Deep Reinforcement Learning in a Modern 3D Video Game

Salzburg University of Applied Sciences

Samuel Arzt [1], Gerhard Mitterlechner [2], Markus Tatzgern [3], Thomas Stütz [4]

Motivation

- Traditionally, Reinforcement Learning (RL) applied to handcrafted features and complicated reward signals.
  Modern RL: end-to-end learning, high-dimensional inputs, large state / action spaces. Video-games often used as testbeds.
  Previous work includes agents trained on retro-games, e.g. Atari2600 [1, 7] or Doom [8, 9].
  Ultimate goal: explore new possibilities to solve real-world problems.
- Main questions of this work:
  - Can recent RL methods also be applied to visually more complex modern 3D video-games?
  - What is the impact on learning performance of various modifications of the agent's visual inputs in comparison to other modifications to the environment?

Contribution

- Introduction of a modern 3D video-game environment for RL.
- Two modern Deep RL techniques applied in this novel environment:
  - DQN (Deep Q-Network) [1-5]
  - A2C (Advantage Actor Critic) [6]
- Evaluated impact on learning performance of different modifications to the environment:
  - Environment representation (complex 3D, simple 2D)
  - Agent motion control (physics-based, linear)
  - Reward signal (sparse, continuous)
- Learning environment, API and reproducible results available at: *github.com/ArztSamuel/DRL_DeliveryDuel*

Novel Learning Environment – Delivery Duel

- Modern 3D video-game developed using the Unity Engine [10].
- Gameplay:
  - Player controls delivery-van in an open city environment.
  - Objective: deliver pizzas to destinations by throwing them out of the van.
  - After each delivery, return to base.
  - Score rewarded for deliveries and return to base.
- Used RL Frameworks:
  - OpenAI Gym [11]
    (Environment Interface Definition)
  - OpenAI Baselines [12]
    (DQN and A2C implementation)
  - Unity ML-Agents [13]
    (Interface from Unity to Python)

Fig. 1. Comparison of the complex 3D render-mode, i.e. as in the original game, and the simple 2D render-mode, with only important information

References

[1] Volodymyr Mnih et al. "Human-level control through deep reinforcement learning". In: *Nature* 518.7540 (2015), pp. 529-533.

[2] Tom Schaul et al. "Prioritized Experience Replay". In: *International Conference on Learning Representations (ICLR)*. 2018.

[3] Hado Van Hasselt, Arthur Guez, and David Silver. "Deep Reinforcement Learning with Double Q-Learning". In: *AAAI*. 2016, pp. 2094-2100.

[4] Ziyu Wang et al. "Dueling network architectures for deep reinforcement learning". In: *International Conference on Machine Learning (ICML)*. 2016, pp. 1995-2003.

[5] Matteo Hessel et al. "Rainbow: Combining Improvements on Deep Reinforcement Learning". In: *International Conference on Artificial Intelligence (IJCAI)*. 2015, pp. 554-560.

[6] Volodymyr Mnih et al. "Asynchronous methods for deep reinforcement learning". In: *International Conference on Machine Learning (ICML)*. 2016, pp. 1928-1937.

[7] Marc G Bellemare et al. "The Arcade Learning Environment: An evaluation platform for general agents". In: *J. Artif. Intell. Res. (JAIR)* 47 (2013), pp. 253-279.

[8] Guillaume Lample and Devendra Singh Chaplot. "Playing FPS Games with Deep Reinforcement Learning". In: *AAAI*. 2017, pp. 2140-2146.

[9] Michal Kempka et al. "Vizdoom: A doom-based ai research platform for visual reinforcement learning". In: *Computational Intelligence and Games (CIG), 2016 IEEE Conference on.* IEEE. 2016, pp. 1-8.

[10] Unity-Technologies. *Unity3D.* 2018. URL: http://web.archive.org/web/20180412233455/https://unity3d.com/ (visited on 04/14/2018)

[11] Greg Brockman et al. *OpenAI Gym.* 2016. eprint: arXiv: 1606.01540.

[12] Prafulla Dhariwal et al. *OpenAI Baselines*. https://github.com/openai/baselines. 2017.

[13] Unity-Technologies. *ML-Agents Homepage*. 2018. URL: https://web.archive.org/web/20180322013330/https://unity3d.com/machine-learning/ (visited on 04/14/2018).

