

SUPPLEMENTARY MATERIAL FOR THE PAPER "SCENE UNDERSTANDING IN DEFORMABLE OBJECT MANIPULATION VIA TAXONOMY-GUIDED VISION-LANGUAGE MODELS"

Table II reports the performance of all evaluated VLMs across the three DOM tasks, with BERT measuring textual reasoning quality and Hamming distance capturing taxonomy code accuracy.

The figures Fig. 3, Fig. 4, Fig. 5, and Fig. 6, provide illustrative examples of VLM responses guided by T-DOM for the three analyzed DOM tasks, including towel twisting (sliding and twisting actions), meat transport, and cloth edge tracing.

TABLE II: Quantitative Results of Taxonomy-guided VLMs on three DOM tasks

|                                    | BERT $\uparrow$        |                        |                        | HAMMING $\downarrow$   |                        |                        |
|------------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                                    | Gemini2.5Pro           | Qwen-VL-72B            | Qwen-VL-32B            | Gemini2.5Pro           | Qwen-VL-72B            | Qwen-VL-32B            |
| <b>Task 1 - Twisting Towel</b>     |                        |                        |                        |                        |                        |                        |
| Motion                             | <b>0.68</b> $\pm$ 0.24 | 0.49 $\pm$ 0.23        | 0.51 $\pm$ 0.24        | <b>0.10</b> $\pm$ 0.16 | 0.17 $\pm$ 0.17        | 0.32 $\pm$ 0.23        |
| Prehensile Interaction             | <b>0.69</b> $\pm$ 0.09 | 0.57 $\pm$ 0.17        | 0.68 $\pm$ 0.16        | <b>0.13</b> $\pm$ 0.22 | <b>0.13</b> $\pm$ 0.22 | 0.24 $\pm$ 0.19        |
| Non P. Interaction                 | <b>0.76</b> $\pm$ 0.08 | 0.55 $\pm$ 0.14        | 0.67 $\pm$ 0.18        | 0.23 $\pm$ 0.28        | <b>0.17</b> $\pm$ 0.21 | <b>0.17</b> $\pm$ 0.21 |
| Contact Sliding                    | 0.62 $\pm$ 0.14        | <b>0.72</b> $\pm$ 0.30 | 0.59 $\pm$ 0.20        | <b>0.13</b> $\pm$ 0.22 | <b>0.13</b> $\pm$ 0.22 | <b>0.13</b> $\pm$ 0.22 |
| Deformation                        | 0.63 $\pm$ 0.26        | <b>0.77</b> $\pm$ 0.18 | 0.66 $\pm$ 0.17        | 0.20 $\pm$ 0.24        | <b>0.06</b> $\pm$ 0.11 | <b>0.06</b> $\pm$ 0.11 |
| Bending                            | 0.52 $\pm$ 0.18        | <b>0.68</b> $\pm$ 0.25 | 0.63 $\pm$ 0.24        | 0.25 $\pm$ 0.20        | 0.06 $\pm$ 0.11        | <b>0.03</b> $\pm$ 0.08 |
| <b>Task 2 - Transporting Meat</b>  |                        |                        |                        |                        |                        |                        |
| Motion                             | <b>0.65</b> $\pm$ 0.12 | 0.56 $\pm$ 0.20        | 0.59 $\pm$ 0.19        | <b>0.06</b> $\pm$ 0.10 | 0.11 $\pm$ 0.13        | 0.10 $\pm$ 0.12        |
| Prehensile Interaction             | <b>0.59</b> $\pm$ 0.08 | 0.52 $\pm$ 0.18        | 0.48 $\pm$ 0.13        | <b>0.07</b> $\pm$ 0.12 | 0.10 $\pm$ 0.13        | 0.21 $\pm$ 0.18        |
| Non P. Interaction                 | <b>0.61</b> $\pm$ 0.08 | 0.46 $\pm$ 0.07        | 0.51 $\pm$ 0.08        | 0.17 $\pm$ 0.18        | 0.21 $\pm$ 0.14        | 0.24 $\pm$ 0.14        |
| Contact Sliding                    | 0.74 $\pm$ 0.20        | <b>0.81</b> $\pm$ 0.17 | 0.80 $\pm$ 0.15        | 0.11 $\pm$ 0.13        | <b>0.08</b> $\pm$ 0.12 | 0.10 $\pm$ 0.15        |
| Deformation                        | 0.37 $\pm$ 0.22        | <b>0.58</b> $\pm$ 0.37 | 0.45 $\pm$ 0.30        | 0.21 $\pm$ 0.18        | <b>0.17</b> $\pm$ 0.19 | 0.24 $\pm$ 0.16        |
| Bending                            | 0.68 $\pm$ 0.20        | <b>0.77</b> $\pm$ 0.23 | 0.64 $\pm$ 0.30        | <b>0.01</b> $\pm$ 0.06 | 0.06 $\pm$ 0.14        | 0.07 $\pm$ 0.12        |
| <b>Task 3 - Cloth Edge Tracing</b> |                        |                        |                        |                        |                        |                        |
| Motion                             | 0.69 $\pm$ 0.08        | <b>0.70</b> $\pm$ 0.05 | 0.58 $\pm$ 0.06        | 0.37 $\pm$ 0.19        | <b>0.33</b> $\pm$ 0.00 | <b>0.33</b> $\pm$ 0.00 |
| Prehensile Interaction             | 0.65 $\pm$ 0.05        | 0.55 $\pm$ 0.06        | <b>0.72</b> $\pm$ 0.08 | 0.03 $\pm$ 0.08        | <b>0.00</b> $\pm$ 0.00 | 0.13 $\pm$ 0.13        |
| Non P. Interaction                 | 0.69 $\pm$ 0.07        | <b>0.78</b> $\pm$ 0.02 | <b>0.78</b> $\pm$ 0.05 | 0.32 $\pm$ 0.05        | <b>0.17</b> $\pm$ 0.00 | <b>0.17</b> $\pm$ 0.00 |
| Contact Sliding                    | <b>0.65</b> $\pm$ 0.04 | 0.55 $\pm$ 0.05        | 0.54 $\pm$ 0.03        | <b>0.25</b> $\pm$ 0.00 | <b>0.25</b> $\pm$ 0.00 | <b>0.25</b> $\pm$ 0.00 |
| Deformation                        | <b>0.69</b> $\pm$ 0.05 | 0.21 $\pm$ 0.01        | 0.64 $\pm$ 0.23        | <b>0.00</b> $\pm$ 0.00 | 0.25 $\pm$ 0.00        | <b>0.00</b> $\pm$ 0.00 |
| Bending                            | <b>0.65</b> $\pm$ 0.08 | 0.48 $\pm$ 0.10        | 0.64 $\pm$ 0.09        | 0.33 $\pm$ 0.17        | 0.63 $\pm$ 0.13        | <b>0.38</b> $\pm$ 0.13 |

