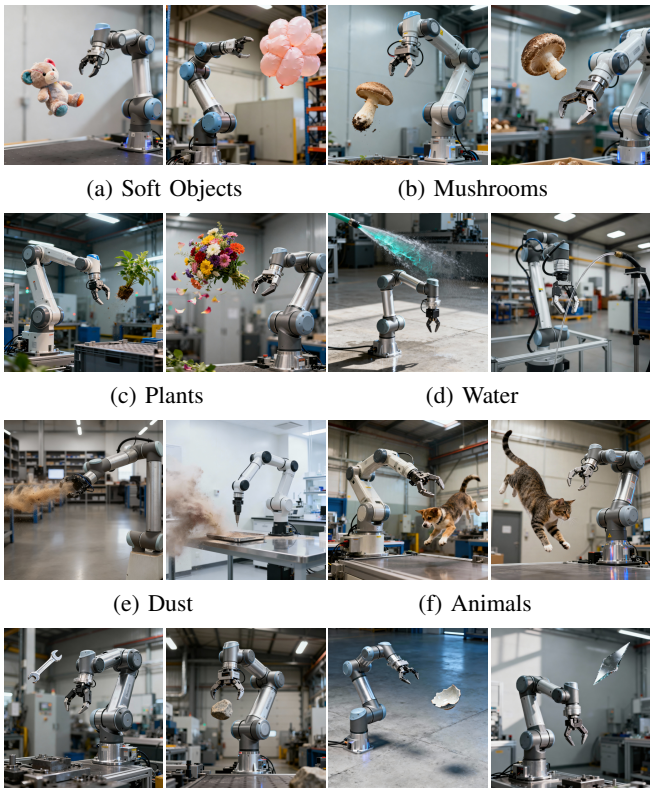
Damage Risk Quantification for Robot Collisions Using Vision-Language Models

Jonas Kiemel[1], Erhan Öztop[2] and Tamim Asfour[1]

***Abstract*— This work investigates the use of Vision-Language Models (VLMs) to estimate the risk of damage from robot-object collisions. Using a curated dataset of 100 images, each depicting a moving object or substance close to collision with a robot, we evaluate how state-of-the-art VLMs quantify the risk of damage to both the robot and the object on a scale from 0 to 10. While numerical outputs vary among models, an analysis across eight object categories shows that VLMs can produce plausible risk quantifications. Our dataset of everyday objects provides reference points for quantified risk values, enabling future VLM applications in damage-aware collision avoidance.**

I. INTRODUCTION

Collision avoidance with moving obstacles is a fundamental challenge in robotics. Estimating collision severity becomes crucial when higher-priority objectives conflict (e.g., cycle-time constraints) or multiple threats converge. For example, an industrial robot performing high-speed part insertion might tolerate light contact with soft packing foam but must avoid a dropped wrench to prevent gripper damage and costly production halts. Estimating damage risks requires advanced reasoning over multiple object properties, such as mass, shape, and material. Substances such as water and dust exemplify cases where more nuanced reasoning is required, as they cause minimal structural harm but may impair sensitive electronic or mechanical components. Vision-language models (VLM) have recently shown impressive reasoning performance across various domains and benchmarks [1]–[3]. This study examines their potential for damage risk quantification using a curated dataset of photorealistic images[3], featuring objects and substances from eight categories, as illustrated in Fig. 1.



(a) Soft Objects (b) Mushrooms (c) Plants (d) Water (e) Dust (f) Animals (g) Bulky Objects (h) Sharp Objects

Fig. 1: We utilize a 100-image dataset with moving objects or substances from 8 categories to evaluate vision-language models (VLMs) on quantifying robot collision damage risks.

II. RELATED WORK

Many existing approaches focus on estimating collision likelihood. Damage severity, the other key component of overall risk, has received less research attention. To quantify injury risks for humans, Haddadin et al. [4] performed crash tests with light-weight manipulators, while Paez-Granados and Billard [5] performed crash tests with service robots and personal mobility devices. Han et al. [6] measured force pain thresholds for safe human-robot interaction. Pose et al. [7] investigated the correlation between collision impact-energy density and human pain intensity. Damage severity metrics incorporating robot stiffness and inertia were employed in in [8] and [9]. This work advances the field by estimating damage severity from scene images using visual reasoning.

III. METHOD

We use VLMs to quantify collision damage risks by providing them with a scene image and prompting them to output a risk value as follows: *Consider the object or substance moving towards the robot. Assume that the robot collides with this object or substance. Estimate the risk of damage to the robot / to the object on a scale from 0 to 10 (0=very low risk of damage, 10=very high risk of damage). Output the number only.* The prompt was kept general to ensure applicability across a wide range of scenes. Since the images provide no explicit velocity information, the risk estimation in this work relies primarily on properties such

This work has been supported by the Japan Society for the Promotion of Science (JSPS) and by the German Federal Ministry of Research, Technology and Space (BMFTR) under the Robotics Institute Germany (RIG).

