Abstract

Creating agents capable of adapting to any context means enabling them to function effectively even when humans and robots share the same environment without physical barriers, a concept known as human-robot collaboration in industrial contexts. While collaboration comes naturally to humans, it poses a significant challenge for robots, as it requires several key elements: understanding human actions, knowing how to interact with people, and responding appropriately to the surrounding environment. Computer vision is essential for enabling robots to collaborate with humans. However, current state-of-the-art approaches are highly data-dependent, requiring vast amounts of annotated data to function effectively. Acquiring such data is often time-consuming, which hinders the adaptation of these solutions to real-world applications.

This thesis demonstrates that the data requirements for human-robot collaboration can be reduced through three main paradigms: leveraging high-level general representations, exploiting few-shot learning, and generating data. These approaches are applied to the three crucial tasks for human-robot collaboration: action recognition, robot grasping, and imitation learning. Understanding actions is a key characteristic for the robot to meet human needs, while robot grasping allows autonomous interaction with objects in the environment. Finally, imitation learning enables robots to perform tasks directly from humans, delivering agents that can adapt to new requirements by imitating human actions.

The thesis demonstrates that the three proposed paradigms allow for the creation of adaptable robots. Using high-level representations, such as skeletal models, as input to action recognition models enables the development of frameworks that generalize to various environments. Few-shot approaches leverage learned representations to identify similarities between a small set of examples and a target image. This capability is particularly beneficial in contexts like semantic robot grasping, as it enables systems to adapt to human needs. Lastly, data generation techniques can enhance performance by expanding the training distribution, which is especially beneficial in imitation learning, where data collection is labor-intensive due to the need for manual robot manipulation. Data generation reduces, or even eliminates, the need for extensive demonstration recording.

Overall, this thesis shows that reducing data requirements is possible, marking a critical step toward real human-robot collaboration.