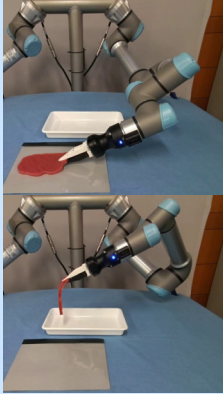
| User Request                                                                                                       | VLM Response                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | Ground Truth                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>[System Prompt for VLM]</p>  | <p><b>Taxonomy Code:</b><br/>100000-0100-000000-0000-0100-0001</p> <p><b>Motion:</b> Left arm: None. Right arm: Dynamic. Reasoning: Right arm lifts object with fast motion. Left arm static. No sliding detected.</p> <p><b>Prehensile:</b> Left arm: None. Right arm: Point grasp. Reasoning: Right gripper visibly constrains object. Left arm not in contact.</p> <p><b>Non-Prehensile:</b> Environment: None. Active Left arm: None. Active Right arm: None. Reasoning: No other contacts visible besides grasp.</p> <p><b>Contact Sliding:</b> Left arm: No Sliding. Right arm: No Sliding. Reasoning: No relative motion at contact points.</p> <p><b>Deformation:</b> Type(s): Tension. Reasoning: Object elongates as it's lifted, showing tension.</p> <p><b>Bending:</b> Structured: Level 0. Unstructured: Level 0. Reasoning: Minor curvature from lifting, no complex folds.</p> | <p><b>Taxonomy Code:</b><br/>010000-0100-000000-0000-1000-0100</p> <p><b>Motion:</b> Left arm: Quasistatic Potential motion upwards and inwards. Right arm: None. Reasoning: The left arm moved upwards to lift the meat and inwards to bring it above the white tray.</p> <p><b>Prehensile:</b> Left arm: Point Grasp. Right arm: No Grasp. Reasoning: the left gripper continue to maintain the grasp on the meat.</p> <p><b>Non-Prehensile:</b> Environment: None. Active Left arm: None. Active Right arm: None. Reasoning: The meat is lifted off from the table using the point grasp.</p> <p><b>Contact Sliding:</b> Left arm: None. Right arm: None. Reasoning: There is no relative motion anywhere.</p> <p><b>Deformation:</b> Compression. Reasoning: The meat is being squeezed by the left gripper.</p> <p><b>Bending:</b> Structured: Level 0. Unstructured: None. Reasoning: The meat is lifted on one end creating a slight bend.</p> |

