**Table 1** Comparison of large vision models and traditional vision models.

|  |  |  |
| --- | --- | --- |
| **Feature/Aspect** | **Large vision models** | **Traditional vision models** |
| **Core architecture** | Transformer-based, global self-attention | Convolution-based, local receptive fields |
| **Context & dependencies** | Global context, excels at long-range dependencies | Local focus, struggles with long-range dependencies |
| **Image handling** | Processes image patches; more robust to variations | Uses sliding filters; sensitive to some variations |
| **Data needs** | Best with large-scale pre-training | Can work with smaller datasets, benefits from pre-training |
| **Multimodal ability** | Stronger, more inherent multimodal integration | Requires more specialized designs for multimodal |
| **Parallelism** | High (sequence processing) | Good for convolutions, poor for sequential tasks |
| **Key advantages** | Global understanding, long dependencies, scalability | Efficient local feature extraction |
| **Key limitations** | Can be data-hungry for pre-training | Limited global view, adversarial vulnerability |

**Table 2**

The currently popular and representative large models.

| Original version | Latest version | Release date  (original) | Release date  (latest) | Types  (original → latest) | Information | References |
| --- | --- | --- | --- | --- | --- | --- |
| OPT | / | May 2rd, 2022 | / | LLM | OPT promotes transparency, reproducibility, and broader community engagement and innovation in NLP research. (open source) | (Zhang et al., 2022b) |
| BLOOM | BLOOMZ | July 12th, 2022 | December 15th, 2022 | LLM | A decoder only model based on Transformer architecture. (open source) | (Le Scao et al., 2022) |
| PMC-LLaMA | / | Aprill 27th, 2023 | / | LLM | Inject medical knowledge into existing LLM using 4.8 million biomedical academic papers. (open source) | (Wu et al., 2023a) |
| PaLM2 | / | May 11st, 2023 | / | LLM | PaLM2 was a neural network-based language model that was considered one of the most advanced language models available at the time of its release in May 2023. | (Anil et al., 2023) |
| BloombergGPT | / | March 30th, 2023 | / | LLM | A LLM for the financial field. | (Wu et al., 2023d) |
| OceanGPT-Basic-7B | OceanGPT-Basic-14B/7B/2B | October 3rd, 2023 | July 4th, 2024 | LLM | OceanGPT is the first LLM in the ocean domain. (open source) | (Bi et al., 2023) |
| DeepSeek LLM | DeepSeek-R1 | November 29th, 2023 | January 20th, 2025 | LLM | DeepSeeke-R1 excels in complex tasks such as mathematics, coding, and natural language reasoning | (Guo et al., 2025a) |
| Llama 2 | Llama 4 | July 20th, 2023 | April 6th, 2025 | LLM → MLLM | A series of large models released by Meta. | / |
| Qwen-7B | Qwen2.5-Omni-7B | August 3rd, 2023 | March 27th, 2025 | LLM → MLLM | A super large model launched by Alibaba Cloud. (open source) | (Bai et al., 2023) |
| Kimi Chat | Kimi k1.5 | October 10th, 2023 | January 20th, 2025 | LLM → MLLM | Kimi k1.5 surpasses GPT-4o by 550% in mathematics, coding, and other capabilities under short-chain thinking mode. | / |
| Gemma | Gemma 3 | February 21st, 2024 | March 12th, 2025 | LLM → MLLM | Gemma 3 is a MLLM released by Google. (open source) | (Team et al., 2024; Team et al., 2025) |
| PaLM-E | / | March 6th, 2023 | / | LVM | PaLM-E can integrate vision and language into robot control. | (Driess et al., 2023) |

(*continued on next page*)

