Zhipeng Bao

(+1)917-318-5651 | zbao@cs.cmu.edu | https://zpbao.github.io

EDUCATION

Carnegie Mellon University — Pittsburgh, PA

PhD in Robotics — *Aug. 2022 – Aug. 2025 (Expected)*

- **Advisor** Prof. Martial Hebert
- **Thesis** *Bridging Generative and Discriminative Learning with Diffusion Models*
- **Committee** Martial Hebert, Deva Ramanan, Jun-Yan Zhu, Alexei Efros (UCB), Yu-Xiong Wang (UIUC), Pavel Tokmakov (TRI)

Master of Science in Robotics — GPA:4.18/4.0 — *Aug. 2019 – Aug. 2021*

- **Advisor** Prof. Martial Hebert

Tsinghua University — Beijing, China

Bachelor of Electronic Engineering — GPA:3.53/4.00 — *Aug. 2015 – July 2019*

- **Honor** Comprehensive Excellence Award of Tsinghua University (2018)

Australian National University — Canberra, Australia

Exchange Program — GPA:6.33/7.00 — *Feb. 2018 – June 2018*

RESEARCH EXPERIENCE

Visual Perception with Generative Models — Aug 2019 – Now

Carnegie Mellon University. Advisor: Prof. Martial Hebert — *Pittsburgh, PA*

- Video understanding with diffusion models (VDM-Understanding, under review).
- Repurpose Text-to-Video diffusion models for referring video segmentation (REM, under review).
- A unified diffusion-based model for joint multi-modal generation and dense perception (Diff-2-in-1, under review).
- A comprehensive study that probes visual encoders for 3D scene understanding (Lexicon3D, NeurIPS'24).
- Multi-task view synthesis with Neural Randiance Fields (MuvieNeRF, ICCV'23; SS-NeRF, WACV'23).
- Use GANs to facilitate multi-task visual learning (MGM, ICML'22).
- A generative system for joint view synthesis and recognition (Bowtie, ICLR'21).

Learning Compositional Representations with Minimal Supervisions — Aug 2021 – Sep 2023

Carnegie Mellon University. Advisor: Prof. Martial Hebert — *Pittsburgh, PA*

- Compositional fine-tuning with Text-to-Image diffusion models (SepEn, SIGGRAPH'24).
- Object discovery from motion-guided tokens (MoTok, CVPR'23).
- Motion-guided object-centric learning (DoM, CVPR'22).

WORKING EXPERIENCE

GenAI, Meta — Menlo Park, CA

Research Intern, Mentor: Dr. Xiaofang Wang — *May 2024 – Aug 2024*

- Designed a novel self-distillation objective to boost MLLMs with stronger visual tokens.
- Submitted a paper to CVPR 2025 (VLM-recognition, under review).

Adobe Research — Seattle, WA

Research Intern, Mentor: Dr. Yijun Li — *May 2023 – Aug 2023*

- Proposed a compositional fine-tuning algorithm for Text-to-Image diffusion models achieving state-of-the-art performance on compositional, meanwhile expressing promising generation capacity after large-scale training.
- Filled a patent and authored a paper (SepEn, SIGGRAPH'24).

Toyota Research Institute — Los Altos, CA

Research Intern, Machine Learning Research Group. Mentor: Dr. Pavel Tokmakov — *June 2021 – Sep. 2022*

- Scaled the recent frameworks for object discovery from toy, synthetic images to complex, real-world videos.
- Filled two patents and authored two papers (DoM, CVPR'22; MoTok, CVPR'23).

DATA 61, CSIRO — Canberra, Australia

Research Intern, Computer Vision research group. Mentor: Dr. Shaodi You — *Feb. 2018 – Sep. 2018*

- Proposed a novel triple-channel model for single image-based facial expression recognition (FER).
- Authored a paper (Facial3D, ICCVW'19).

