

KI Labor - Sommersemester 2022

Reinforcement **L**earning
Sprintwechsel & Vorstellung Assignment

Robin Baumann, Sven Müller, Maximilian Blanck,
Pascal Fecht, **Matthias Richter**, **Tim Bossenmaier**

Karlsruhe, 27. Mai 2022

Schedule



Datum	Thema	Inhalt	Präsenz
18.03.22	Allg.	Organisation, Teamfindung, Vorstellung CV	Ja
25.03.22	CV	Q&A Sessions	Nein
01.04.22	CV	Sprintwechsel, Vorstellung Assignment	Ja
08.04.22	CV	Q&A Sessions	Nein
05.04.22	Ostern		
22.04.22	CV / NLP	Abgabe CV, Vorstellung NLP	Ja
29.04.22	NLP	Q&A Sessions	Nein
06.05.22	NLP	Sprintwechsel, Vorstellung Assignment	Ja
13.05.22	NLP	Q&A Sessions	Nein
20.05.22	NLP / RL	Abgabe NLP, Vorstellung RL	Ja
27.05.22	RL	Sprintwechsel, Vorstellung Assignment	Nein
03.06.22	Ausfall		
10.06.22	Pfingsten (H-KA zu)		
17.06.22	RL	Q&A Sessions (Brückentag)	Nein
24.06.22	RL	Abgabe RL, Abschluss KI Labor	Ja

Agenda

› **Theorie**

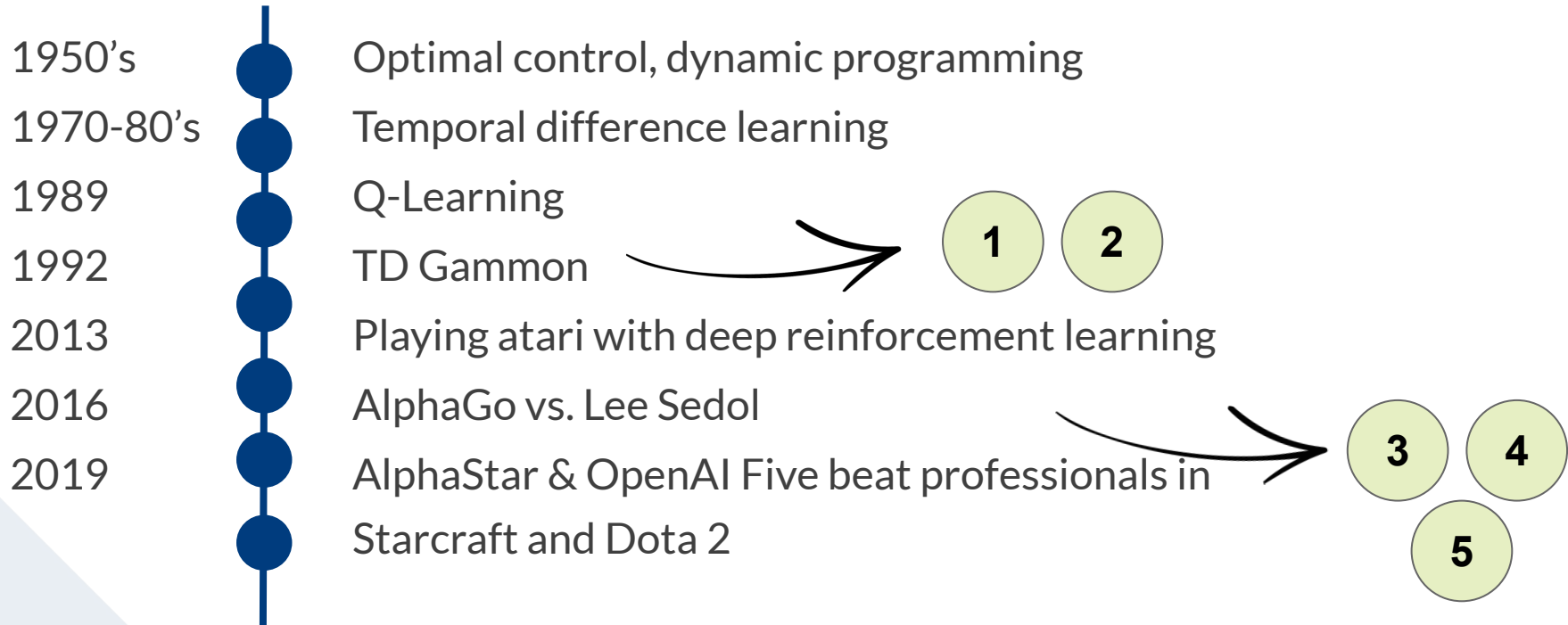
- Deep Q-Network
- Experience Replay
- Target Model
- Vorverarbeitung für Pixel-basierte Atari Games (Framestacking, etc.)

