




LE ZHANG

 (+86)17317812540  z118300720002@gmail.com  <https://zhangle.netlify.app/>

EDUCATION

Mila & Université de Montréal

Sep. 2022 – June. 2024 (exp)

M.Sc in Computer Science

Montreal, Canada

Fudan University

Sep. 2018 – June. 2022

Bachelor of Engineering in Biomedical Engineering

Shanghai, China

Research Interest: Natural Language Processing, Computer Vision, Cross Vision Language Learning, Machine Learning

RESEARCH EXPERIENCE

Shanghai AI LAB

Feb 2022 – Aug 2022

Research Intern, Advisor: Prof. Siqi Sun

Shanghai

- Propose a new generative protein sequence transformer model - MSAT5
- Introduce a task utilizing millions of protein sequences to pretrain MSAT5

Georgia Tech's Social and Language Technologies (SALT) lab

May 2021 – Sep 2021

Summer Research Intern, Advisor: Prof. Diyi Yang

Remote

- Proposed **TreeMix**, a novel compositional data augmentation method for natural language understanding
- Evaluated on a wide range of benchmarks including GLUE and empirically validated that TreeMix outperforms current state-of-the-art data augmentation methods

Shanghai QiZhi Institute

Feb 2021 – Apr 2021

Research Intern, PI: Prof. Hang Zhao

Shanghai, China


- Expanded the task formulation and datasets introduced in *Visual Indicates Sound* by adding multi-modal features
- Designed a framework based on patch-wise contrastive learning and introduced a sound embedding module to utilize sound signals as an indicator for the transformation
- Designed a crowd-source labeling HTML page for Amazon Mechanical Turk

Fudan Data Intelligence and Social Computing Lab (DISC)

Sep 2020 – Present Day

Research assistant, Advisor: Prof. Zhongyu Wei

Shanghai, China

- Built a multi-modal event dataset: Collected multi-modal documents (image or video & text) from official and commercial press accounts on  Sina-weibo using crawler. Predefined categories and sub-types for social events e.g. {Military: [conflict, weapon operation, transportation]} , pre-process the dataset and hire crowd-source labelers to label the dataset.
- Programmed a crowd-source labeling pipeline and a labeling system based on Django.

PUBLICATION & MANUSCRIPTS

1. **Le Zhang**, Zichao Yang, Diyi Yang. TreeMix: Compositional Constituency-based Data Augmentation for Natural Language Understanding. **Submitted to 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2022)**[page][code]
2. Jingfeng Yang, **Le Zhang** , Diyi Yang. SUBS: Subtree Substitution for Compositional Semantic Parsing **Submitted to 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2022)**[page]

SELECTED AWARDS

- Outstanding graduate from Fudan University
- 2021-2022 Second Prize of the Scholarship for Outstanding Students at Fudan University
- 2020-2021 Huatai Securities Technology Scholarship at Fudan University (the First Prize, top 1%)
- 2019-2020 Huawei Scholarship at Fudan University (the First Prize, top 1%)
- 2018-2019 Second Prize of the Scholarship for Outstanding Students at Fudan University

SKILLS

Programming Languages: Python, C/C++, JavaScript, HTML/CSS, SQL, Bash, MATLAB

Tools and Frameworks: Git, \LaTeX , Pytorch, Huggingface, Docker