Fig. 3: Example of a VLM response guided by T-DOM for analyzing the transport object action in the meat transport task.

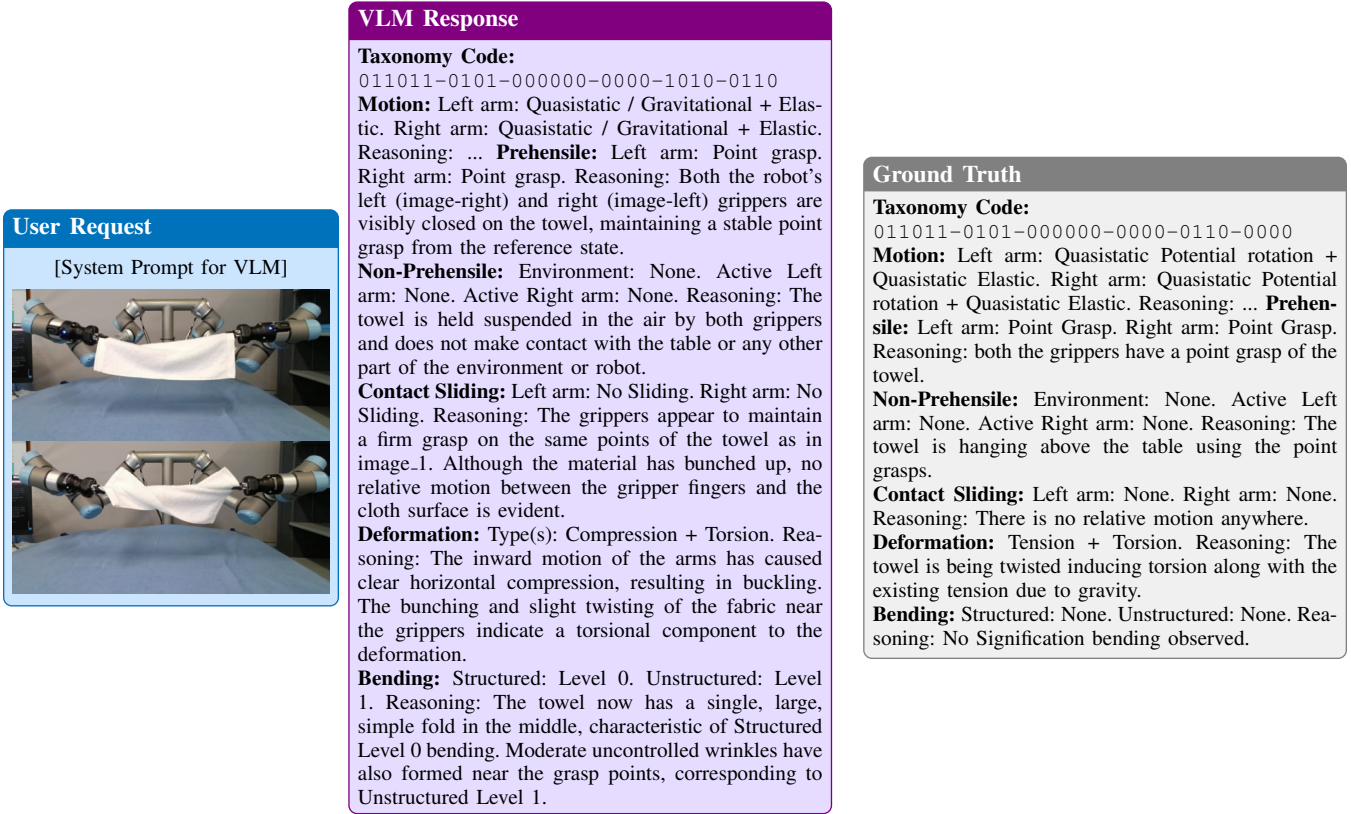


Fig. 4: Example of a VLM response guided by T-DOM