› **Praxis**

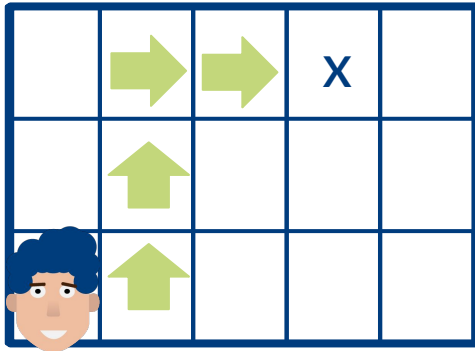
- CartPole Gym mit Deep Q-Learning (Aufgabe 3)
- Pong (Pixel-basiert) mit Deep Q-Learning (Aufgabe 4)
- Assignment

Reinforcement Learning

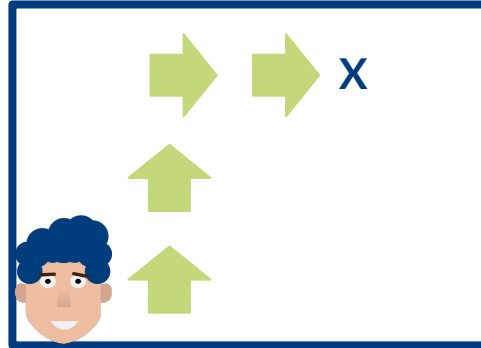
Meilensteine im Reinforcement Learning



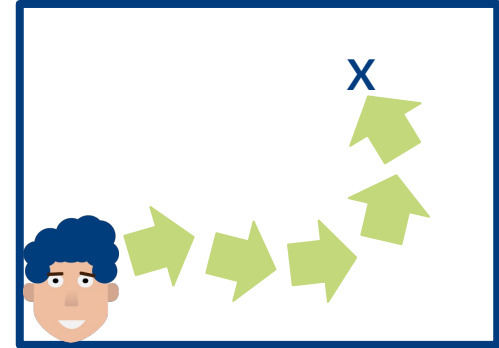
Recap



Zustände & Aktionen
diskret



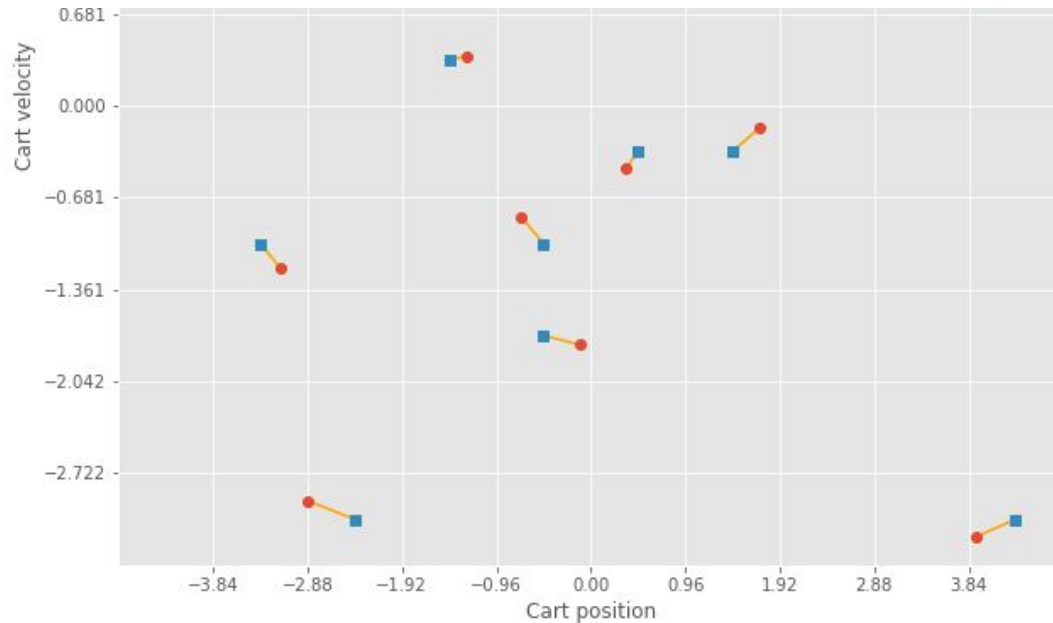
Zustände kontinuierlich &
Aktionen diskret



