
25
50
75
100
125
150
175
200



Average Cumulative Reward
Generations
Generation 24
Generation 145
Generation 0
Generation 34
Generation 56



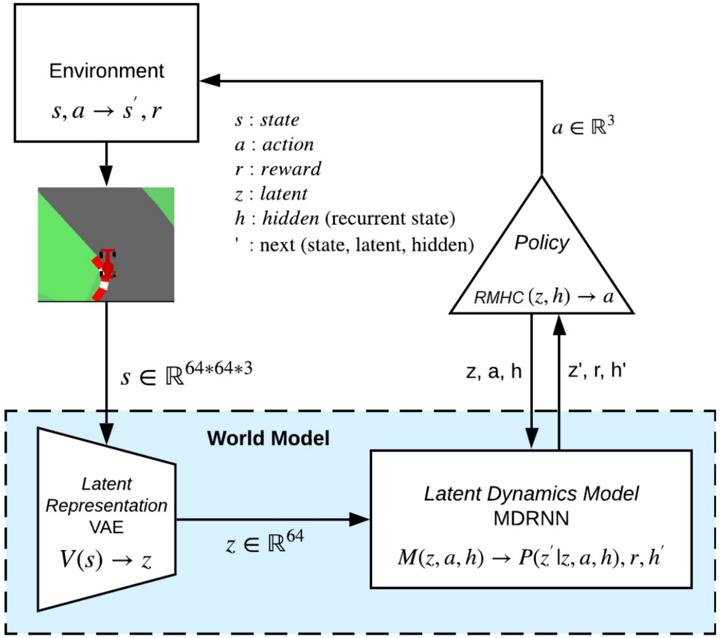
(a)

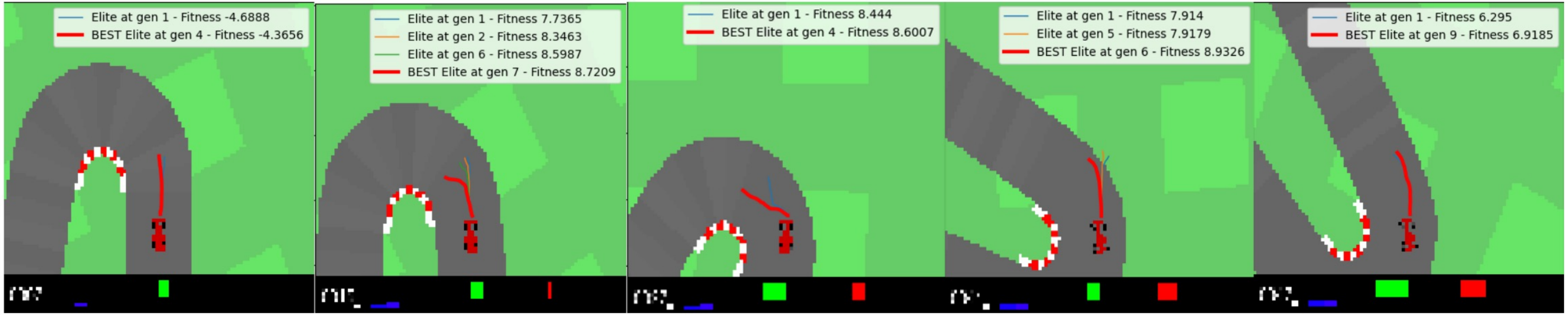


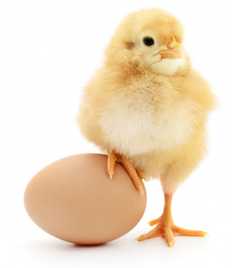
(b)

