All you can Creep!

Lukas Dennert
Sofie Henghuber
Balthasar Schüss
Denny Steigmeier

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Autonome Systeme Praktikum | Wintersemester 2020/21 | 16.02.2021

Domäne = Wurm

3x Segmente

- Bodenkontakt
- Geschwindigkeit
- Winkelgeschwindigkeit
- Inverse Transformations-Richtung
- Segmentrotation

Reward

- = Velocity * Orientation
- Random Agent Reward = 1
- Baseline Avg. = 800
- Anfangs geringer Information Gain

3x Gelenke

- Bodenkontakt
- Geschwindigkeit
- Winkelgeschwindigkeit
- Winkel
- je 3 Aktionsmöglichkeiten

1x Kopf

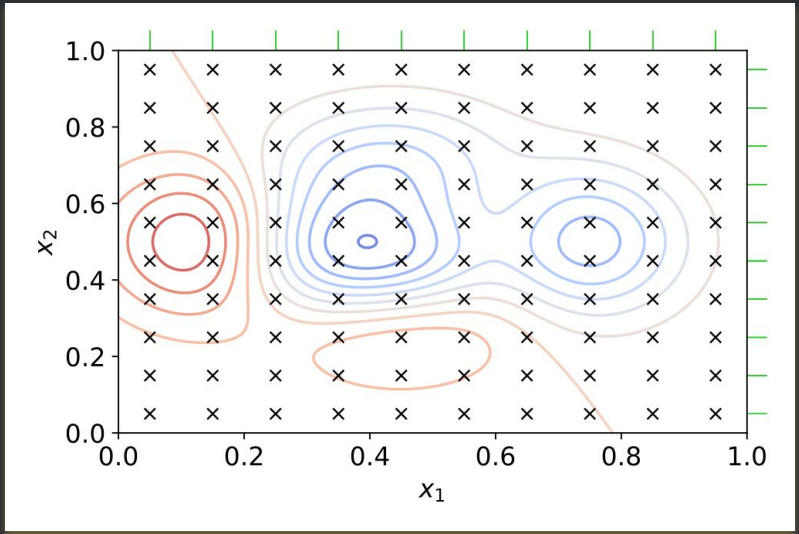
- Sichtdistanz
- Zielgeschwindigkeit
- Rotation
- Zieldistanz

