Representation-Invariant Latent Spaces

Prof. Gerhard Neumann,
Niklas Freymuth, Vaishak Shaj

The prediction quality of Deep Neural Networks (DNNs) varies greatly depending on the representation of the data they are trained on. Additionally, it can be very difficult to translate between different representations, even if they contain similar information. As an example, the top part of Figure 1 shows both a human and a robot arm pushing an object to a certain target position. While both contain roughly the same semantic information ('the object is pushed towards the target position'), their representations are totally different. Motivated by the vast amount of videos available online, we want to develop a Reinforcement Learning agent using DNNs that is able to efficiently learn from human video demonstrations.

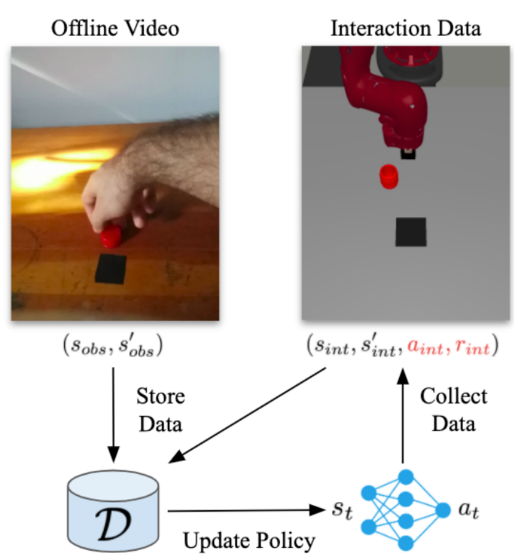

Figure 1: How can we integrate human demonstrations into a Reinforcement Learning framework [3]?

In this project, the goal is to develop an approach (a network architecture and/or a training regime) that results in a *representation-invariant latent space*, i.e., some hidden layer of the network that contains useful information about the data regardless of its input representation. To this end, we will leverage approaches related to Variational Auto Encoders (VAEs) [2] and the fields of Transfer Learning (see e.g., [4]) and Representation Learning [1]. For example, a first approach could be to train a VAE jointly on different representations and shape its latent space by requiring that provided data pairings map to the same point in this space (c.f. [3]). This latent space will then be used to learn from demonstrations in both simulated and physical robot tasks.

References

[1] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[2] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *stat*, 1050:10, 2014.

[3] Karl Schmeckpeper, Oleh Rybkin, Kostas Daniilidis, Sergey Levine, and Chelsea Finn. Reinforcement learning with videos: Combining offline observations with interaction. *Conference on Robot Learning (CoRL)*, 2020.

[4] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.