Evolutionary Planning in Latent Space

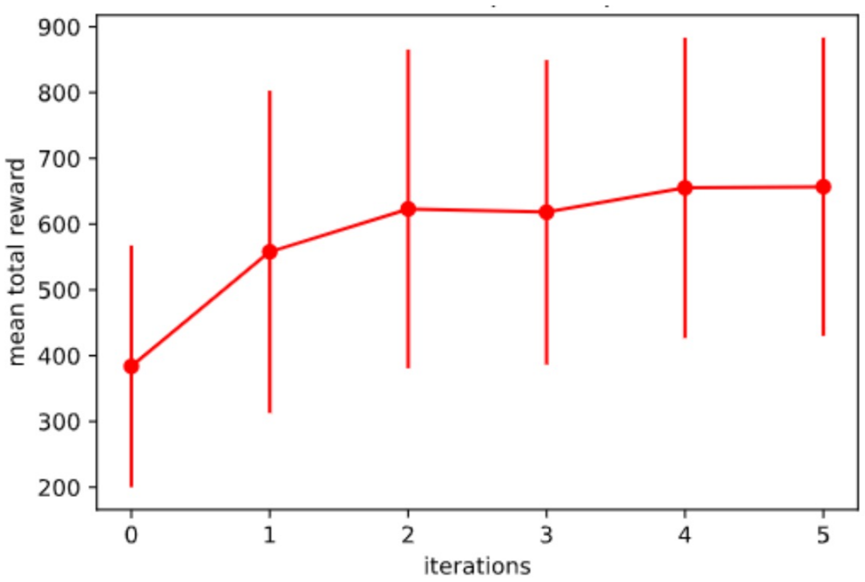
(Olesen, Nguyen, Palm, Risi; EvoApps 2020)

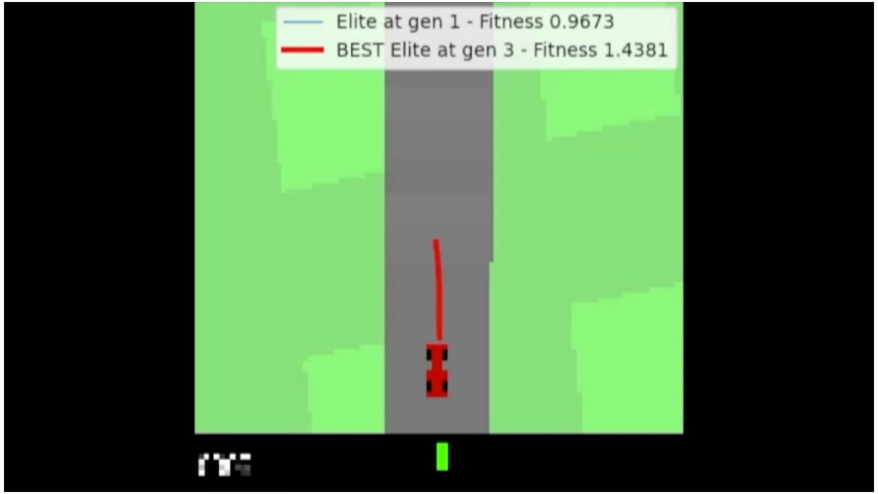






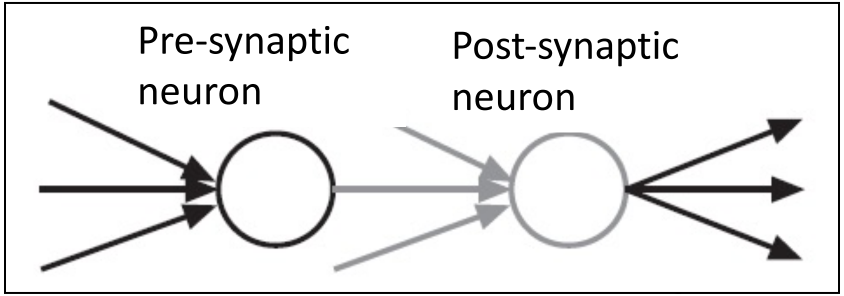
The world model can be iteratively refined