Hyperparametersuche

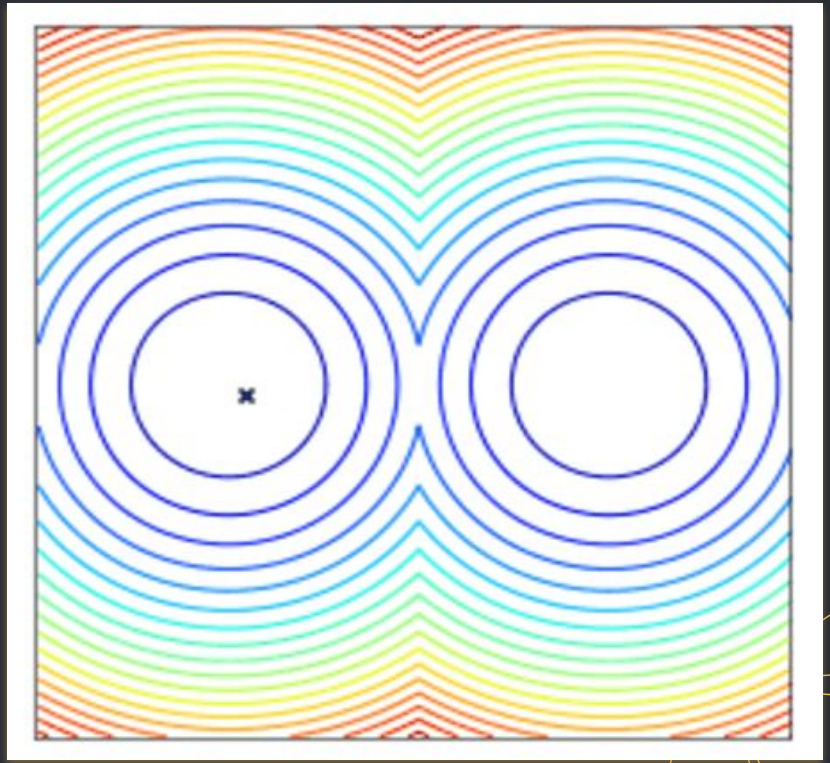
Evolutionärer Algorithmus →

Gridsearch

↓

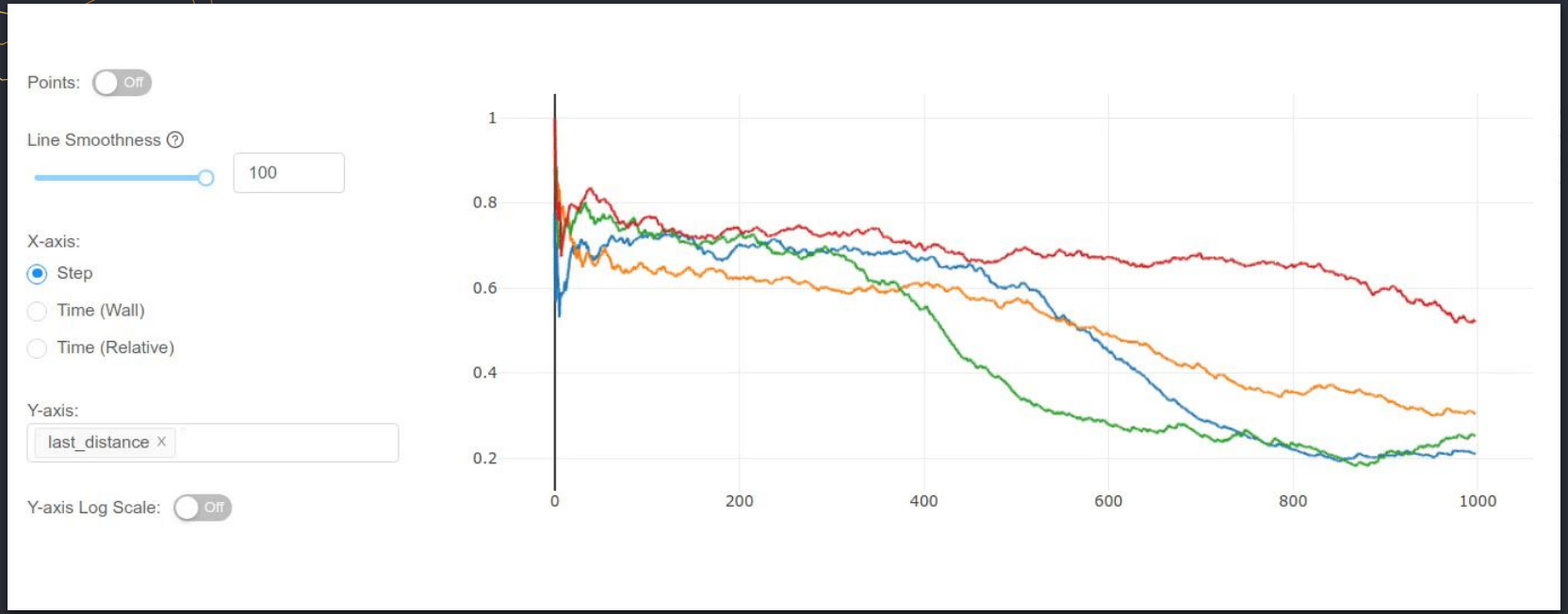


https://upload.wikimedia.org/wikipedia/commons/b/b6/Hyperparameter_Optimization_using_Grid_Search.svg



https://github.com/facebookresearch/nevergrad/blob/master/docs/resources/TwoPointsDE.gif

ML-Flow

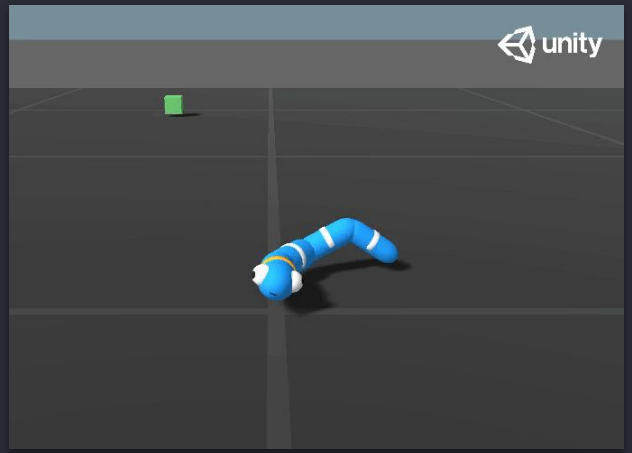


Remote Server, 2000+ training runs, Slurm parallelization & OOM

Evaluierung: A2C

Hidden neurons: 64, gamma: 0.995, max_scale_clamping: 0.3, batch_size: 10, scale_activation: sigmoid

