

Sr. No.	Title	Year	Description	Link
01	LVLm-eHub	2023	An extensive benchmark for assessing large vision-language models (LVLMs) using both quantitative results and user comments. Six essential LVLM capabilities are evaluated quantitatively using 16 tasks including more than 40 text-related visual datasets.	<a href="https://arxiv.org/abs/2306.09265">https://arxiv.org/abs/2306.09265</a>
02	LAMM	2023	An open-source MLLM project with the goal of establishing a complete ecosystem for MLLM training and evaluation, it contains a benchmark and dataset for image and point-cloud instruction tweaking.	<a href="https://arxiv.org/abs/2306.06687">https://arxiv.org/abs/2306.06687</a>
03	What a MESS	2023	A benchmark for cross-domain assessment of semantic segmentation which facilitates thorough performance analysis over a variety of domain-specific datasets, including medical, engineering, and agricultural.	<a href="https://arxiv.org/abs/2306.15521">https://arxiv.org/abs/2306.15521</a>
04	YOLOWeeds	2023	In this study, 25 YOLO models are benchmarked and the CottoWeedDet12 dataset consisting of 5648 photos of 12 weed classes in cotton fields is introduced.	<a href="https://arxiv.org/html/2403.03390v1">https://arxiv.org/html/2403.03390v1</a>
05	FACET	2023	Fairness in Computer Vision Evaluation Benchmark (FACET) is an extensive benchmark dataset comprising 32,000 images which was used for object detection, segmentation, and image classification. Person-related attributes, like skin tone and hair type, were manually annotated by expert reviewers, and fine-grained classes, like disk jockey and guitarist, were labelled.	<a href="https://arxiv.org/abs/2309.00035">https://arxiv.org/abs/2309.00035</a>
06	Battle of the Backbones	2023	BoB analyses more than 1500 training runs to assess pretrained models, such as vision-language and self-supervised learning, on a variety of computer vision tasks. Although supervised CNNs are the best, self-supervised learning has potential, indicating that future work on SSL should concentrate on more complex structures and bigger datasets.	<a href="https://arxiv.org/pdf/2310.19909">https://arxiv.org/pdf/2310.19909</a>
07	Benchmarking Object Detectors with COCO	2024	This study discusses a semi-automatic procedure that is utilized to obtain high-quality mask annotations for the COCO-ReM benchmark. A comparison of mask AP trends between COCO-ReM and COCO-2017 highlights shifts in model ranks and highlights the role that mask quality plays in improving object detectors.	<a href="https://arxiv.org/abs/2403.18819">https://arxiv.org/abs/2403.18819</a>
08	BenchLLM	2023	It compares Large Multimodal Models (LMMs) to different approaches. According to the findings, even if an LMM excels in one style, their performance declines in other kinds. It improves LMMs' thinking to be asked to forecast style before content. Skilful LMMs ought to distinguish mistakes from stylistic differences.	<a href="https://arxiv.org/abs/2312.02896">https://arxiv.org/abs/2312.02896</a>