

Department of Computer Science,
Ira A. Fulton Schools of Engineering,
Arizona State University,
699 S Mill Ave, Tempe 85281.

(480) 207-0035 ☎
@Jacob292020 🐦
zy.fang@asu.edu ✉
<https://www.public.asu.edu/~zfang29/> 🌐

(Jacob) Zhiyuan Fang

- EDUCATION** Ph.D. Candidate, [Arizona State University](#), 2017 - 2022. GPA: 4.0/4.0
Computer Science, *School of Computing, Informatics, and Decision Systems Engineering*.
- Visiting Student, [University of California, Irvine](#), June - Sep, 2016.
School of Information and Computer Sciences
- B.E., [Southern University of Science and Technology](#), 2013 - 2017. GPA: 3.6/4.0
Computer Engineering, *College of Engineering*

RESEARCH INTERESTS

-
- *Vision and Language, Representation Learning* under *weak/un-supervised* learning schema.
 - *Image/video Captioning, Textual Grounding* and etc.
 - *Efficient VL learning*, including building lightweight, data-efficient and real-world applicable Vision and Language models.

* indicates equal contribution.

PUBLICATIONS

- “*Injecting Semantic Concepts into End-to-End Image Captioning.*” **Zhiyuan Fang**, Jianfeng Wang, Xiaowei Hu, Lin Liang, Zhe Gan, Lijuan Wang, Yezhou Yang, Zicheng Liu, Nov, 2021 (Under Review)
- “*Compression Visual-linguistic Model via Knowledge Distillation.*” **Zhiyuan Fang**, Jianfeng Wang, Xiaowei Hu, Lijuan Wang, Yezhou Yang, Zicheng Liu. (International Conference on Computer Vision (ICCV), 2021)
- “*SEED: Self-Supervised Distillation for Visual Representation.*” **Zhiyuan Fang**, Jianfeng Wang, Lijuan Wang, Lei Zhang, Yezhou Yang, Zicheng Liu, (International Conference on Representation Learning (ICLR), May, 2021)
- “*Video2commonsense (v2c): Learning to transform video scenes to commonsense knowledge.*” **Zhiyuan Fang**, Tejas Gokhale, Chitta Baral, and Yezhou Yang, (*Empirical Methods in Natural Language Processing (EMNLP)*), Long Paper, 2020)
- “*ViTAA: Visual-Textual Attributes Alignment in Person Search by Natural Language.*” Zhe Wang*, **Zhiyuan Fang***, Jun Wang and Yezhou Yang, (*European Conference on Computer Vision (ECCV)*, 2020)

- “Temporal Language Grounding with Referring Attention and Weak Supervision.”* **Zhiyuan Fang**, Shu Kong, Zhe Wang, Charless Fowlkes, and Yezhou Yang, April, 2020
- “Modularized Textual Grounding for Counterfactual Resilience.”* **Zhiyuan Fang**, Shu Kong, Charless Fowlkes, Yezhou Yang, (Conference on Computer Vision and Pattern Recognition (CVPR), 2019)
- “Blocksworld Revisited: Learning and Reasoning to Generate Event-Sequences from Image Pairs.”* Tejas Gokhale, Shalajja Sampat, **Zhiyuan Fang**, Yezhou Yang and Chitta Baral, (CVPRW on Language and Vision, 2019)
- “Weakly Supervised Attention Learning for Textual Phrases Grounding.”* **Zhiyuan Fang**, Shu Kong, Tianshu Yu, Yezhou Yang, (CVPRW on Vision and Language, 2018)
- “Range Loss for Deep Face Recognition with Long-tail.”* Xiao Zhang, **Zhiyuan Fang**, Yandong Wen, Zhifeng Li, Qiao Yu, (International Conference on Computer Vision (ICCV), 2017)
- “A Multi-resolution Analysis Approach for Spinal Metastasis Detection using Siamese Neural Network.”* Juan Wang, **Zhiyuan Fang**, Ning Lang, Huishu Yuan, Lydia Su, Pierre Baldi, (Computers in Biology and Medicine, 2017) (Honor Paper, 5%)
- “A Behavior Mining Based Hybrid Recommender System.”* **Zhiyuan Fang**, Zhang Lingqi, and Chen Kun. 2016 IEEE International Conference on Big Data Analysis (ICBDA). IEEE, 2016.
-

**PREVIOUS
EXPERIENCES**

Research Intern @ Microsoft Cloud & AI, Redmond Feb, 2021. - Sep, 2021

Keywords: VL Pre-training, Knowledge Distillation, Vision Transformer

Collaborators: Jianfeng Wang, Zhe Gan, Xiaowei Hu, Lin Liang, Lijuan Wang, Zicheng Liu

Project1: We explore and propose to conduct the Vision and Language Model Pre-training assisted by the Knowledge Distillation (KD) technique. Empirical results show that the VL representations learned via KD show better results than the regular VL pre-training methods when transferring to downstream VL tasks and are more data-efficient.

