

# Binjie Zhang

Google Scholar

Email: [zbj19@mails.tsinghua.edu.cn](mailto:zbj19@mails.tsinghua.edu.cn)

Mobile: +86 15216801797

## EDUCATION

- **Tsinghua University (THU)** China  
• *Master of Computer Science and Technology* Aug 2019 - Jul 2022  
Supervisor: *Prof. Chun Yuan*  
Research Interests: *Compatible Representation Learning; Cross-modality video understanding*
- **East China University of Science and Technology (ECUST)** China  
• *Bachelor of Information Engineering* Aug 2015 - Jul 2019  
Cumulative GPA: 3.77 / 4.00, Ranking: 1 / 92

## PUBLICATIONS

- [1] **B. Zhang**, Y. Ge, Y. Shen, Y. Li, C. Yuan, X. Xu, Y. Wang, and Y. Shan. "Hot-Refresh Model Upgrades with Regression-Alleviating Compatible Training in Image Retrieval." International Conference on Learning Representations (ICLR). 2022.
- [2] **B. Zhang**, Y. Ge, Y. Shen, S. Su, F. Wu, C. Yuan, X. Xu, Y. Wang, and Y. Shan. "Towards Universal Backward-Compatible Representation Learning." The 31st International Joint Conference on Artificial Intelligence (IJCAI). 2022. (long oral)
- [3] S. Su\*, **B. Zhang\***, Y. Ge, X. Xu, Y. Wang, C. Yuan, and Y. Shan. "Privacy-Preserving Model Upgrades with Bidirectional Compatible Training in Image Retrieval." European Conference on Computer Vision (ECCV). 2022. (\* co-first authors, under review)
- [4] **B. Zhang**, Y. Li, C. Yuan, D. Xu, X. Zhao, and Y. Shan. "Cascade Cross-Modality Alignment Network for Video Temporal Grounding." IEEE Transactions on Multimedia (TMM). (under review)

## WORK EXPERIENCE

- **Intern in Tencent ARC Lab** Shenzhen, China  
• *Work on compatible representation learning, supervised by Dr. Yixiao Ge and Dr. Yantao Shen.* 2020 - Present
- **Intern in Tencent ARC Lab** Shenzhen, China  
• *Focus on cross-modality video understanding, supervised by Dr. Yu Li and Dr. Ying Shan.* 2019 - 2020

## HONORS AND AWARDS

- **Annual College Personage Award** (Highest student honor in ECUST) Mar, 2018
- **National Scholarship for Undergraduates (twice)** (Ministry of Education of China) Nov 2017 & Nov 2016
- Top Level Comprehensive Curriculum Scholarship Oct 2017 & Oct 2016
- Arawana Scholarship (Arawana charity foundation of China sponsored) Oct 2017
- Social Work Award & Outstanding Student & Excellent Class Cadres Sep 2017 & Sep 2016
- The Interdisciplinary Contest in Modeling (ICM) - Meritorious Winner(9%) Feb 2018
- International Genetically Engineered Machine Competition (Gold Award) Nov 2017
- China Computer Programming and Design Competition (Championship) Aug 2016

## PROJECTS

- **Hot-Refresh Model Upgrades with Regression-Free Compatible Training in Image Retrieval.**<sup>[1]</sup>: We for the first time study the model regression problem in hot-refresh model upgrades of image retrieval systems with compatible representation learning. In addition, we introduce a simple yet effective training regularization and a uncertainty based backfilling strategy to alleviate model regression.
- **Towards Universal Backward-Compatible Representation Learning.**<sup>[2]</sup>: Previous works of compatible learning only investigated a close-set training scenario, and were limited by more realistic and challenging open-set scenes. We first explore all possible compatible training settings, and further propose a universal backward-compatible training paradigm with a novel structural prototype refinement algorithm.
- **Privacy-Preserving Model Upgrades with Bidirectional Compatible Training in Image Retrieval.**<sup>[3]</sup>: The task of privacy-preserving model upgrades in image retrieval desires to reap the benefits of rapidly evolving new models without accessing the raw gallery images. Although pioneering works introduced backward compatible training as a possible solution, its improvement in sequential model upgrades is gradually limited by the fixed old gallery embeddings. To this end, we propose a new model upgrade paradigm, termed Bidirectional Compatible Training.
- **Cascade Cross-Modality Alignment Network for Video Temporal Grounding.**<sup>[4]</sup>: We propose an intuitive attention-based two-stage structure named Cascade Cross-Modal Alignment Network, which incorporates global information of text and video in the coarse stage, and captures local correlation in the fine stage. In addition, we introduce several task-specific losses to explicitly align multi-modal embedding spaces for the first time.