0 Episoden

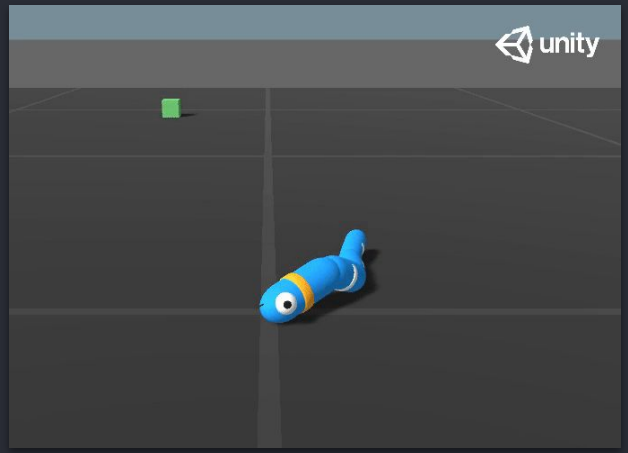


5000 Episoden ⇒ Reward: 159

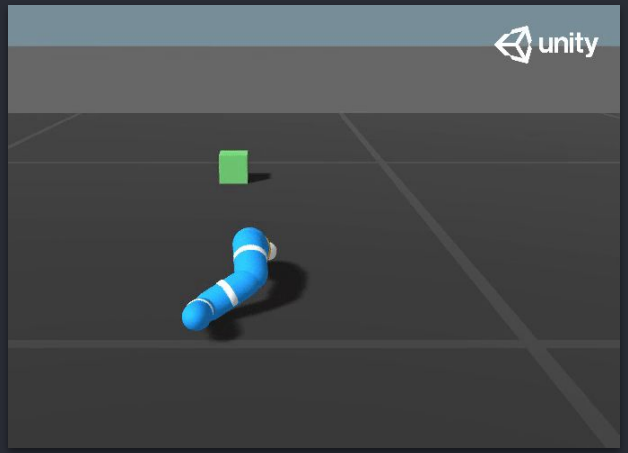
7500 Episoden ⇒ Reward: 341

Evaluierung: PPO

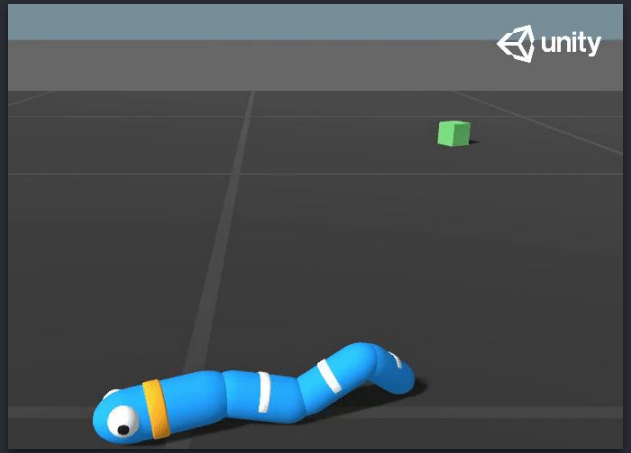
Hidden neurons: 512, gamma: 0.995, batch_size: 2024, buffer_size: 20240, scale_activation: sigmoid



0 Episoden

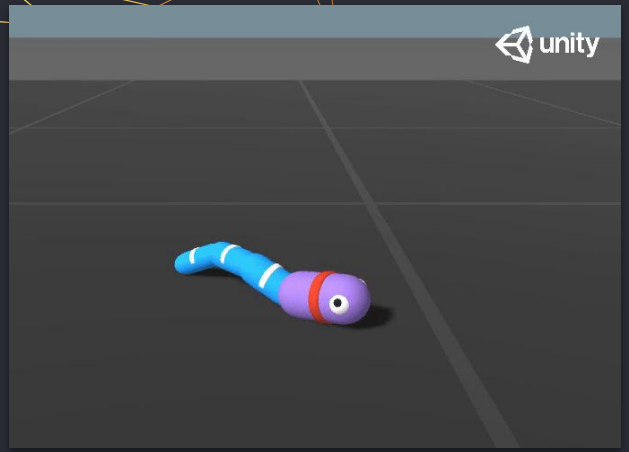


10000 Episoden ⇒ Reward: 300

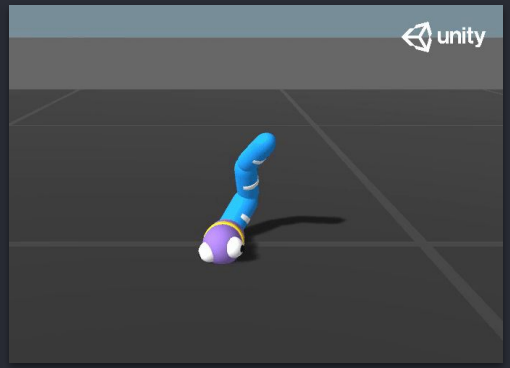


25000 Episoden ⇒ Reward: 1000

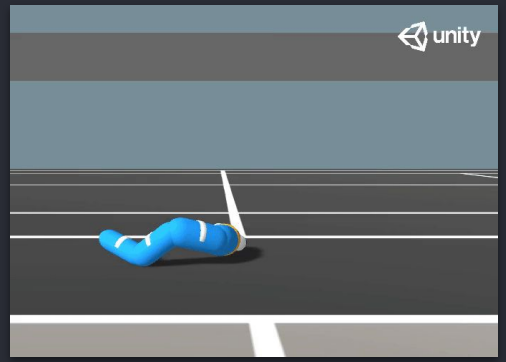
Modifikationen



~60% größerer Kopf

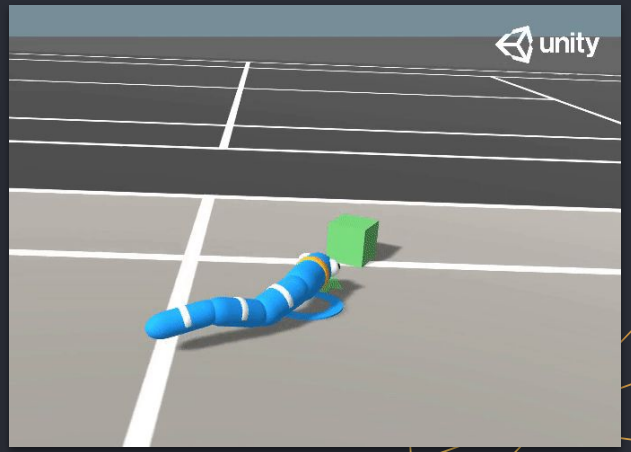


Kopf zu groß → zu schwer (“Kopflastig”)

