Towards Efficient Crop Phenotyping using LVLMs

Rukshan Karannagoda[1], Yadan Luo[1], Zijian Wang[1], Kaiyu Guo[1], Scott C Chapman[2], Mahsa Baktashmotlagh[1]

[1]*School of Electrical Engineering and Computer Science, The University of Queensland.*
[2]*School of Agriculture and Food Sustainability, The University of Queensland.*

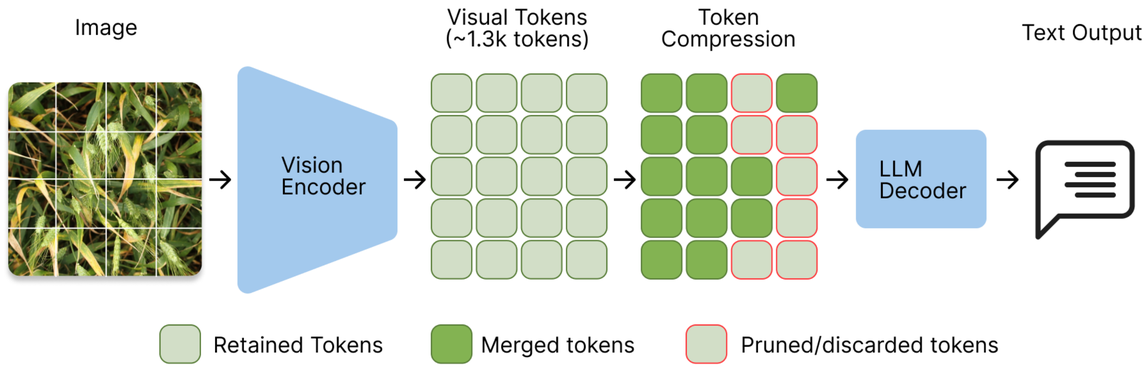
AAGI Analytics for the Australian Grains Industry

ABSTRACT

Large Vision Language Models (LVLMs) face computational challenges when processing high-resolution agricultural imagery. We evaluate training-free token compression strategies combined with tiling and augmentation, demonstrating 2-8 times speedup, 10-20% less memory consumption and increased accuracy across agricultural classification, semantic segmentation, and grounding tasks. Our findings highlight opportunities and limitations for efficient LVLM-based plant phenotyping.

1 Motivation

- Large Vision-Language Models (LVLMs) that can "look at" crop images and describe what they see are powerful tools for measuring plant traits and spotting diseases at scale.
- However they are heavy, slow, and expensive to run, which puts them out of reach for the low-resource settings where they'd be most useful.
- Getting these models to run efficiently, without sacrificing the accuracy is a live challenge for the sector.
- Token Compression removes redundant visual tokens to drive down the cost of inference while preserving model accuracy.

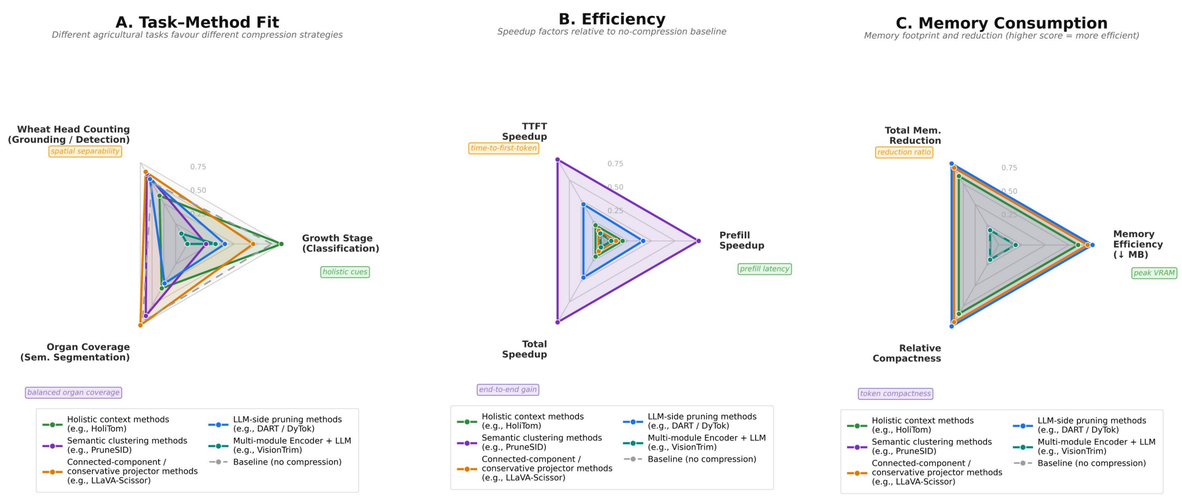


2 Method

We developed training-free techniques to make LVLMs practical for real-world crop phenotyping, focusing on wheat growth stage classification, organ percentage estimation and counting tasks.

- We compared different LVLM architectures paired with visual token compression techniques that strip out redundant information from images so the model has less to process.
- Image Tiling was used to increase counting accuracy.
- Image augmentation was used to find the best techniques that are robust to weather and sensor perturbations.