for analyzing the twist action in the twist towel task.

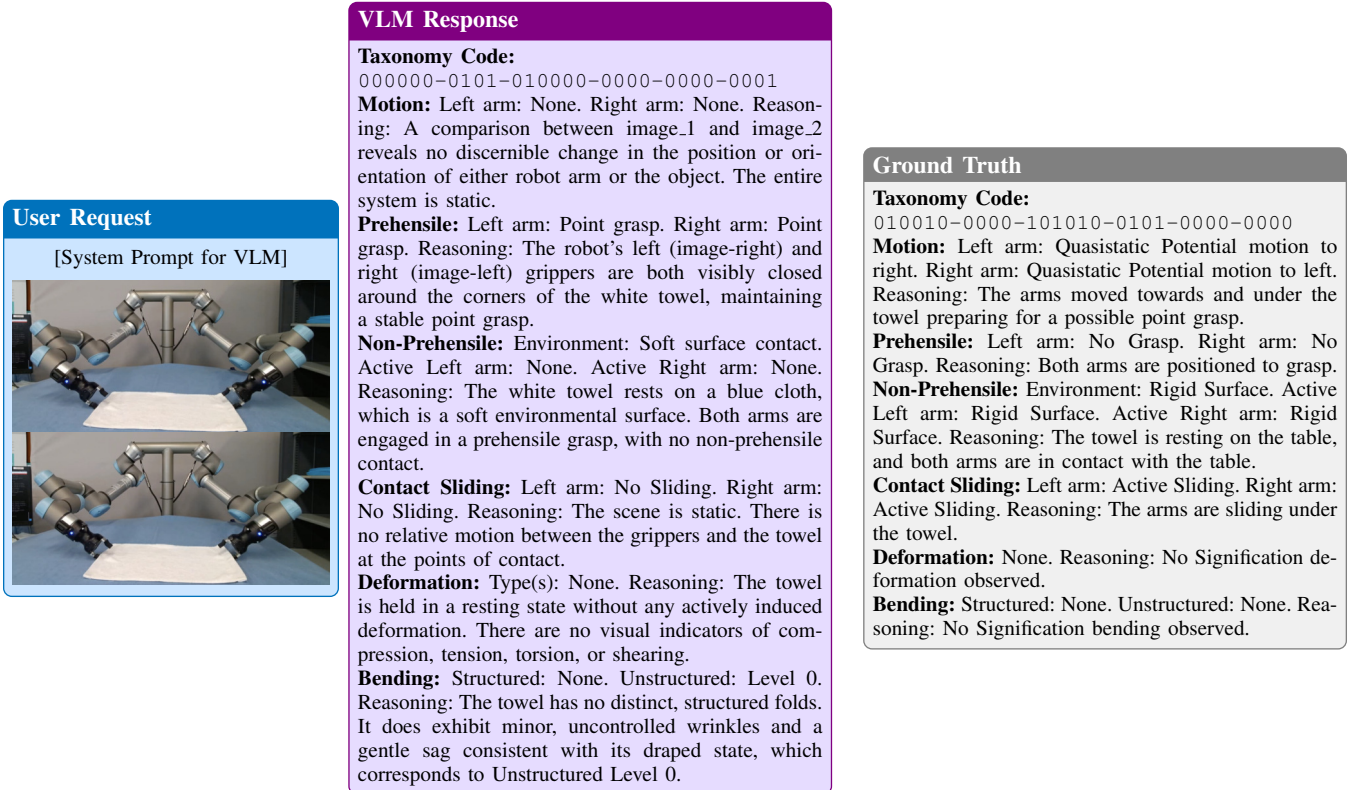


Fig. 5: Example of a VLM response guided by T-DOM for analyzing the sliding action in the towel twisting task.