Peer-reviewed and In-submission Articles

Boosting Multimodal LLMs via Visual Token Supervision, Under Review [pdf]
Zhipeng Bao, Miao Liu, Ankit Ramchandani, Mengjiao Wang, Felix Juefei-Xu, Xide Xia, Tong Xiao, Yu-Xiong Wang, Martial Hebert, Ning Zhang, and Xiaofang Wang

Video Diffusion Models Learn the Structure of the Dynamic World, Under Review [pdf]
Zhipeng Bao, Anurag Bagchi, Yu-Xiong Wang, Pavel Tokmakov, and Martial Hebert

ReferEverything: Towards Segmenting Everything We Can Speak of in Videos, Under Review [pdf]
Anurag Bagchi, Zhipeng Bao, Yu-Xiong Wang, Pavel Tokmakov, and Martial Hebert

Diff-2-in-1: Bridging Generation and Dense Perception with Diffusion Models, Under Review [pdf]
Shuhong Zheng, Zhipeng Bao, Martial Hebert, and Yu-Xiong Wang

Lexicon3D: Probing Visual Foundation Models for Complex 3D Scene Understanding, NeurIPS 2024 [pdf]
Yunze Man, Shuhong Zheng, Zhipeng Bao, Martial Hebert, Liangyan Gui, and Yu-Xiong Wang

Separate-and-Enhance: Compositional Finetuning for Text-to-Image Diffusion Models, SIGGRAPH 2024 [pdf]
Zhipeng Bao, Yijun Li, Krishna Kumar Singh, Yu-Xiong Wang, and Martial Hebert

Multi-task View Synthesis with Neural Radiance Fields, ICCV 2023 [pdf]
Shuhong Zheng*, Zhipeng Bao*, Martial Hebert, and Yu-Xiong Wang

Object Discovery from Motion-guided Tokens, CVPR 2023 [pdf]
Zhipeng Bao, Pavel Tokmakov, Yu-Xiong Wang, Adrien Gaidon, and Martial Hebert

Beyond RGB: Scene-Property Synthesis with Neural Radiance Fields, WACV 2023 [pdf]
Mingtong Zhang, Shuhong Zheng, Zhipeng Bao, Martial Hebert, and Yu-Xiong Wang

Generative Modeling for Multi-task Visual Learning, ICML 2022 [pdf]
Zhipeng Bao, Martial Hebert, and Yu-Xiong Wang

Discovering Objects that Can Move, CVPR 2022 [pdf]
Zhipeng Bao, Pavel Tokmakov, Allan Jabri, Yu-Xiong Wang, Adrien Gaidon, and Martial Hebert

Bowtie Networks: Generative Modeling for Joint Few-shot Recognition and Novel-View Synthesis, ICLR 2021 [pdf]
Zhipeng Bao, Yuxiong Wang, and Martial Hebert

Single-Image Facial Expression Recognition Using Deep 3D Re-Centralization, ICCVW 2019 [pdf]
Zhipeng Bao, Shaodi You, Lin Gu, and Zhenglu Yang

A Joint Method for Marker-Free Alignment of Tilt Series in Electron Tomography, ISMB 2019 [pdf]
Renmin Han, Zhipeng Bao, Xiangrui Zeng, Tongxin Niu, Fa Zhang, Min Xu, and Xin Gao

Patents

Compositional finetuning with T2I diffusion models. 2024
Zhipeng Bao, Yijun Li, Krishna Kumar Singh

Object detection based on motion-guided tokens. 2024 [link]
Zhipeng Bao, Pavel Tokmakov, Yuxiong Wang, Adrien David Gaidon, Martial Hebert

Self-supervised compositional feature representation for video understanding. 2023 [link]
Zhipeng Bao, Pavel Tokmakov, Allan Jabri, Yuxiong Wang, Adrien David Gaidon, Martial Hebert

Services

Reviewer: CVPR, ICCV, ECCV, NeurIPS, ICLR, ICML

Technical Skills

Languages: Python, MATLAB, Java, C/C++, HTML, R
Tools & Frameworks: Git, Pytorch, Tensorflow, Latex, DeepSpeed, Torchrun, SageMaker, SLURM, Jupyter Notebook