Learning multi-granularity task primitives from construction videos for human-robot collaboration

Zaolin PAN, Yantao YU
The Hong Kong University of Science and Technology, Hong Kong, China

To foster human understanding of robots, on the other hand, it is crucial to understand what a robot learns from the visual world. The visual data captured in the field will inevitably contain noise and interference because of the complex and dynamic environment. If robots make decisions based on this irrelevant information, it can compromise the versatility of robots, undermine worker trust in robots, and even endanger worker safety. One viable option to address this problem is to provide visual explanations for the perception processes of robots (Anjomshoae *et al.*, 2019). Visual explanations of what robots have learned from the visual world to solve the task allow us to assess the reliability and reasonableness of robot decisions. This helps increase trust, smoothness, and productivity in

2.1 Construction Task Modelling

Task modelling

3. Methodology

To bridge these gaps, this study proposes a ConvNet-based method for learning multi-

Step: A step is an element of completing a subtask and the goal state of a sequence of activities, which can be implemented with multiple sequences of activities.

Subtask: A subtask is the basic unit of a task. It is achieved through a defined sequence of steps.

Task: A task represents the work to be conducted by a HRC team. It specifies the initial state and the goal state. It can be decomposed into a set of fixed-order subtasks.

The hierarchical relationship of the above task primitives is shown in Fig. 1, which represents a real-world scaffn ng condsaruhion

6.	use ladder (32)	adjust sole plate (34)	transom ledger setting (22)	prepare wire (31)	plan bracing setup (14)
7.	set putlog (30)	set transom ledger (22)	toe board setting (82)	install ladder (63)	guardrail setup (43)
8.	set coupler (13)	set mushroom coupler (28)	wire setting (31)	install tread board (19)	base setup (168)
9.	set sole plate (17)	secure tread board by wire (58)			

Luo, X. *et al.* (2019) "Vision-based detection and visualisation of dynamic workspaces," *Automation in Construction*, 104, pp. 1–13. doi: https://doi.org/10.1016/j.autcon.2019.04.001.
Luo, X. *et al.* (2020) "Combining deep features and activity context to improve recognition of activities of workers in groups," *Computer-*