Zustände & Aktionen
kontinuierlich

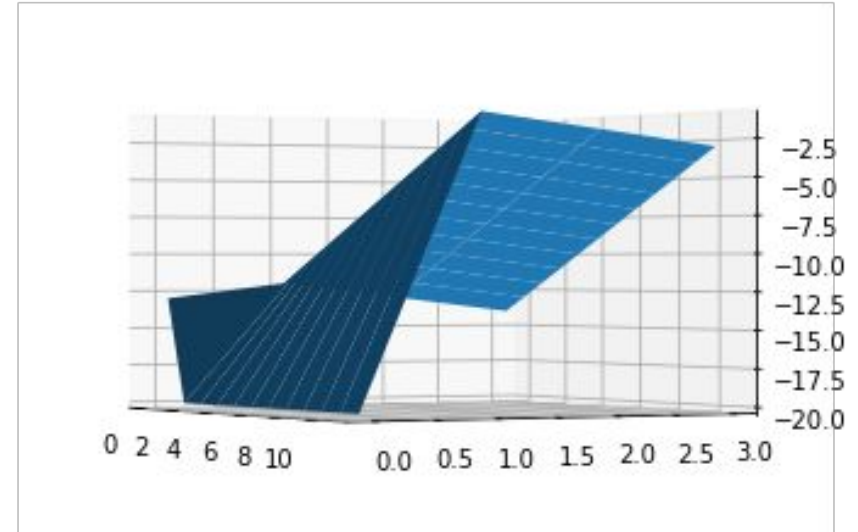
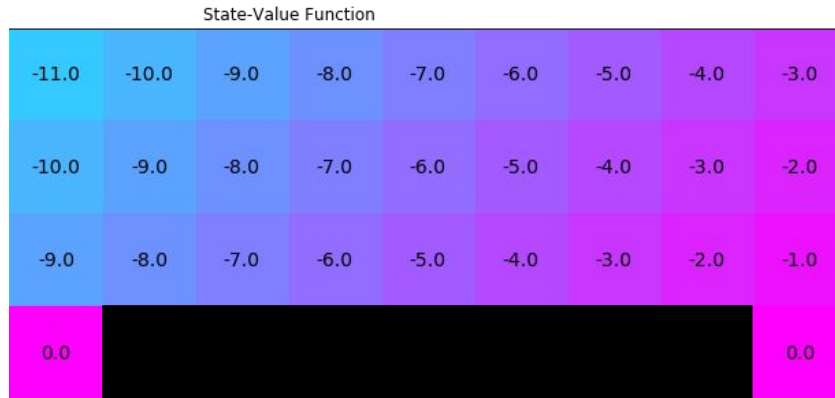
Was bisher geschah ...