Results

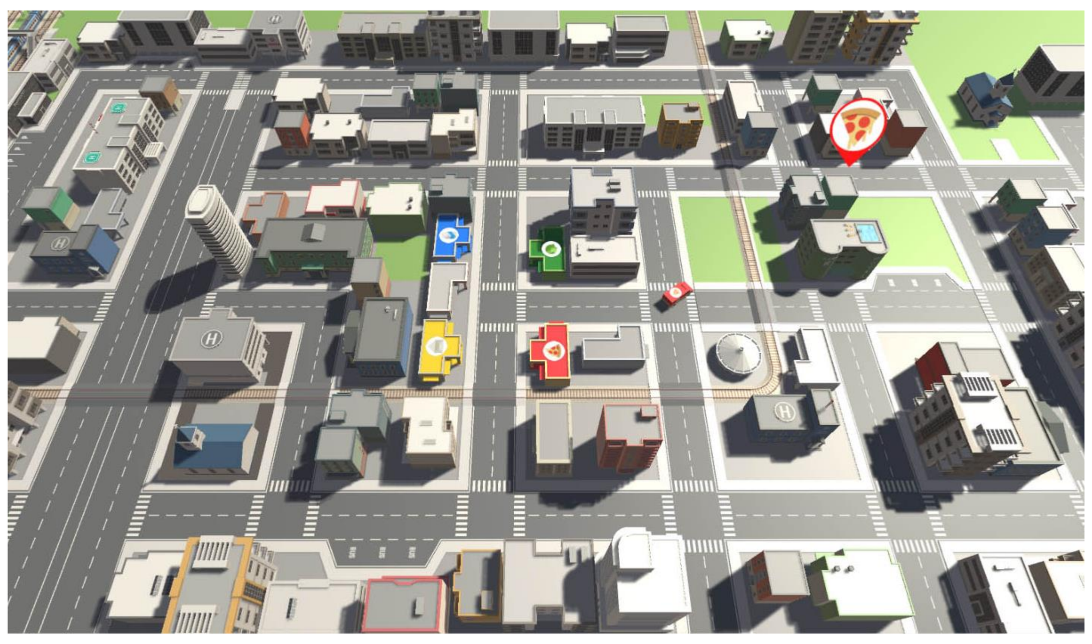

Fig. 2. In-game view of the 3D video-game Delivery Duel. The player has to deliver pizzas to destinations located in an open city environment. Items are delivered by throwing them out of the player's delivery-van

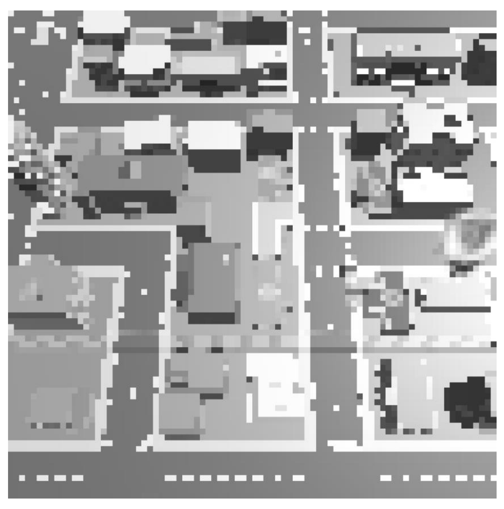

Fig. 3. Example input image of the neural network. The input consists of a stack of four frames down-sampled to 84x84 pixels and grey-scaled, as in [1]

- Agents trained for five million steps on each possible environment configuration.
- Agents compared by mean reward achieved over the last 100 episodes.
- Overall DQN performed better than A2C.
- Continuous reward and linear movement more important than visual complexity.
- No agent able to complete task when confronted with combination of complex rendering and physical motion control.
- Implies that employed CNN able to cope with additional visual complexity, however only as long as interaction problem remains simple enough.
- Possible future transfer learning from linear to physical movement.