**Table 2** (*continued*).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Original version | Latest version | Release date  (original) | Release date  (latest) | Types  (original → latest) | Information | References |
| InternImage | InternImage-H | November 10th, 2022 | October 4th, 2023 | LVM | A LVM based on deformable convolution. (open source) | (Wang et al., 2023) |
| PanguCVLM 3.0 | PanguCVLM 5.0 | July 7th, 2023 | June 21th, 2024 | LVM | A LVM that simulates and automates human visual processes. | / |
| LLaVA | LLaVA-NeXT (Stronger) | April 17th, 2023 | May 10th, 2024 | LVM | LLaVA has the ability to align and fuse the visual information of images with the semantic information of text. (open source) | (Liu et al., 2023b) |
| mPLUG-Owl | mPLUG-Owl3 | April 27th, 2023 | August 20th, 2024 | LVM → MLLM | mPLUG-Owl is developed by Alibaba DAMO Academy. (open source) | (Ye et al., 2023; Ye et al., 2024) |
| SPARK 1.0 | SPARK 4.0 Turbo | May 6th, 2023 | October 24th, 2024 | LVM → MLLM | A new generation of cognitive intelligence model with Chinese as its core. | / |
| Claude 3 | Claude 3.7 Max | March 4th, 2024 | March 18th, 2025 | MLLM | A MLLM that primarily focuses on code processing. | / |
| ERNIE 4.0 | ERNIE 4.5 | October 17th, 2023 | March 16th, 2025 | MLLM | ERNIE is a new generation of Baidu's large model for knowledge enhancement. | / |
| ImageBind | / | May 9th, 2023 | / | MLLM | ImageBind is the first AI model that can bind information from six modes. | (Girdhar et al., 2023) |
| GPT-4 | GPT-4.5 | March 14th, 2023 | February 28th, 2025 | MLLM | GPT-4.5 significantly enhances its knowledge reserves and emotional intelligence by expanding unsupervised learning and reasoning capabilities. | (Bubeck et al., 2023) |
| Skywork | Skywork 4.0 | April 17th, 2023 | January 6th, 2025 | MLLM | Skywork is a series of large models developed by the Kunlun · Skywork team. | (Wei et al., 2023) |
| Gemini | Gemini 2.5 | December 6th, 2023 | March 26th, 2025 | MLLM | Gemini is a MLLM launched by Google DeepMind. | (Team et al., 2023) |
| Sora | / | February 15th, 2024 | / | MLLM | Sora can create realistic and imaginative scenes from text instructions. | (Peebles and Xie, 2023) |
| Hunyuan-t1 | Hunyuan-t1 (official) | February 17th, 2025 | March 21st, 2025 | MLLM | Hunyuan-t1 is a deep-thinking model independently developed by Tencent. | / |

**Table 3** Large models with agricultural potential.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Based | Method | Problem | Application prospect | References |
| LLM | ChatGPT | GPT-3.5-turbo | Agricultural information extraction | Rapid querying of agricultural information | (Peng et al., 2023) |
| AugGPT | Text data augmentation | Few-shot learning for agricultural data | (Dai et al., 2023) |
| / | Aurora | Atmospheric prediction | Predicting weather in agriculture | (Bodnar et al., 2024) |
| LVM | / | MAE, DINO, DINOv2 | Plant phenotyping tasks | Monitoring crop health | (Chen et al., 2023a) |
| PaLM, ViT | PaLM-E | Robot control | Agricultural intelligent machines or robots | (Driess et al., 2023) |
| MLLM | SAM | TAM | Object tracking and segmentation in videos | Monitor animals in agricultural farming | (Yang et al., 2023a) |

**Table 4** Agricultural large models.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Name | Achievement | Significance | References |
| LLM | TimeGPT | Predicting soil moisture | Contributes to sustainable agricultural practices | (Deforce et al., 2024) |
| ChatGPT | Designed a tomato-picking robot | Simplify the design process of agricultural robots | (Stella et al., 2023) |
| FMFruit | Identifying multiple types of fruits | Research on robotic harvesting and fruit detection | (Li et al., 2024) |
| AgriGPT | Multimodal agricultural knowledge Q&A | Promote precision agriculture practices | (Liu et al., 2023a) |
| ShenNong | Development of specialized large models for multiple agricultural domains | Driving agricultural intelligence and comprehensive efficiency improvement | / |
| ChatAgri | Cross-linguistic classification of agricultural texts | Provide decision support for precision agriculture | (Zhao et al., 2023a) |
| LVM | SpectralGPT | Process spectral remote sensing data | Greatly enhanced the processing capability of agricultural spectral data | (Hong et al., 2024) |
| SAM | Chicken segmentation and tracking | Facilitates segmentation and tracking tasks in agriculture | (Yang et al., 2023c) |
| Agricultural field boundary delineation | Beneficial for PA, crop monitoring, and yield estimation | (Tripathy et al., 2024) |
| MLLM | ITLMLP | Cucumber disease recognition | Agricultural disease recognition | (Cao et al., 2023) |
| AIE-SEG | High-precision segmentation of agricultural imagery | Enables automated field monitoring and yield estimation​ | (Xu et al., 2023) |