Diskretisierung



Funktionsapproximation

Beispiel Cliffwalking

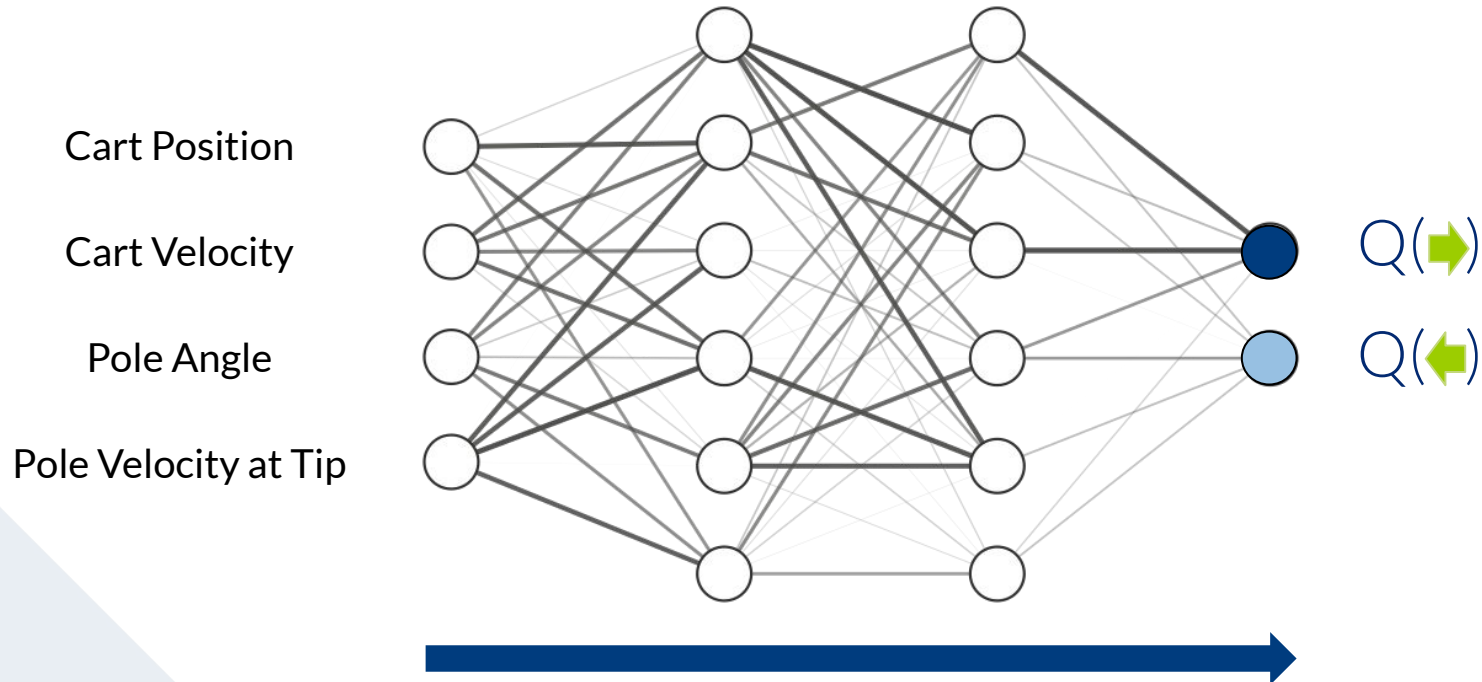


$$\hat{v}(s, \theta) \approx v_{\pi}(s), \theta \in \mathbb{R}^d$$

\hat{v} (Approximation) v_{π} (Value-Function) θ (Gewichtsvektor)

