

Large Vision-Language Models

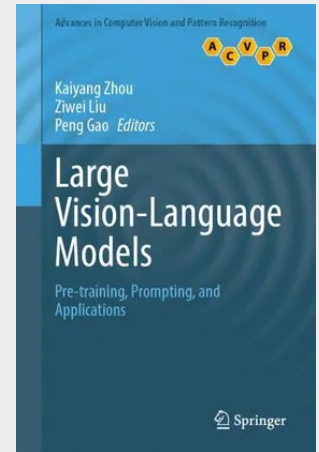
Pre-training, Prompting, and Applications

The rapid progress in the field of large multimodal foundation models, especially vision-language models, has dramatically transformed the landscape of machine learning, computer vision, and natural language processing. These powerful models, trained on vast amounts of multimodal data mixed with images and text, have demonstrated remarkable capabilities in tasks ranging from image classification and object detection to visual content generation and question answering. This book provides a comprehensive and up-to-date exploration of large vision-language models, covering the key aspects of their pre-training, prompting techniques, and diverse real-world computer vision applications. It is an essential resource for researchers, practitioners, and students in the fields of computer vision, natural language processing, and artificial intelligence.

begins by exploring the fundamentals of large vision-language models, covering architectural designs, training techniques, and dataset construction methods. It then examines prompting strategies and other adaptation methods, demonstrating how these models can be effectively fine-tuned to address a wide range of downstream tasks. The final section focuses on the application of vision-language models across various domains, including open-vocabulary object detection, 3D point cloud processing, and text-driven visual content generation and manipulation.

Beyond the technical foundations, the book explores the wide-ranging applications of vision-language models (VLMs), from enhancing image recognition systems to enabling sophisticated visual content generation and facilitating more natural human-machine interactions. It also addresses key challenges in the field, such as feature alignment, scalability, data requirements, and evaluation metrics. By providing a comprehensive roadmap for both newcomers and experts, this book serves as a valuable resource for understanding the current landscape, limitations, and future directions of VLMs, ultimately contributing to the advancement of artificial intelligence.

The rapid progress in the field of large multimodal foundation models, especially vision-language models, has dramatically transformed the landscape of machine learning, computer vision, and natural language processing. These powerful models, trained on vast amounts of multimodal data mixed with images and text, have demonstrated remarkable capabilities in tasks ranging from image classification and object detection to visual content generation and question answering. This book provides a comprehensive and up-to-date exploration of large vision-language models, covering the key aspects of their pre-training, prompting techniques, and diverse real-world computer vision applications. It is an essential resource for researchers, practitioners, and students in the fields of computer vision, natural language processing, and artificial intelligence. Large Vision-Language Models begins by exploring the fundamentals of large vision-language models, covering architectural designs, training techniques, and dataset construction methods. It then examines prompting strategies and other adaptation methods, demonstrating how these models can be effectively fine-tuned to address a wide range of downstream tasks. The final section focuses on the application of vision-language models across various domains, including open-vocabulary object detection, 3D point cloud processing, and text-driven visual content generation and manipulation. Beyond the technical foundations, the book explores the wide-ranging applications of vision-language models (VLMs), from enhancing image recognition systems to enabling sophisticated visual content generation and facilitating more natural human-machine interactions. It also addresses key challenges in the field, such as feature alignment, scalability, data requirements, and evaluation metrics. By providing a comprehensive roadmap for both newcomers and experts, this book serves as a valuable resource for understanding the current landscape, limitations, and future directions of VLMs, ultimately contributing to the advancement of artificial intelligence.



192,59 €

179,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9783031949685

Medium: Buch

ISBN: 978-3-031-94968-5

Verlag: Springer

Erscheinungstermin: 31.08.2025

Sprache(n): Englisch

Auflage: Erscheinungsjahr 2025

Serie: Advances in Computer Vision and Pattern Recognition

Produktform: Gebunden

Gewicht: 832 g

Seiten: 429

Format (B x H): 160 x 241 mm

