Supplementary

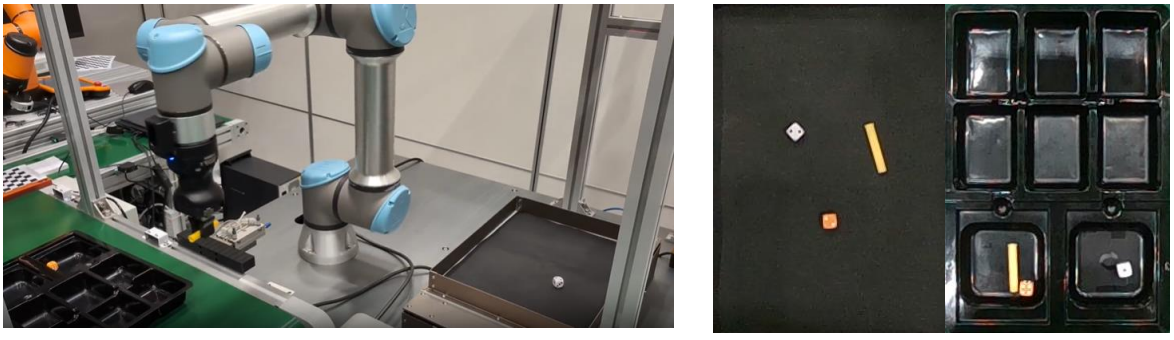
Supp. A Real World Tabletop Manipulation

As the proposed algorithm can increase the model performance with fewer examples, it is further applied to real-world tabletop manipulation. Using natural language instructions and examples selected from the proposed algorithm, a UR5e robot equipped with a gripper performs different picking and placing objects tasks, as depicted in Fig. S1(a). We used Gemini 1.0 Pro for this evaluation.

Two RGBD cameras (RealSense D435i) are mounted above the picking and placing area to guide the robot. For each test, we provide a query to the system that specify objects and grouping criteria (e.g., color, type). The available placing locations are predefined and handled by the API we constructed. We designed this grouping strategy to better replicate product packaging applications, which is achieved with the function `parse_obj_name_in_group()`.

The `put_first_on_second()` function executes the pick-and-place tasks. During execution, the system first analyzes pick area. The object recognition model used in this evaluation is trained via supervised learning. We first remove the background of the image using SegmentAnything (Kirillov *et al.*, 2023a). Then, the feature vectors of each object is extracted by the DINO feature model (Caron, Touvron, Misra, Jégou, *et al.*, 2021). Finally, we train the model on the feature vectors by a linear Support Vector Machine. Using this specific model, the product items can be recognized. Four sets of demo and results are available online[1], with Fig. S1(b) showing an example of separating objects by color.

This LLM approach, compared to a traditional method, offers more intuitive and flexible robot programming. Notably, a reduction in example usage from 6 to approximately 3.71 per instruction showcases an increase in efficiency, accelerates development cycle, and code reusability.



(a) **(b)**

Figure S1: **(a)** The arm executes the generated code **(b)** Input: objects distributed in the picking area (left); Output: sorted objects in the tray according to the given instruction (right).

[1] https://hkflair-f0086.github.io/prompt-selection-augmentation/