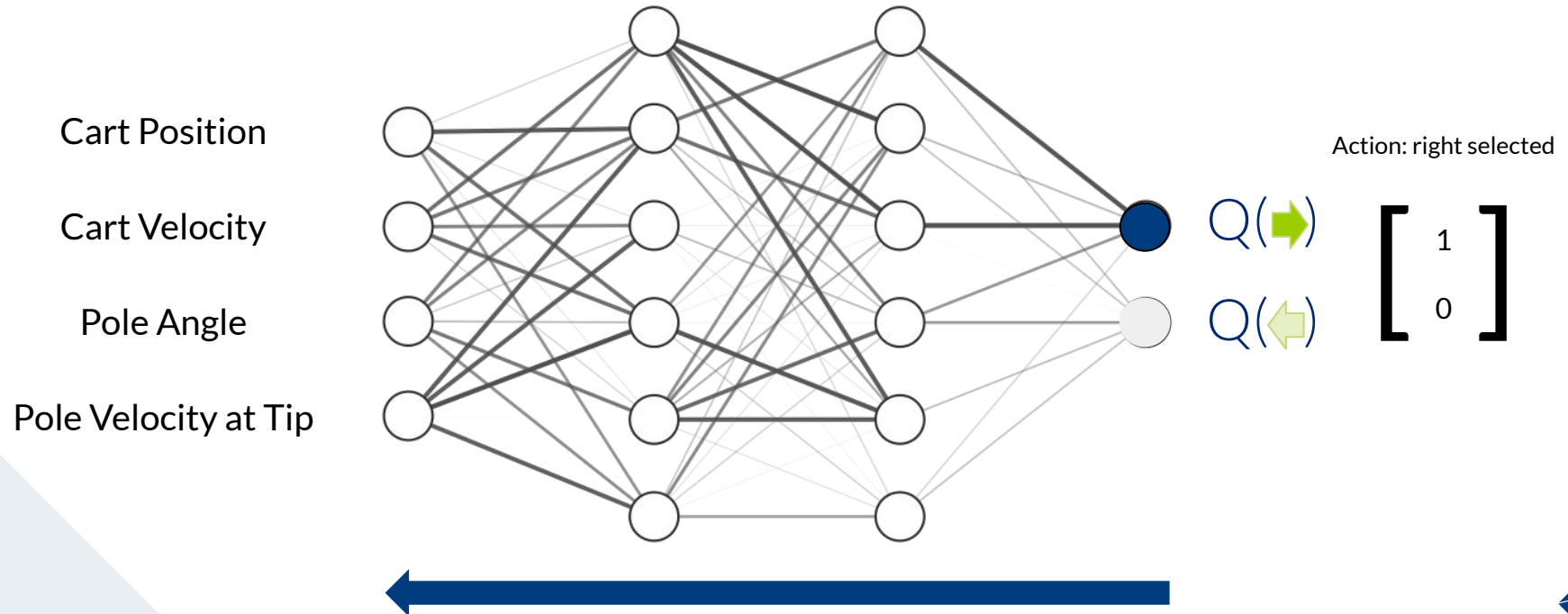
Funktionsapproximation

Beispiel CartPole: **Forward-Pass**



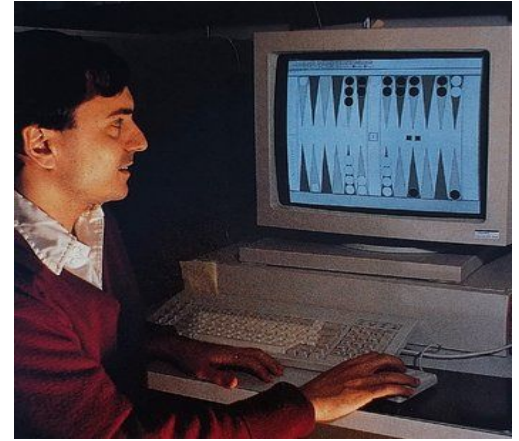
Funktionsapproximation

Beispiel CartPole: **Backward-Pass**



RL mit Neuronalen Netzen

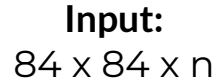
- › TD-Gammon (1995)
 - erste erfolgreiche Anwendung von NNs in RL
 - ähnliche Ansätze für andere Spiele nicht erfolgreich
 - etwa 20 Jahre keine Fortschritte in diesem Bereich
- › Ansatz erfährt durch Deep Learning und Fortschritte in Computer Vision wieder Aufmerksamkeit
 - RL kann Ende-zu-Ende gelernt werden (direkter Input von Bildern anstelle von erstellten Features) -> großer Fortschritt



Atari 2600 Spiele



Pixel-basierte Atari Games



Funktionsapproximation

Deep Q-Network

- › Approximation der Q-Funktion mit NN
 - Optimierung mit stochastic gradient descent (SGD)
 - Minimierung des Abstands zwischen Schätzer und Target

$$\underline{L_i(\theta_i)} = \mathbb{E}_{\underline{(s,a,r,s') \sim U(D)}} [\underline{(y_i - \hat{q}(s, a, \theta_i))^2}]$$

↑ ↑
Loss Weights
in Iteration i

↑
Experience
Replay

↑
Target
Q-Value

↑
approximierter
Q-Value

Target Network

- › Kopie der eigentlichen Architektur mit fixen Gewichten
- › Update alle c Iterationen

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s') \sim U(D)} [\underbrace{(y_i - \hat{q}(s, a, \theta_i))^2}_{\text{Q-Network}}]$$

Q-Network

$$r + \gamma \max_{a'} Q(s', a', \theta_i^-)$$

Target Network

Experience Replay

- Problem
 - Starke Korrelation der States erschwert das Lernen
- Lösung
 - Letzte N Experiences werden in **Replay Memory** gespeichert
 - Random Uniform Sampling

$e_{t+N} = (s_{t+N}, a_{t+N}, s_{t+N+1}, r_{t+N+1})$
...
$e_t = (s_t, a_t, s_{t+1}, r_{t+1})$

Replay Memory

Deep Q-Learning

Algorithmus

Algorithm 1 Deep Q-learning with Experience Replay

Initialize replay memory \mathcal{D} to capacity N

Initialize action-value function Q with random weights

for episode = 1, M **do**

 Initialize sequence $s_1 = \{x_1\}$ and preprocessed sequenced $\phi_1 = \phi(s_1)$

for $t = 1, T$ **do**

 With probability ϵ select a random action a_t

 otherwise select $a_t = \max_a Q^*(\phi(s_t), a; \theta)$

 Execute action a_t in emulator and observe reward r_t and image x_{t+1}

 Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$

 Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in \mathcal{D}

 Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from \mathcal{D}

 Set $y_j = \begin{cases} r_j & \text{for terminal } \phi_{j+1} \\ r_j + \gamma \max_{a'} Q(\phi_{j+1}, a'; \theta) & \text{for non-terminal } \phi_{j+1} \end{cases}$