3 LVLM Benchmarking



We benchmark across 3 quantitative phenotyping benchmarks against Qwen 2.5-VL 7B and 3B models and yield equal or better performance.

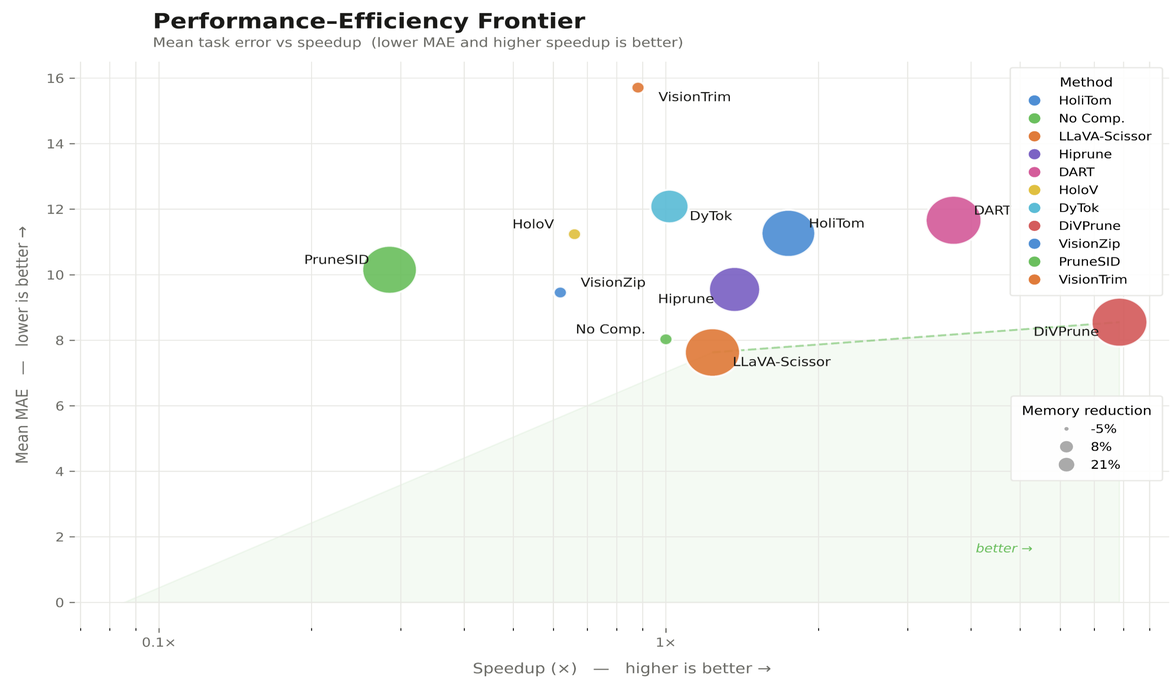
We find that task type (growth stage classification, organ percentage estimation, counting) predicts token compression method fit.

4 Tiling and Augmentation



Fig. 3. Performance of token reduction methods against weather and sensor perturbations: organ coverage estimation vs. counting.

- Image Tiling drastically increases per object token count and the accuracy of counting and localization tasks.
- Semantic clustering-based token reduction methods are resilient against weather and sensor perturbations.



5 Conclusion and Future Work

Conclusion: These results give researchers and downstream developers a clear picture of which LVLM and/or tiling + compression combinations are worth deploying in the field, removing guesswork from the model selection process.

The wheat-specific token compression pipeline is a starting point that other groups can adapt to their own crops and traits.

Future Work: We develop spatially aware token compression, boundary-aware tiling, and adaptive augmentation strategies, and evaluate their generalisability across crops and efficiency on edge hardware.

6 References

[1] Lu, G., Li, S., Mai, G., Sun, J., Zhu, D., Chai, L., Sun, H., Wang, X., Dai, H.,Liu, N., Xu, R., Petti, D., Li, C., Liu, T., Li, C.: Agi for agriculture (2023), https://arxiv.org/abs/2304.06136

[2] Zhao, B., Jin, W., Del Ser, J., Yang, G.: Chatagri: Exploring potentials of chatgpt on cross-linguistic agricultural text classification. Neurocomputing 557, 126708(2023)

[3] Yang, S., Chen, Y., Tian, Z., Wang, C., Li, J., Yu, B., Jia, J.: Visionzip: Longer is better but not necessary in vision language models. In: IEEE/CVF CVPR 2025, Nashville, TN, USA,June (2025).https://doi.org/10.1109/CVPR52734.2025.01843

[4] Yuan, B., Song, S., Fernandez, J., Luo, Y., Baktashmotlagh, M., Wang,Z.: WisWheat: A three-tiered vision-language dataset for wheat management. https://doi.org/10.48550/arXiv.2506.06084

GRDC Grains Research & Development Corporation | Adelaide University | Curtin University | The University of Queensland Australia

The support of the Grains Research and Development Corporation (GRDC) provided through the Analytics for the Australian Grains Industry (AAGI) project is acknowledged.