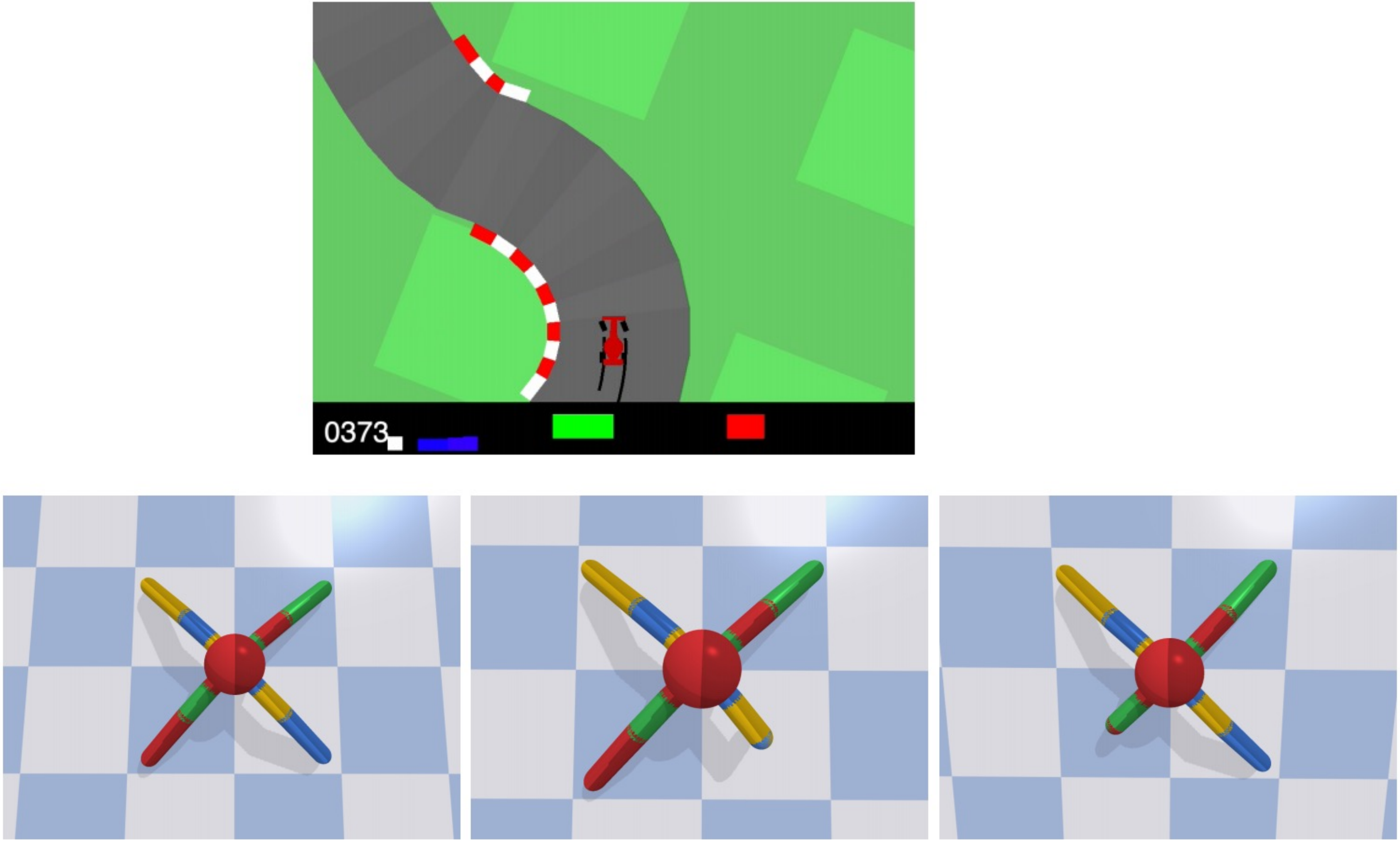


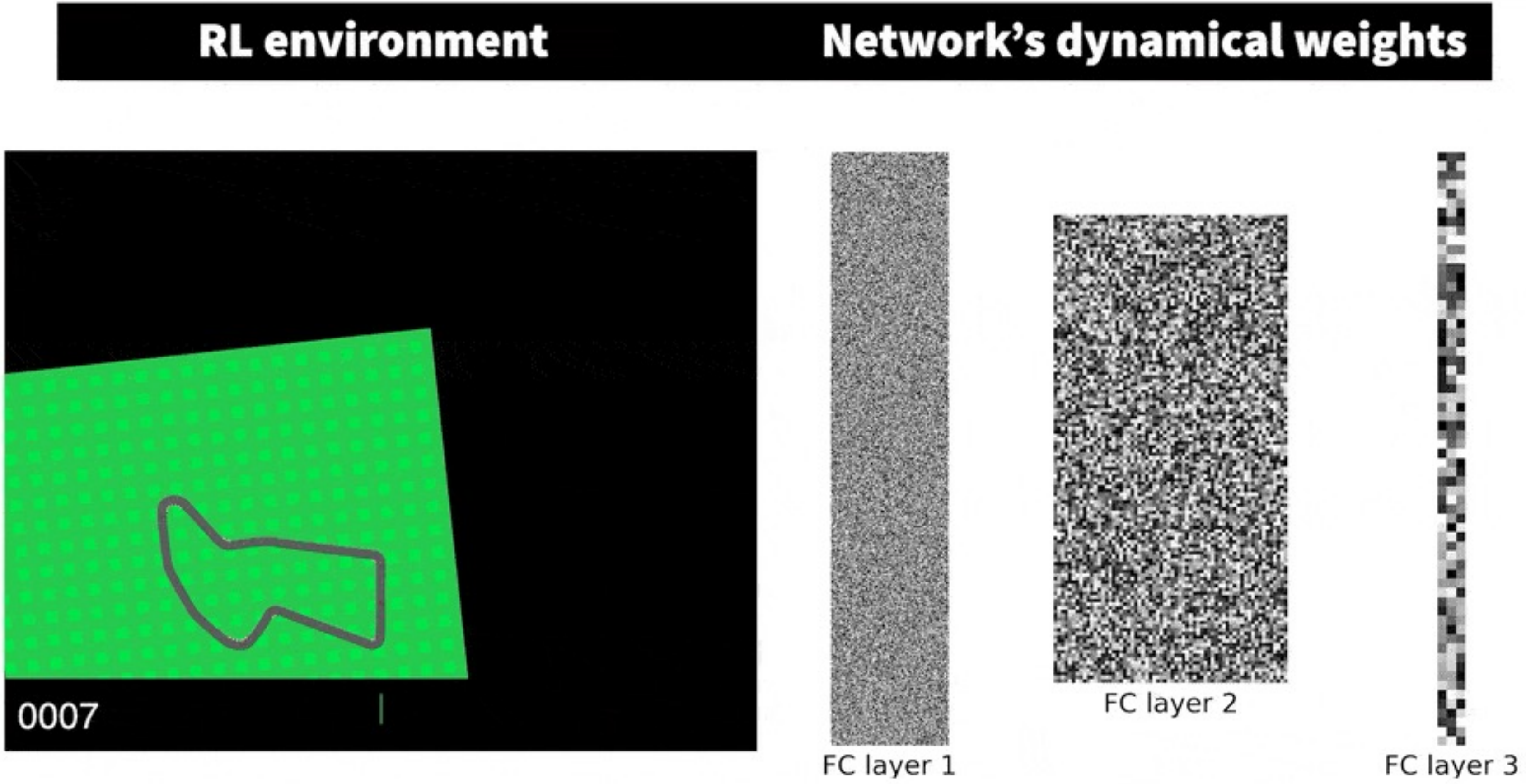
Meta-Learning through Hebbian Plasticity in Random Networks (Najarro & Risi, NeurIPS 2020)

- Networks trained through RL or evolution often static during lifetime ➔ limits their adaptation capabilities
- Start network with random weights instead and only evolve Hebbian learning rules

