

Hao Tan

1-(919) 519-6899 | haotan@cs.unc.edu | [Google Scholar](#) | github.com/airsplay

EDUCATION

University of North Carolina, Chapel Hill

Ph.D. in Computer Science. Advisor: Prof. Mohit Bansal

Chapel Hill, NC

Aug. 2016 – Aug 2021(Expected)

Shanghai Jiao Tong University

Bachelor in Computer Science, ACM honors class

Shanghai, China

Sep. 2012 – June 2016

RESEARCH INTERNSHIPS

Hugging Face

Supervisor: Thomas Wolf

Dec 2020 – Present

Paris, France

- Ongoing work on language and vision interactions. We study whether the pre-trained language model (e.g., BERT) could be applied to vision models.

Bloomberg AI

Supervisors: Chen-Tse Tsai, Yujie He, and Anju Kambadur

June 2020 – Oct 2020

Manhattan, NY

- Summarize salient chart information with natural language to automate financial news creation; use OCR and image regions simultaneously; endow the neural speaker with domain-specific knowledge.

Google Research

Supervisors: Vihan Jain, Eugene Ie, and Jason Baldridge

May 2019 – Aug. 2019

Mountain View, CA

- Train vision-and-language navigator in a curriculum: from short paths to long paths; build language generative models and instruction-splitting models to bridge the gap in currently-available data.

Adobe Research

Supervisors: Franck Deroncourt, Zhe Lin, and Trung Bui

May 2018 – Aug. 2018

San Jose, CA

- Execute and generate image-editing instructions; collect data and build a novel relational speaker which focuses on the transformations instead of objects.

Microsoft Research Asia

Supervisor: Xin Tong

Aug. 2015 – Feb. 2016

Beijing, China

- Reconstruct the strand-based hair model from the video; optimize it with spatial and temporal structural constraint; train a fully convolutional network that automatically segments the hair region.

SELECTED PROJECTS

Vokenization: Improving Language Understanding via Visual Supervision

Nov 2019 – May 2020

- Build a visually-supervised language model by predicting token-related images.
- Introduce the “vokenization” technique to contextually map language tokens to their related images (vokens).
- Consistent improvements over self-supervised alternatives on several language (Eng) tasks: GLUE (SST2 QNLI QQP MNLI), SQuAD, and SWAG.

LXMERT: Vision-and-Language Pre-Training Framework

Jan. 2019 – May 2019

- Develop the first pre-training framework for image-and-text tasks.
- Build novel cross-modal neural transformer model with object-based visual encoding.
- Pre-train on large-scale datasets with mask prediction, cross-modal matching, and image question answering tasks
- State-of-the-art results on three vision-and-language tasks: VQA (+2%), GQA (+3%), and NLVR2 (+22%).

TECHNICAL SKILLS

Programming Languages: Python, C/C++, Java, Matlab.

Deep-Learning Frameworks: PyTorch (4 projects), Tensorflow (3 projects), Caffe (2 projects).

PROFESSIONAL SERVICE

Workshop Organizer: SPLU-RoboNLP 2021.

Conference Reviewer: ACL, EMNLP, NAACL, AACL, INLG, ICLR, AAAI.

Workshop Program Committee: NeurIPS ViGIL 2019 and 2020, NeurIPS HAMLETS 2020.

ACHIEVEMENTS AND AWARDS

- **Bloomberg Fellowship**, Top 5% of applicants, Bloomberg Data Science Ph.D. Fellowship. 2019-2021
- **Rank 1**, VizWiz Challenge Leaderboard; aim to help visually-impaired people. Jul 2019
- **Rank 1**, GQA Challenge 2019; with the standalone model. May 2019
- **Top 3**, Visual Question Answering (VQA) Challenge 2019, May 2019
- **Rank 1**, Natural Language for Visual Reasoning for Real (NLVR2). May 2019
- **Rank 1**, Room-to-Room Vision-and-Language Navigation (VLN). Oct 2018
- **Rank 1**, Natural Language for Visual Reasoning. Apr 2018

PUBLICATIONS

1. **Hao Tan** and Mohit Bansal. “Vokenization: Improving Language Understanding via Contextualized, Visually-Grounded Supervision.” In *EMNLP 2020*.
2. Qinxin Wang, **Hao Tan**, Sheng Shen, Zhewei Yao, and Michael Mahoney. “MAF: Multimodal Alignment Framework for Weakly-Supervised Phrase Grounding.” In *EMNLP 2020*.
3. Xiang Zhou, Yixin Nie, **Hao Tan**, and Mohit Bansal. “The Curse of Performance Instability in Analysis Datasets: Consequences, Source, and Suggestions.” In *EMNLP 2020*.
4. Hyounghun Kim, Abhay Zala, Graham Burri, **Hao Tan**, and Mohit Bansal. “ArraMon: A Joint Navigation-Assembly Instruction Interpretation Task in Dynamic Environments.” In *Findings of EMNLP 2020*.
5. **Hao Tan***, Yubo Zhang*, and Mohit Bansal. “Diagnosing the Environment Bias in Vision-and-Language Navigation.” In *IJCAI 2020*.
6. Haonan Chen, **Hao Tan**, Alan Kuntz, Mohit Bansal, and Ron Alterovitz. “Enabling Robots to Understand Incomplete Natural Language Instructions Using Commonsense Reasoning.” In *ICRA 2020*.
7. Hyounghun Kim, **Hao Tan**, and Mohit Bansal. “Modality-Balanced Models for Visual Dialogue.” In *AAAI 2020*.
8. **Hao Tan** and Mohit Bansal. “LXMERT: Learning Cross-Modality Representations from Transformers.” In *EMNLP 2019*. **Oral (12.0%)**
9. **Hao Tan**, Franck Deroncourt, Zhe Lin, Trung Bui, and Mohit Bansal. “Expressing Visual Relationships via Language.” In *ACL 2019*. **Oral (8.0%)**
10. **Hao Tan**, Licheng Yu, and Mohit Bansal. “Learning to Navigate Unseen Environments: Back Translation with Environmental Dropout”. In *NAACL 2019*. **Oral (12.0%)**
11. **Hao Tan** and Mohit Bansal. “Object Ordering with Bidirectional Matchings for Visual Reasoning.” In *NAACL 2018*. **Oral (6.7%)**
12. **Hao Tan** and Mohit Bansal. “Source-Target Inference Models for Spatial Instruction Understanding”. In *AAAI 2018*. **Oral (8.2%)**
13. Licheng Yu, **Hao Tan**, Mohit Bansal, and Tamara L. Berg. “A Joint Speaker-Listener-Reinforcer Model for Referring Expressions.” In *CVPR 2017*. **Spotlight (8.0%)**