[1]High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics (IAR), Karlsruhe Institute of Technology (KIT), Germany, jonas.kiemel@kit.edu, asfour@kit.edu

[2]Symbiotic Intelligent Systems Research Center (SISReC), The University of Osaka, Japan erhan.oztop@otri.osaka-u.ac.jp

[3]https://github.com/translearn/damageRisk

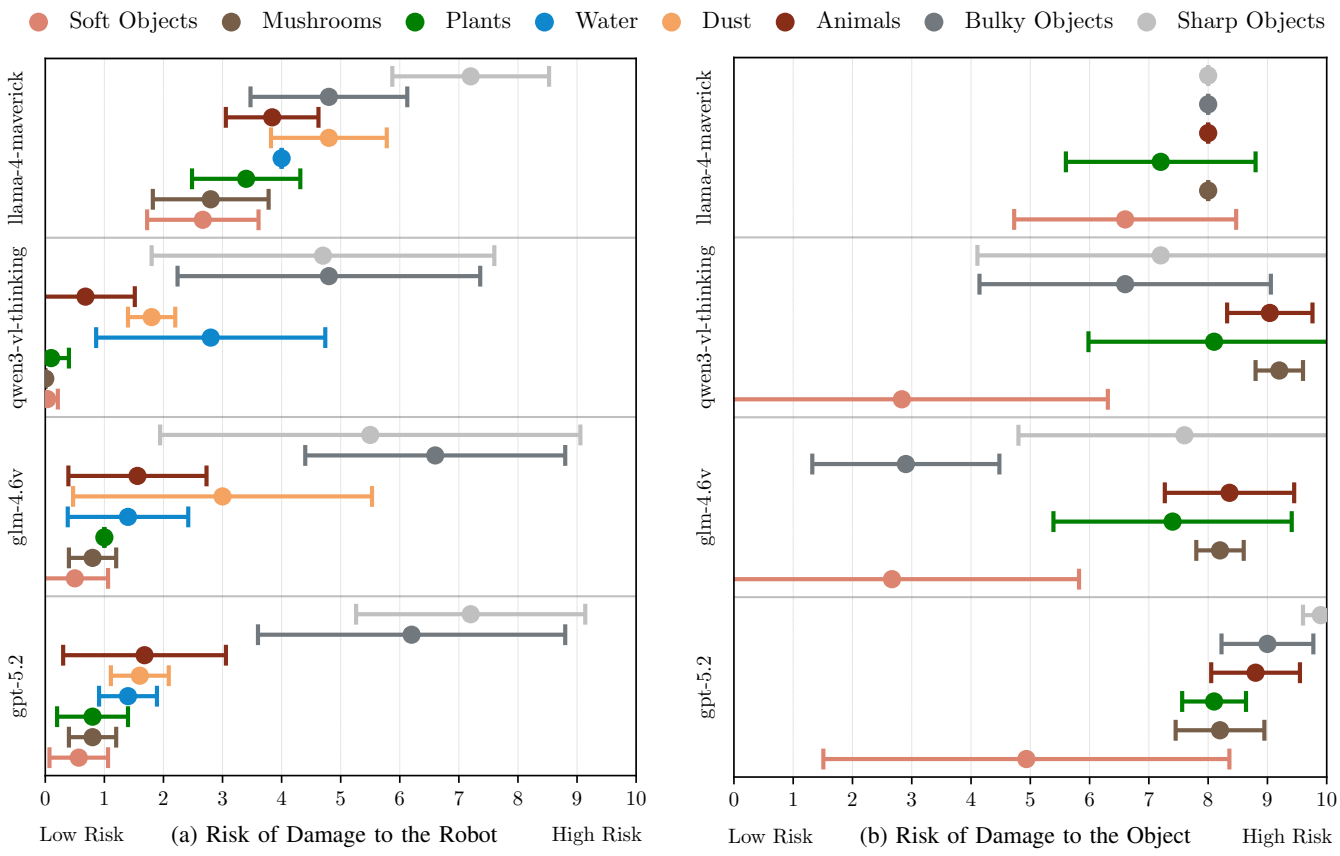


Fig. 2: The results of the damage risk quantification for the robot (a) and object (b) across categories. Dots represent mean values. Error bars indicate ±1 standard deviation. Damage to water and dust omitted for logical reasons.

as the shape, estimated mass, and apparent rigidity of the object and the robot. Relative velocity, if known, can serve as a scaling factor for the estimated risk values, but was not utilized in the current study.

IV. Evaluation

Our evaluation compares the VLMs *Llama-4-Maverick-17B-128E-Instruct* [10], *Qwen3-VL-235B-A22B-Thinking* [11], *GLM-4.6V* [12], [13], and *GPT-5.2* [14] using a curated dataset of photorealistic scene images generated with Seedream 4.0 [15]. All VLMs except GPT-5.2 are open-source.