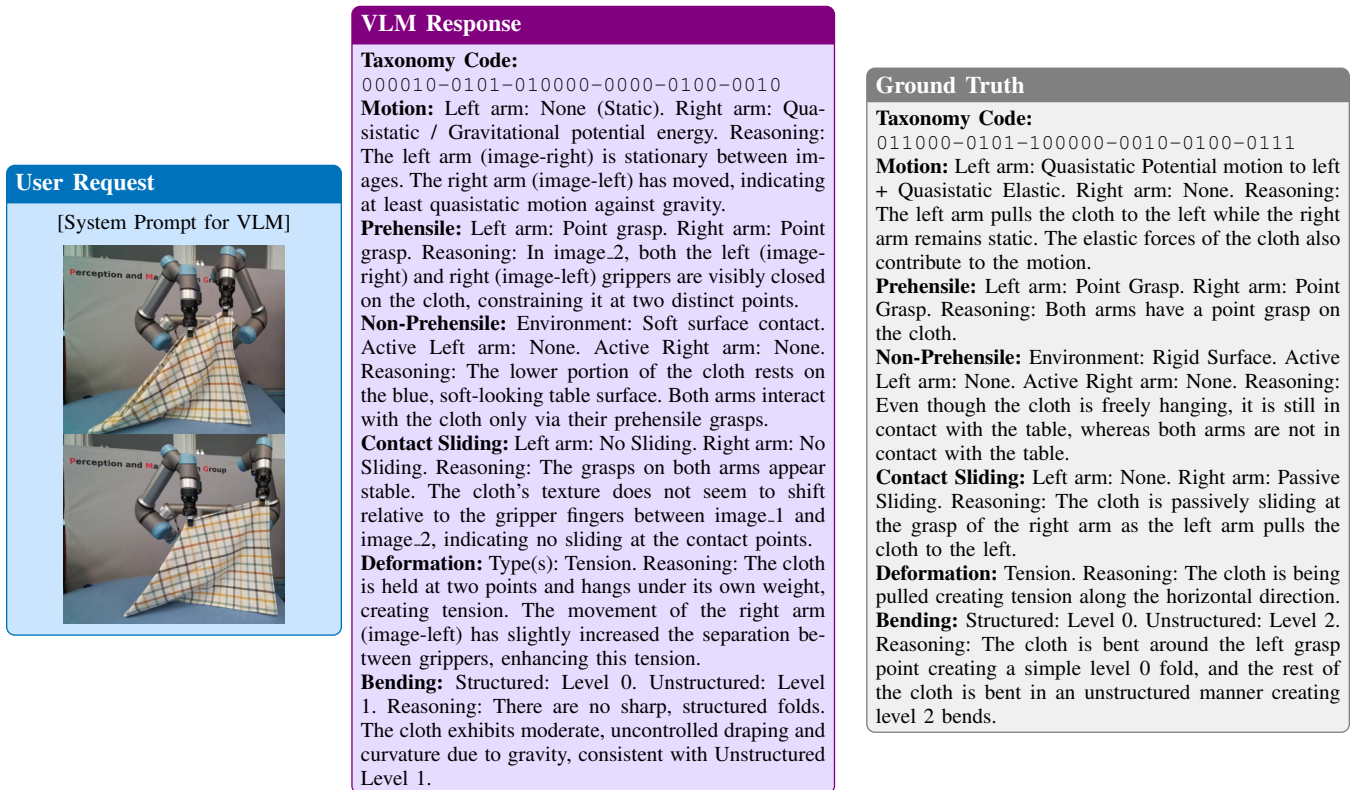


Fig. 6: Example of a VLM response guided by T-DOM for analyzing the action of tracing in the cloth edge tracing task.