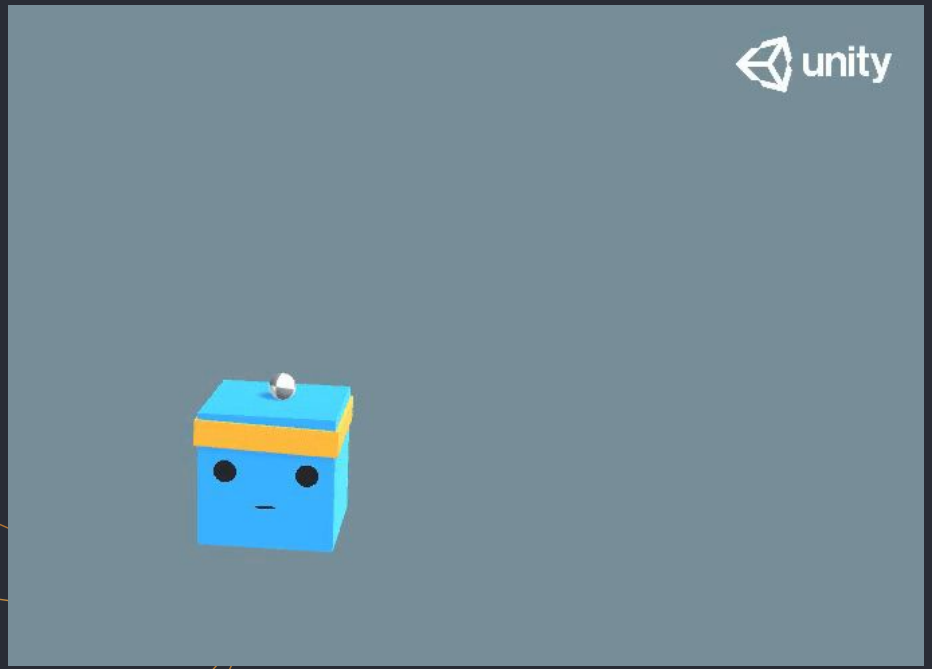


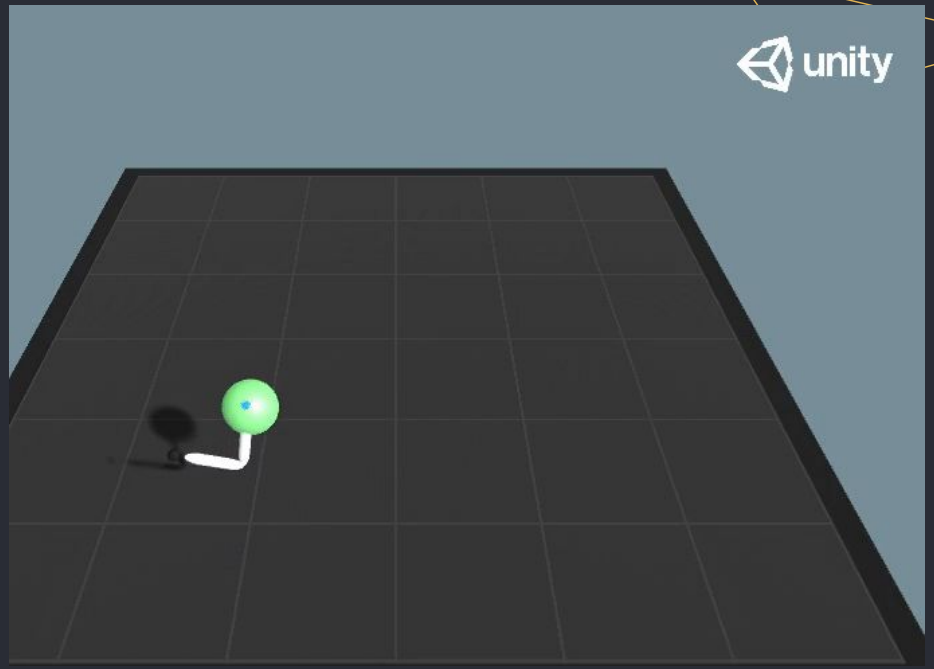
Environment zu stark gekippt

2° Breite und 2° Länge gekippt



Andere Domänen





Time Scale = 3x

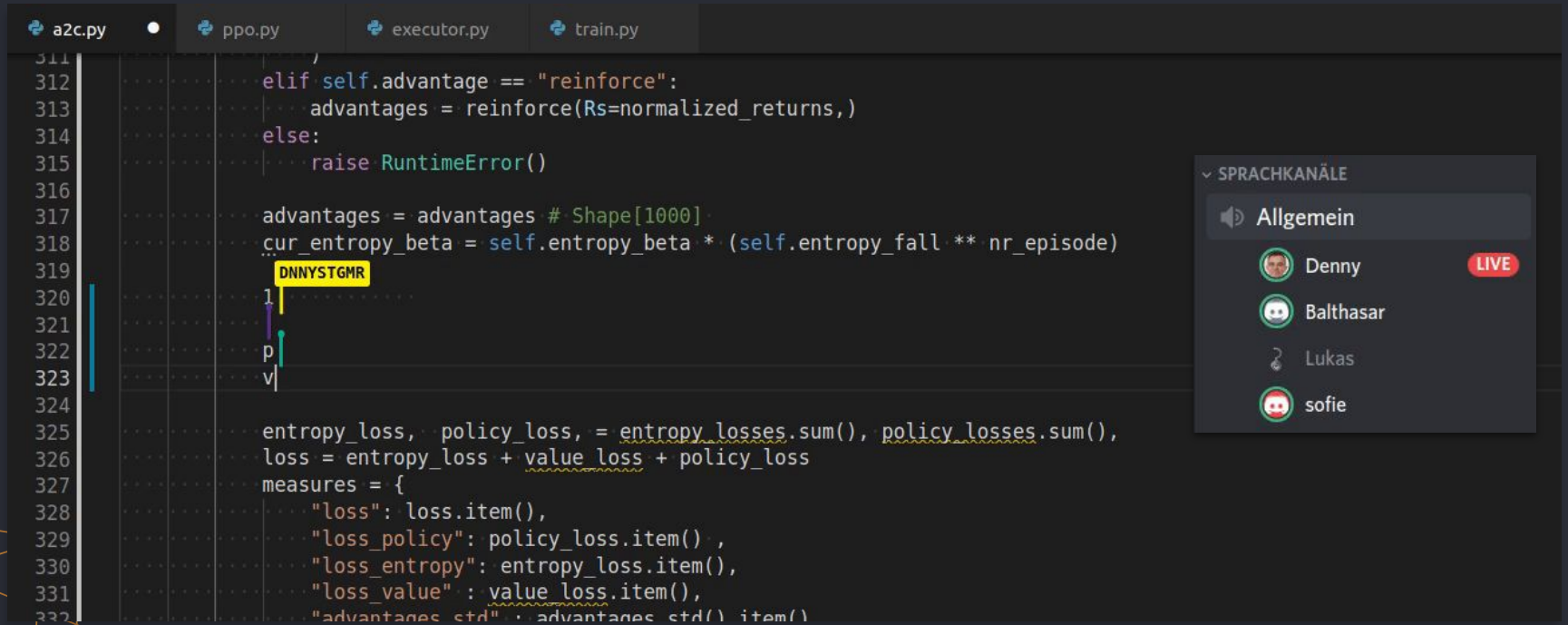
Vergleich

PPO	A2C
Schneller und stabilere Performance ⟶ Baseline übertroffen	A2C Implementierung nicht stabil genug
512 Hidden Neurons vorteilhaft	64 Hidden Neurons ausreichend
Wichtige Parameter: Batch- und Buffersize (Vielfaches)	Deutliche Modifikationen: Gamma, Batchsize (A3C), max clamping action_scale
Sigmoid activation für action_scale	Sigmoid activation für action_scale

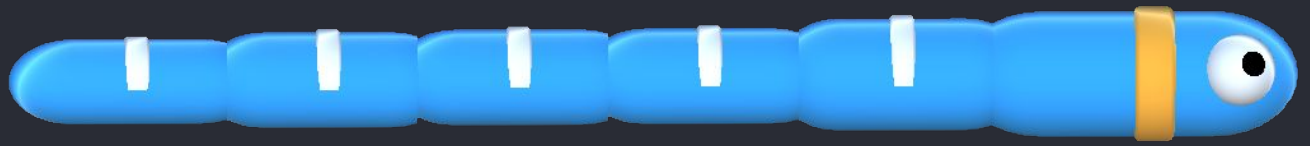
Algorithmen aus Papern nicht 100% anwendbar
Modifikationen für komplexere Aufgaben

Aufgabenverteilung (letztendlich)