Project2: We develop ViTCAP, a novel detector-free image captioning model using Vision Transformer architecture. Experiments show that the full transformer architecture achieves CIDEr score 138.2 on COCO-Karpathy split with only 10% flops of detector-based captioning model.

Research Intern @ Microsoft Cloud & AI, Redmond May, 2020. - Sep, 2020

Keywords: Self-supervised Representation Learning, Knowledge Distillation

Collaborators: Jianfeng Wang, Lei Zhang, Lijuan Wang, Zicheng Liu

Project: We explore and study the self-supervised visual representation learning (SSL) in the schema of Knowledge Distillation for small visual backbone models. In particular, we find that the performances of SSL closely relate to the size learnable parameters of the visual backbone that small visual backbones typically lead to degraded representations. We propose to exploit a larger visual backbone as the Teacher model (trained and kept frozen by SSL method) and transfer its knowledge to the small Student architecture in an unsupervised manner.

Visiting Student@ MMLab of Chinese Academy of Sciences Sep 2016 - Feb 2017

Keywords: *Face Recognition, Image Classification*

Collaborators: Yu Qiao, Zhifeng Li

Project: We design and adapt contrastive loss for the face recognition task when training the training data is severely long-tail distributed.

Visiting Student@ University of California, Irvine June 2016 - Sep 2016

Keywords: *Deep learning, Biomedical Image processing,*

Collaborators: Juan Wang, Pierre Baldi

Project: We leverage the the image localization technique for identifying the spinal metastasis automatically using a deep Siamese Neural Network.

**HONORS
& AWARDS**

Outstanding Reviewer, British Machine and Vision Conference, 2019.
 Engineering Graduate Fellowship, Ira Fulton School of Engineering, 2019
 Honor Paper Award (5%), Computers in Biology and Medicine, 2018
 Outstanding Bachelor Thesis Award *SUSTech, 2017*
 Extraordinary Oral Presentation, IEEE, ICBDE, 2016

**ACADEMIC
SERVICES**

Journal Reviewer: *TIP, IEEE; Patter Recognition, Elsevier; TCSVT, IEEE*
 Conference Reviewer/Committee: *NeurIPS, CVPR, ICCV, ECCV, AAAI, ICLR, BMVC, ICRA, WACV, ACCV etc.*
 Emergency Reviewer: *BMVC 2019, CVPR 2020*
 Academic Seminar Organizer: *ASU-APG Vision and Language Seminar, Spring 2020*

Conference Workshop Organizer:

CVPR 2022: O-DRUM: Open-Domain Retrieval Under Multi-Modal Setting

TALKS:

Video2Commonsense: Generating Commonsense Descriptions to Enrich Video Captioning, Conference on Empirical Methods in Natural Language Processing, 2020.

Weakly Supervised Attention Learning for Textual Phrases Grounding, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, June 2019.

Machine Commonsense: From Video to Hidden Commonsense Knowledge, Telluride Neuromorphic Cognition Engineering Workshop, 2019.

**TEACHING
ASSOCIATIVE:**

<i>FIN 330: Data Mining and Data Analysis</i>	<i>SUSTech, 2017 Spring</i>
<i>CSE 205: Object Oriented Programming</i>	<i>ASU, 2017 Fall, 2018 Spring</i>
<i>CSE 310: Data Structure and Algorithm</i>	<i>ASU, 2017 Fall, 2018 Fall, 2021 Fall</i>
<i>CSE 571: Artificial Intelligence</i>	<i>Coursera-ASU, 2019 Spring</i>

REFEREES

Yezhou Yang (Ph.D. Advisor) ([WEBSITE](#))
Assistant Professor, Arizona State University
Tel: 301-661-3865
Email: yz.yang@asu.edu

Shu Kong ([WEBSITE](#))
Postdoc Researcher, Carnegie Mellon University
Email: aimerykong@gmail.com

Jianfeng Wang
Principle Researcher, Microsoft
Email: jianfw@microsoft.com

Tianshu Yu ([WEBSITE](#))
Assistant Professor, Chinese University of Hong Kong
Email: yutianshu@cuhk.edu.cn