 Perform a gradient descent step on $(y_j - Q(\phi_j, a_j; \theta))^2$ according to equation 3

end for

end for

Beispiele für Preprocessing

Warp Frame

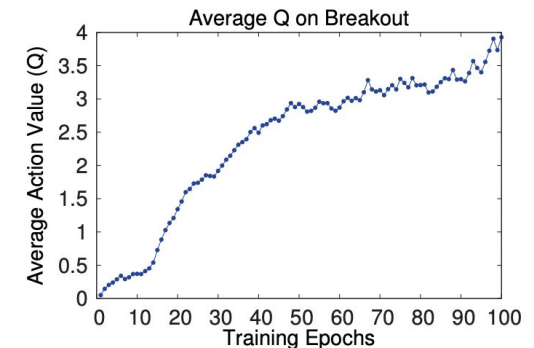
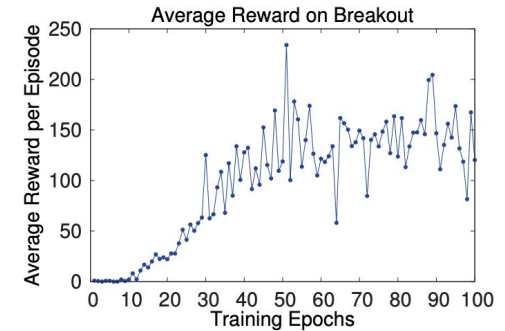
- › Konvertierung der Frames in Graustufen
- › Downsampling / Cropping

Framestacking

- › mehrere aufeinanderfolgende Frames als Eingabe, um Bewegung nachvollziehen zu können und für kürzeres Training
- › Ausführung der gewählten Aktion für gesamten Framestack

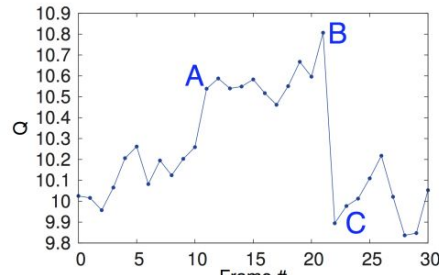
Evaluierung von Deep RL

- › Im Vergleich zu Supervised Learning deutlich herausfordernder (kein Vergleich von training und validation möglich)
- › Zwei grundlegende Metriken:
 - **Average Reward:** mittlerer, erzielter Reward je Episode
 - **Average Q-Value:** mittlerer Q-Wert für eine vor dem Training zufällig gewählte Menge an Zuständen je Episode