Visual Studio Code mit Live Share

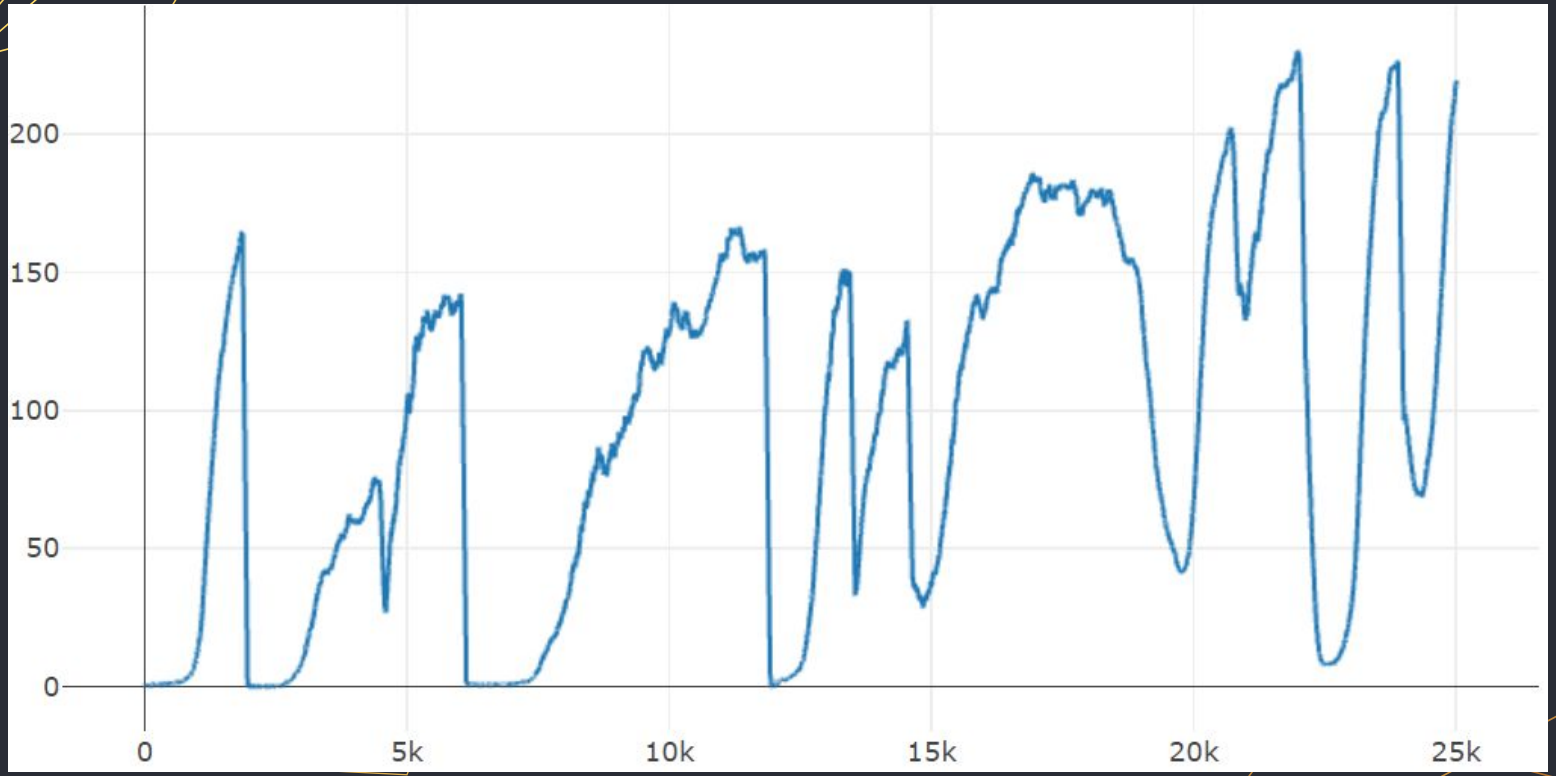


Thank you, for your attention.