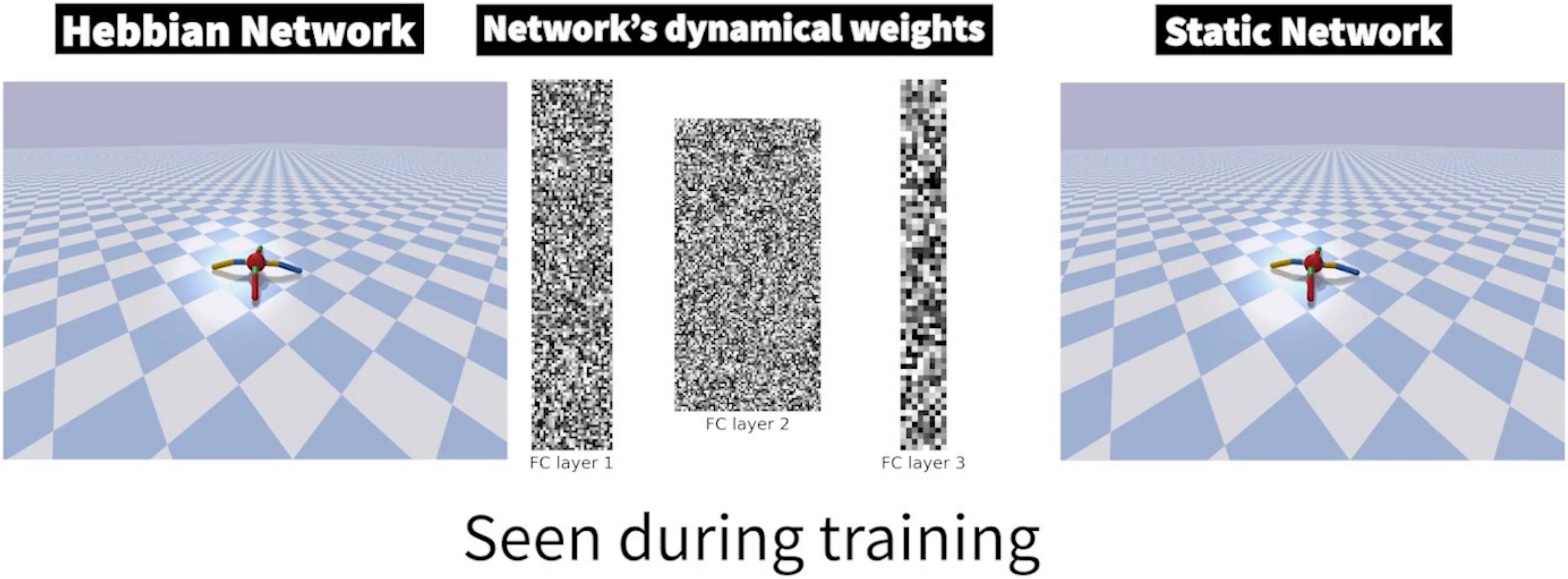


$$\Delta w_{ij} = \eta_w \cdot (A_w o_i o_j + B_w o_i + C_w o_j + D_w)$$


0373

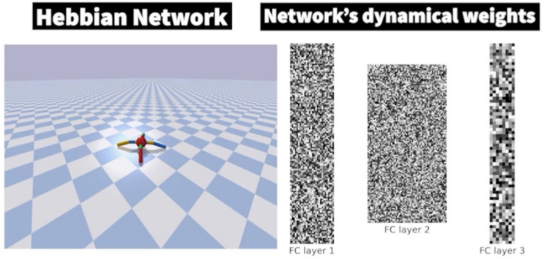


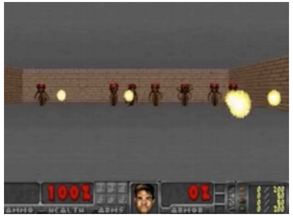
450K trainable parameters

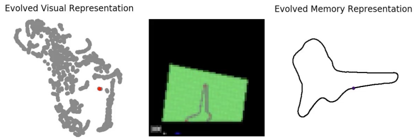


3-layer feedforward network with [128, 64, 8] nodes ➔ 61,440 Hebbian coefficients

Wrapping up



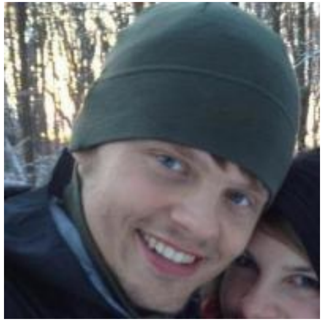




- Neuroevolution is now able to scale to more complex 3D environments
- Hebbian meta-learning allows quick adaptation
- Hybrid approaches for the win

Acknowledgements




Google Research Award

DANMARKS FRIE FORSKNINGSFOND
INDEPENDENT RESEARCH FUND DENMARK

amazon research awards

Thank you for your attention! Questions?

- www.sebastianrisi.com
- Email: sebastian.risi@gmail.com

@risi1979

- Meta-Learning through Hebbian Plasticity in Random Networks
  Elias Elias Najarro and Sebastian Risi. In: Thirty-fourth Conference on Neural Information Processing Systems (NeurIPS 2020)
- Deep Innovation Protection: Confronting the Credit Assignment Problem in Training Heterogeneous Neural Architectures
  Sebastian Risi and Kenneth O. Stanley
  Proceedings of the Thirty-Fith AAAI Conference on Artificial Intelligence (AAAI-2021)
- Deep Neuroevolution of Recurrent and Discrete World Models
  Sebastian Risi and Kenneth Stanley
  To appear in: *Proceedings of the Conference on Genetic and Evolutionary Computation* (GECCO 2019). New York, NY: ACM. (code)
- Evolutionary Planning in Latent Space
  Thor V.A.N. Olesen, Dennis T.T. Nguyen, Rasmus Berg Palm, Sebastian Risi
  Proceedings of the 24th International Conference on the Applications of Evolutionary Computation (EvoApplications 2021)