Autonomous robotic manipulation in novel situations is widely considered as an unsolved problem. Solving it would have enormous impact on society and economy. It would enable robots to help humans in factories, warehouses, but also at homes in everyday tasks.

This project will advance the state of the art in learning algorithms which will enable a robot to autonomously perform complex dexterous manipulation in unfamiliar situations in the real-world with novel objects and significant occlusions. In particular to grasp, push or pull rigid objects, but also to some extent, non-rigid and articulated objects that can be found at homes or in warehouses. Importantly, similarly to humans, a robot will also be able to envision results of its actions, without actually performing them. A robot could plan its actions, e.g. to tilt a low object to be able to grasp it (Fig. 1a), to bend a cloth before gripping and then pulling (Fig. 1b), or to open a door (Fig. 1c).
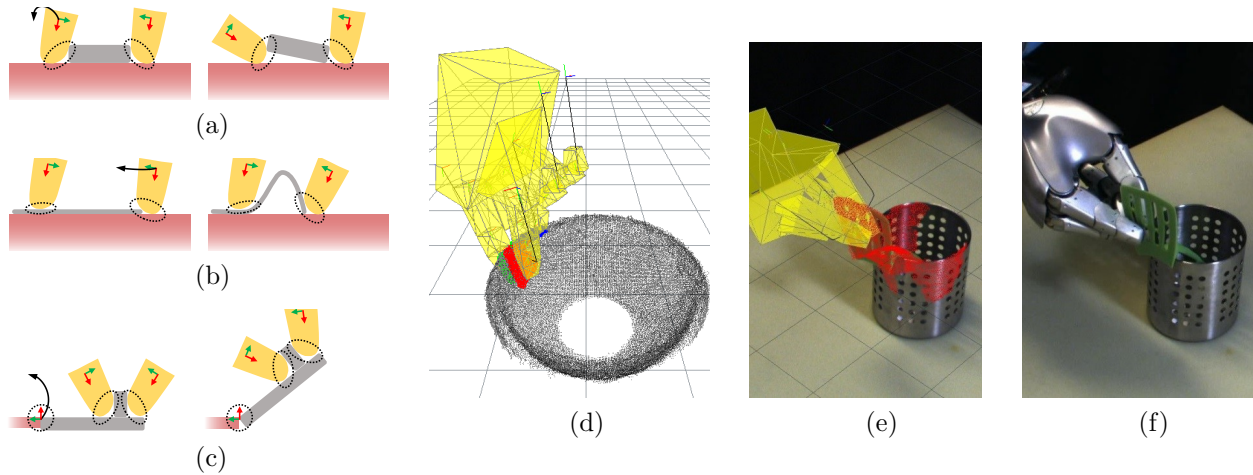

Figure 1: *Examples of manipulation scenarios* (a-c) In columns: grasping a low rigid object (a) or a sheet of paper/cloth (b), from a table/flat surface/another object; opening a door (c).
Point-based contact models (d-e): demonstration of point-based feature contacts, one per each robot link (d), transfer of contacts to a novel point cloud and performed grasp (e).

End-to-end/black-box approaches have made it possible to learn some of these kinds of complex tasks with almost no prior knowledge, for example by learning RGB image to motor command mapping directly. While this is impressive, typically the end-to-end manipulation learning requires prohibitive amounts of training data, scales poorly to different tasks, does not involve prediction and struggles to provide interpretation of its behaviour.

On the other hand, it was shown that exploiting the problem structure can help learning. For example, in bin picking scenarios, many successful approaches assume at least depth images or trajectory planners for robot control. In autonomous driving, adding representations such as depth estimation, or semantic scene segmentation enable to achieve the highest task performance.

We introduced a first algorithm that was capable to learn to grasp novel objects from just one demonstration. It models grasps as a product of object-hand link contacts (Fig. 1d-1f), which accelerates learning and provides a flexible problem structure. Similarly, we modelled object-object contacts, e.g. to facilitate placement, or to predict object behaviour in pushing and grasping. However, our current models either do not involve prediction, they are insensitive to task context and prone to occlusions, or they assume object exemplars in prediction.

In this project, we will overcome all these limitations, and create a novel robotic manipulation algorithm, that can reliably learn contacts and predict their changes after applying manipulation actions in complex scenarios (Fig. 1a-1c). A robot will be trained from demonstration and can further improve its predictions online, while performing its manipulation tasks.