Fig. 2 shows the resulting risk estimates for robot and object damage across the dataset categories. Dots represent mean values per category, with error bars showing standard deviation. The analysis of damage risks to the robot shown in Fig. 2a yields the following results:

- All models assign the lowest damage risk to soft objects, mushrooms, and plants, as these lack rigidity to cause structural harm to the robot.
- The highest risk values are assigned to bulky and sharp objects that can easily deform or puncture robot components.
- Medium damage risk is assigned to animals, water, and dust. Damage risk assessments for water and dust vary significantly across VLMs. This can be attributed to the demand for more profound contextual reasoning and ambiguities in images, such as uncertainty about the robot's waterproofing or dust resistance.

Fig. 2b shows estimated risk values for damage to the object. The following results can be derived:

- All models assign the lowest damage risk to soft objects. The risk estimates within this category vary considerably, as the dataset includes items such as plush toys, which are unlikely to be damaged, but also soft objects like balloons, which may burst upon contact (see Fig. 1a).
- High damage risks are assigned to animals, plants, and mushrooms, reflecting the fact that living organisms are susceptible to severe harm from collisions with robots.
- The estimated damage risks for bulky and sharp objects vary across models. Unlike the other models, GLM-4.6V assigns low damage risks to bulky objects, which is reasonable since items such as wrenches and stones are sturdy and structurally robust against damage.

V. Conclusion and Future Work

Our analysis showed that VLMs can generate plausible damage risk quantifications solely from scene images. Given their strong visual reasoning capabilities, VLMs are a promising tool for damage severity estimation in robotics. In future work, VLM-generated risk estimates can be integrated into robotic control systems to support damage-aware decision-making and collision avoidance.

REFERENCES

[1] X. Yue, Y. Ni, K. Zhang, T. Zheng, R. Liu, G. Zhang, S. Stevens, D. Jiang, W. Ren, Y. Sun, *et al.*, "MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 9556–9567.

[2] Y. Liu, H. Duan, Y. Zhang, B. Li, S. Zhang, W. Zhao, Y. Yuan, J. Wang, C. He, Z. Liu, *et al.*, "MMBench: Is Your Multi-modal Model an All-Around Player?" in *European Conference on Computer Vision*. Springer, 2024, pp. 216–233.

[3] W. Yu, Z. Yang, L. Li, J. Wang, K. Lin, Z. Liu, X. Wang, and L. Wang, "MM-Vet: Evaluating Large Multimodal Models for Integrated Capabilities," in *Proceedings of the 41st International Conference on Machine Learning*, ser. ICML'24. JMLR.org, 2024.

[4] S. Haddadin, A. Albu-Schäffer, and G. Hirzinger, "Safety Evaluation of Physical Human-Robot Interaction via Crash-Testing," in *Robotics: Science and systems*, vol. 3, 2007, pp. 217–224.

[5] D. Paez-Granados and A. Billard, "Crash test-based assessment of injury risks for adults and children when colliding with personal mobility devices and service robots," *Scientific reports*, vol. 12, no. 1, p. 5285, 2022.

[6] D. Han, M. Y. Park, J. Choi, H. Shin, R. Behrens, and S. Rhim, "Evaluation of force pain thresholds to ensure collision safety in worker-robot collaborative operations," *Frontiers in Robotics and AI*, vol. 11, p. 1374999, 2024.

[7] B. Povse, D. Koritnik, T. Bajd, and M. Munih, "Correlation between impact-energy density and pain intensity during robot-man collision," in *2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*. IEEE, 2010, pp. 179–183.

[8] D. Kulić and E. A. Croft, "Real-time safety for human–robot interaction," *Robotics and Autonomous Systems*, vol. 54, no. 1, pp. 1–12, 2006.

[9] K. Ikuta, H. Ishii, and M. Nokata, "Safety Evaluation Method of Designand Control for Human-Care Robots," *The International Journal of Robotics Research*, vol. 22, no. 5, pp. 281–297, 2003.

[10] Meta AI, "The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation," apr 2025. [Online]. Available: https://ai.meta.com/blog/llama-4-multimodal-intelligence/

[11] S. Bai et al., "Qwen3-VL Technical Report," 2025. [Online]. Available: https://arxiv.org/abs/2511.21631

[12] Zhipu AI, "GLM-4.6V: Open Source Multimodal Models with Native Tool Use," dec 2025. [Online]. Available: https://z.ai/blog/glm-4.6v

[13] GLM-V Team, "Glm-4.5v and glm-4.1v-thinking: Towards versatile multimodal reasoning with scalable reinforcement learning," 2025. [Online]. Available: https://arxiv.org/abs/2507.01006

[14] OpenAI, "Introducing GPT-5.2," dec 2025. [Online]. Available: https://openai.com/index/introducing-gpt-5-2

[15] Seedream Team, Y. Chen, Y. Gao, L. Gong, M. Guo, Q. Guo, Z. Guo, X. Hou, W. Huang, Y. Huang, *et al.*, "Seedream 4.0: Toward Next-generation Multimodal Image Generation," *arXiv preprint arXiv:2509.20427*, 2025.