

BO WAN

Email • Github • Google Scholar • Homepage

EDUCATION

KU Leuven

Ph.D. candidate in Computer Engineering, supervised by Prof. Tinne Tuytelaars

Leuven, Belgium

Oct. 2020 - Present

ShanghaiTech University

M.S. in Computer Science, supervised by Prof. Xuming He

Shanghai, China

Sep. 2017 - Aug. 2020

Peiking University

B.S. in Economics (Double Degree)

Beijing, China

Sep. 2015 - Jun. 2017

Beijing University of Posts and Telecommunications

B.S. in Telecommunication Engineering

Beijing, China

Sep. 2013 - Jun. 2017

INTERN EXPERIENCE

Google DeepMind

Student researcher intern, work on location-aware vision-language pre-training.

Zurich, Swiss

Jul. 2023 - Dec. 2023

Google Brain

Student researcher intern, work on vision-language multitask learning.

Zurich, Swiss

Oct. 2022 - Feb. 2023

Tencent AI Lab

Research intern, work on vision and language understanding in videos.

Shenzhen, China

Oct. 2020 - Dec. 2020

Microsoft

Algorithm intern, explore algorithms for Bing search Question Answering system in Chinese.

Beijing, China

Apr. 2017 - Jul. 2017

RESEARCH INTEREST

Vision-Language Pretraining is a fundamental building block for multimodal large language models (MLLM). During my internship at Google DeepMind, I developed a location-aware VL pretraining scheme that significantly enhanced the model's ability to understand fine-grained regional details. Additionally, my other research explored the unsupervised construction of a unified vision and language grammar structure, advancing the integration of multimodal alignments. This work was recognized as ICLR Oral Presentation.

LLM/MLLM are essential to revolutionary AI products like and ChatGPT. During my internship at Google Brain, I investigated the deployment of a unified language decoder for VL multitasking, with a pretrained and locked image encoder. We experimented with training a small Transformer decoder from scratch and using a pretrained LLM (T5), which led to many insightful findings.

Visual Generation is a critical capability for ChatGPT alongside text generation. My previous research focused on the development of controllable video generation with diffusion models, which involved animating images in accordance with customized object motions to align with user expectations.

PUBLICATIONS

LocCa: Visual Pretraining with Location-aware Captioners

B. Wan, M. Tschannen, Y. Xian, F. Pavetic, I. Alabdulmohsin, X. Wang, A.S. Pinto, A. Steiner, L. Beyler, X. Zhai

Arxiv 2024

Animate Your Motion: Turning Still Images into Dynamic Videos

M. Li, B. Wan*, S. Moens, T. Tuytelaars*

Arxiv 2024

UniPT: Universal Parallel Tuning for Transfer Learning with Efficient Parameter and Memory

H. Diao, B. Wan, Y. Zhang, X. Jia, H. Lu, L. Chen

CVPR 2024

Exploiting CLIP for Zero-shot HOI Detection Requires Knowledge Distillation at Multiple Levels

B. Wan, T. Tuytelaars

WACV 2024

Weakly-supervised HOI Detection via Prior-guided Bi-level Representation Learning

B. Wan, Y. Liu*, D. Zhou, T. Tuytelaars, X. He*

ICLR 2023

A Study of Autoregressive Decoders for Multi-Tasking in Computer Vision

L. Beyer*, **B. Wan***, G. Madan*, F. Pavetic*, A. Steiner*, A. Kolesnikov,
A.S. Pinto, E. Bugliarello, X. Wang, Q. Yu, L. Chen, X. Zhai*

Arxiv 2023

Unsupervised Vision-Language Grammar Induction with Shared Structure Modeling

B. Wan, W. Han, Z. Zheng, T. Tuytelaars

(Oral) ICLR 2022

Relation-aware Instance Refinement for Weakly Supervised Visual Grounding

Y. Liu*, **B. Wan***, L. Ma, X. He

CVPR 2021

Bipartite Graph Network with Adaptive Message Passing for Unbiased Scene Graph Generation

R. Li, S. Zhang, **B. Wan**, X. He

CVPR 2021

Single Image 3D Object Estimation with Primitive Graph Networks

Q. He, D. Zhou, **B. Wan**, X. He

ACMMM 2021

Learning Cross-modal Context Graph Networks for Visual Grounding

Y. Liu*, **B. Wan***, X. Zhu, X. He

AAAI 2020

Pose-aware Multi-level Feature Network for Human Object Interaction Detection

B. Wan*, D. Zhou*, Y. Liu, R. Li, X. He

(Oral) ICCV 2019

* indicates equal contribution. Please c.f. Google Scholar for full publications.

AWARDS AND HONORS

- | | | |
|--|--------------------|-----------|
| • Mathematical Contest In Modelling | Meritorious Winner | Apr. 2016 |
| • The Chinese Mathematics Competitions | Second Prize | Aug. 2015 |

ACADEMIC SERVICE

Reviewer for TPAMI, ICLR, NeurIPS, ICML, CVPR, ICCV, ECCV, WACV, ICIP