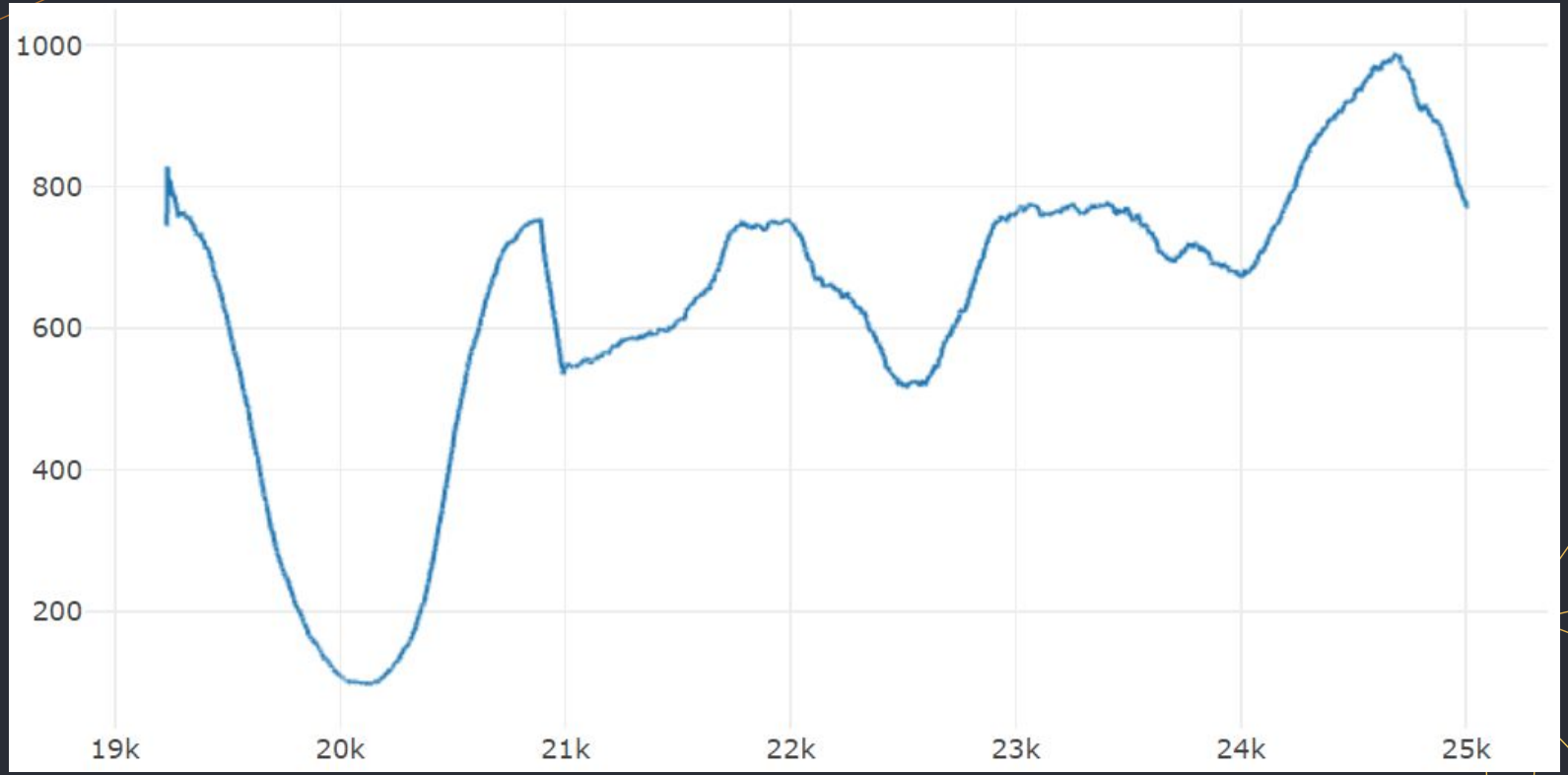


Credit: Presentation template by Slidesgo & icons by Flaticon

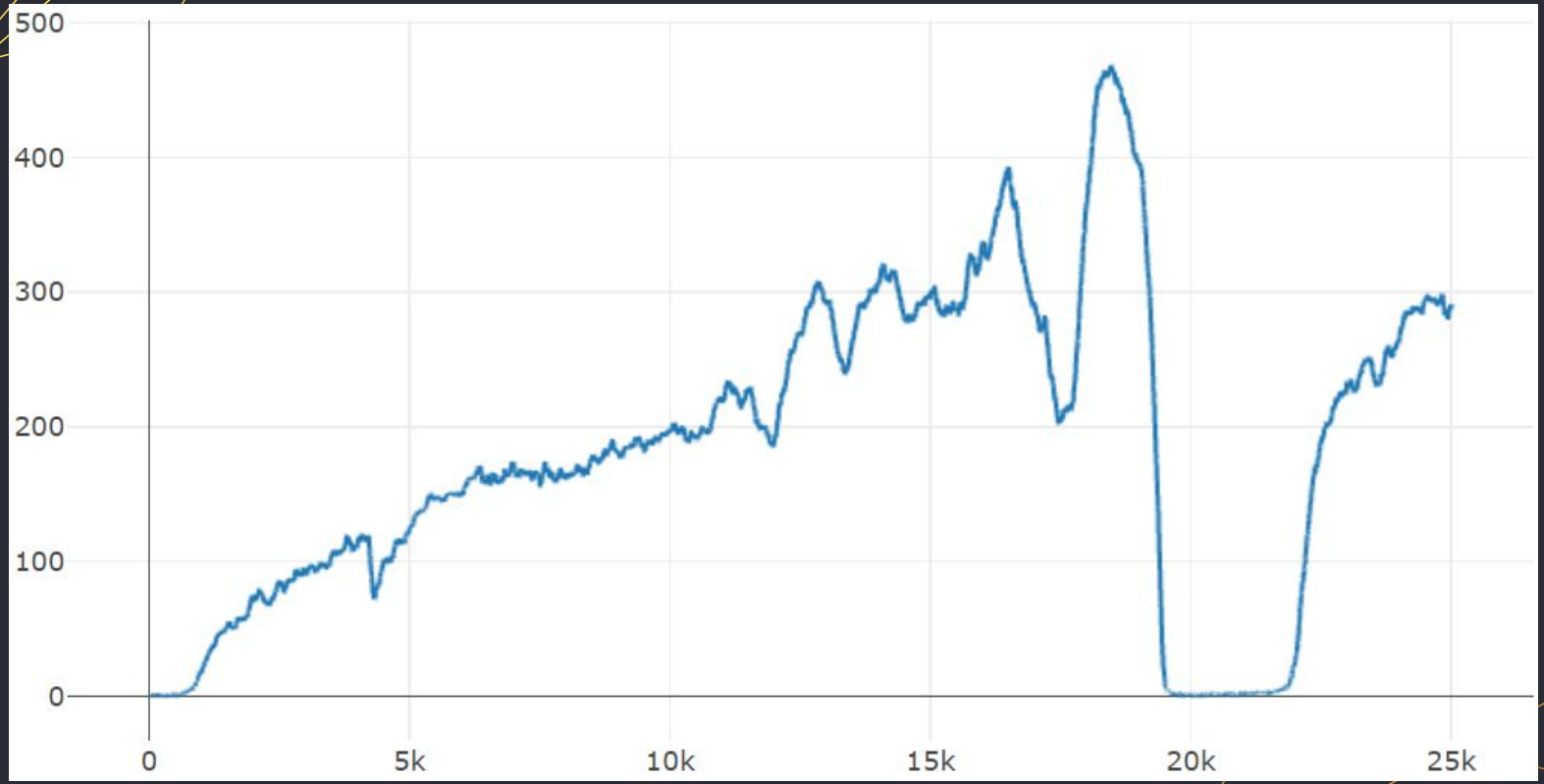
PPO Instabilität



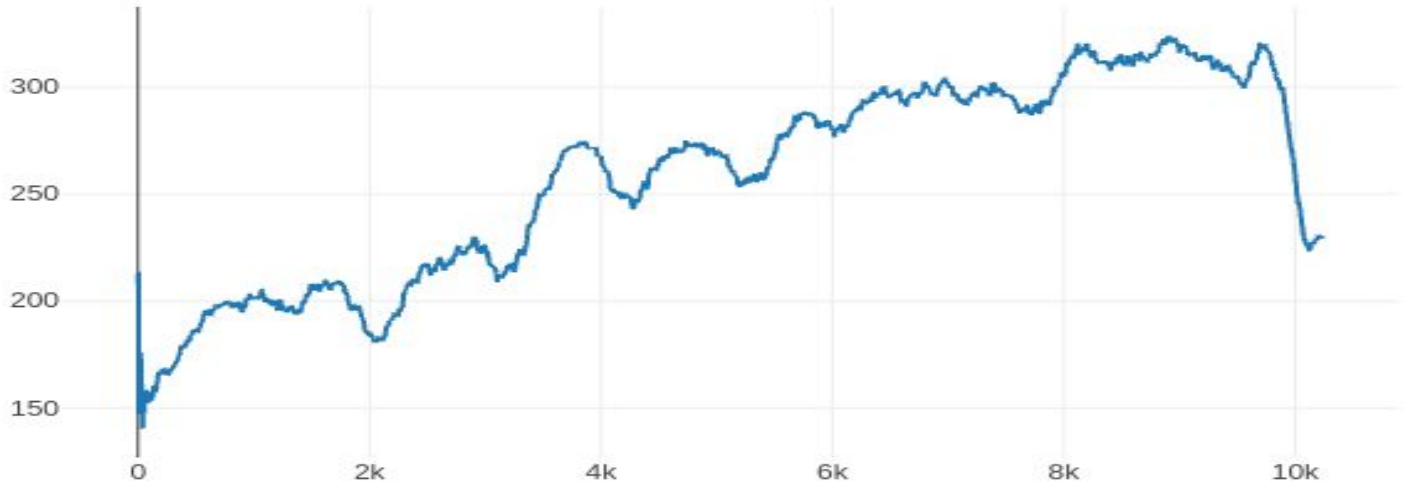
PPO 25000 Episoden Model



PPO Bighead



A2C Instabilität



Aufgabenverteilung (ursprünglich)

Xtreme-Pair Programming mit Hauptverantwortlichen (HV, NV)

Einbindung in Unity, Konzeption Experimente - Denny, Sofie

A2C - Lukas, Denny

Hyperparameter Tuning - Sofie, Balthasar

Evaluation und Visualisierung - Balthasar, Lukas

(Future Work: Off-Policy Alternative / A3C)

Aufgabenverteilung (letztendlich)

Topic	Name	Info
A2C		
Split- & Multihead NN	Sofie	
Activation	Balthasar	Sigmoid, Softplus, Softmax, TanH, ReLu
Min-Max-Clamping	Balthasar	
Loss & Entropy	Balthasar	
Advantages	Sofie	
Return	Sofie	
A2C vs A3C	Sofie	
PPO		
Actor & Critic NN	Lukas	
Memory, Buffer, Batches	Lukas	
Hyperparameter	Denny	
Reward	Denny	
log_prob & prob_ratio	Denny	
weighted_probs & clipping	Lukas	
Slurm	Denny	Slurm Runner
Parameter Search	Sofie	Grid Search, Evolutionary Algorithm
Environments + Unity	Lukas	
Ml-Flow	Balthasar	Measures, Artifacts
Save and Load Models	Balthasar	

4 Segments