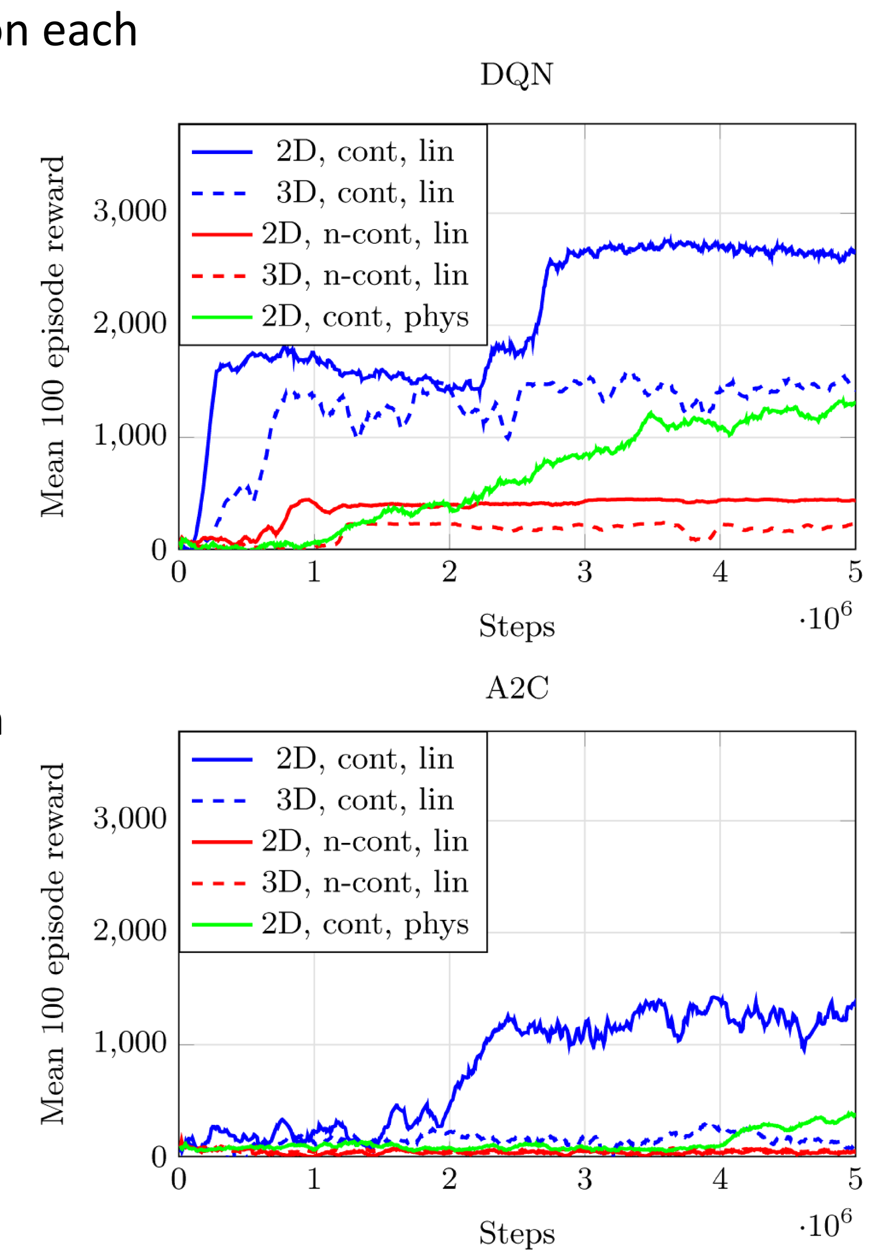

Fig. 4. Comparison of the performance of each DQN (*top*) and A2C (*bottom*) agent on different environment configurations. 2D / 3D = simple / complex render-mode; (n-)cont = (non-)continuous reward; lin / phys = linear / physical motion control

Conclusions

- Novel modern 3D game environment for RL presented and made publicly available.
- Two recent Deep RL methods (DQN and A2C) successfully applied.
- Agents able to complete task even with complex 3D visual input.
- More continuous reward greatly improves learning performance.
- Agents able to cope with complex physical motion control.
- Applied methods fail when confronted with combination of complex visual input and complex motion control.
- Encourages future experiments on this environment, using other algorithms and transfer / curriculum learning or using the multiplayer of Delivery Duel.

Acknowledgements

Delivery Duel was developed in collaboration of Katrin-Anna Zibuschka, Lukas Machegger and Samuel Arzt, who approved to make the game publicly available for scientific purposes. Their work and approval is greatly appreciated.

Furthermore we thank the University of Applied Sciences Salzburg for the provided assistance, including scientific advice and research equipment, which has been a great help in conducting this work.

[1] arzt.samuel@live.de, [2] gmitterlechner.lba@fh-salzburg.ac.at, [3] markus.tatzgern@fh-salzburg.ac.at, [4] thomas.stuetz@fh-salzburg.ac.at, [1, 2, 3, 4] Salzburg University of Applied Sciences