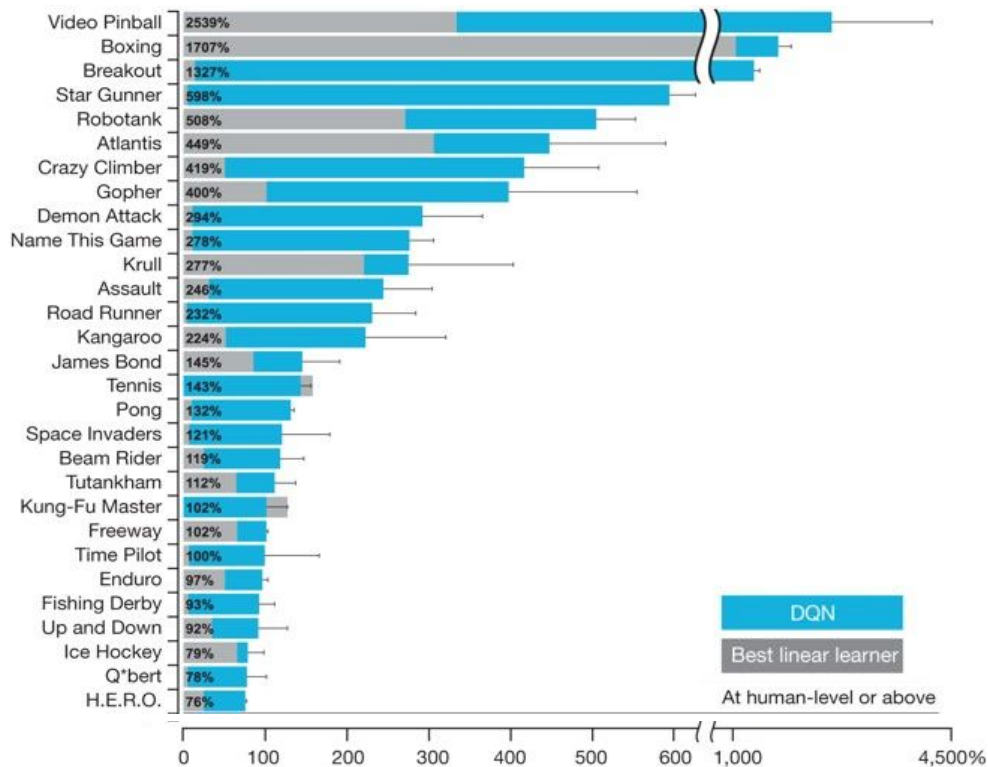
Deep Q-Network

Beispiel für eine gelernte Q-Funktion



Deep Q-Network

Pixel-basierte Atari Games



Aufgaben

Aufgaben

- › Aufgabe 3: CartPole Gym mit Deep Q-Learning (freiwillig)
- › Aufgabe 4: Pong mit Deep Q-Learning (freiwillig)
- › Aufgabe 5: Assignment (Bewertungsgrundlage)

Aufgabe 3: CartPole Gym mit Deep Q-Learning

- › **Freiwillige Bearbeitung** als Vorbereitung auf Assignment
- › Jupyter Lab Notebook



Aufgabe 4: Pong mit Deep Q-Learning

- › **Freiwillige Bearbeitung** als Vorbereitung auf Assignment
- › Jupyter Lab Notebook



Aufgabe 5: Assignment

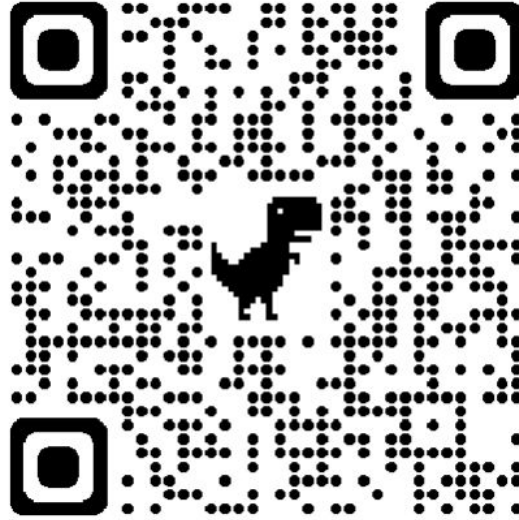
- › **Assignment dient als Bewertungsgrundlage = Pflicht**
- › zwei Atari Spiele zur Auswahl
 - Enduro (Notebook 5a)
 - Breakout (Notebook 5b)
- › Jupyter Lab Notebook
- › Freie Wahl des Ansatzes



Infomaterial

- › Kostenlose "Standard"-Lektüre für den Einstieg in RL:
Reinforcement Learning: An Introduction (Sutton and Barto), siehe <http://incompleteideas.net/book/RLbook2018.pdf>
- › Ausführlich und gut erklärter Einstieg in RL (Video-Lektionen):
UCL Course on RL (David Silver, Google DeepMind), siehe <https://www.davidsilver.uk/teaching/>
- › *Algorithms in Reinforcement Learning* von Csaba Szepesvári, siehe <https://sites.ualberta.ca/~szepesva/papers/RLAlgsInMDPs.pdf>
- › Blog mit Videos zum Einstieg in RL und Q-Learning, DQN und vieles mehr:
Reinforcement Learning – Introducing Goal Oriented Intelligence, siehe <https://deeplizard.com/learn/video/nyjbcRQ-uQ8>
- › *David Silver - AlphaGo, AlphaZero, and Deep Reinforcement Learning*, siehe Lex Fridman Podcast #86 <https://lexfridman.com/david-silver/>

Feedback



<https://forms.gle/4UeEVTWmHGpRGYkm9>

Vielen Dank

Matthias Richter

matthias.richter@inovex.de

Tim Bossenmaier

tim.bossenmaier@inovex.de

