Shuailong Zhu

(+41)779875785 | shuailong.zhu@epfl.ch

Education

Master of Computer Science, EPFL — Sep.2022–present

Bachelor of Engineering, Zhejiang University | Major in EE — Aug.2018–Jun.2022

Honorable Program: Mixed Class, Chu Kochen Honors College, Zhejiang University

- GPA: 92/100, 3.99/4, 4.67/5
- Rank: 1/97
- **Major Courses:**
  Mathematics: Calculus I, Mathematical Analysis II (96/100), Linear Algebra (96/100), Probability and Mathematical Statistics (98/100), Complex Variable Functions (96/100), Matrix Theory (91/100), Optimization Algorithms (93/100)
  Information & Computer Science: Discrete Mathematics and Application (92/100), Data Structure (94/100), Object-oriented Programming (98/100), Computer Organization & Design (93/100), Operation System (94/100), Artificial Intelligence (93/100), Information Theory (95/100), Introduction to Quantum Information (96/100)

Research Experience

Controllability, Convergence and Stability of Soft Prompt Tuning — Aug.2021-Feb.2022

Advisor: Prof. Lingpeng Kong — *Intern, Shanghai AI Lab, The University of Hongkong*

https://github.com/One-punch24/Prompt_Learning_Intern

- **Target**: Multiple forms of soft prompt tuning have been proposed as a parameter-efficient way to better utilize the pretrained language model. My target is to explore the probability of bringing more detailed control with it in NLG and analyze its performance, convergence, and generalization ability in both NLU and NLG tasks.
- **Dispersive soft prompts for fine-grained control in conditioned text generation**: I design soft prompt series and separately embed them into the language model to instruct more fine-grained control on text generation and put up with the autoregression training scheme to tune the prompt series.
- **Instance-aware prompt injection**: Due to the lack of instance-level information in prompt tuning, I design a preprocess-block for better instance information fusion. The improvement of convergence speed can be observed in NLU tasks.
- **Explore the convergence speed and generalization ability of parameter-efficient tuning**: While comparing the convergence speed and generalization ability of Fine Tuning, Bitfit, Prompt Tuning, and Prefix Tuning for both NLG tasks (E2E and WebNLG) and NLU tasks (GLEU), I find that Prompt Tuning, Prefix Tuning, and BitFit exhibit similar behavior, which sacrifice convergence speed but contribute to better generalization ability in pretrained language model, while maintaining ideal performance.

Dynamic Masking in CNN Compression for FPGA Accelerator — Sep.2020-Apr.2021

Advisor: Prof. Kejie Huang — *Zhejiang University*

- **Target:** Our goal is to dynamically predict masks for the output feature map of ReLU and Pooling in CNN to reduce redundant computation.
- **Pre-computation branch for predicting mask:** There are two branches in the network: the master computation branch and the pre-computation branch. I use a low-bit quantization in the pre-computation branch to predict the mask dynamically for redundant computation positions of the output feature map, allowing the master branch to skip those masked positions.
- **Design retraining scheme:** The network performance degrades with the decrease of the number of weight bits in pre-computation branch. Thus, I design a retraining strategy, which only allows for gradient backpropagation in those unmasked positions of the feature map. The retraining proves to be able to recover the performance to the level of the uncompressed network (i.e., the performance increases about 2.1% in CIFAR dataset) while the number of weight bits in pre-computation branch decreases to 1.

Projects

Disentangled Control for 3D-aware Face Image Generation Feb.2022-May2022
Advisor: Prof. Yiyi Liao *Zhejiang University*

- **Target:** To obtain more fine-grained control in image synthesis while maintaining view-consistency.
- **Method & Current Performance:** With the implicit neural representation allowing for multi-view-consistent rendering, I add a mutual information term to enhance the correlation between synthetic images and the latent control code, and introduce weak supervision from traditional face model parameters, which successfully disentangles attributes like expression.

Malware Detection based on k-order-Markov Image Construction Jun.2021-Aug.2021
Advisor: Prof. Houman Homayoun *GREAT Program, UC Davis*
https://github.com/One-punch24/Malware_Markov_Image-ViT

Awards

First Class Scholarship of Zhejiang University (top 2/97 in major)	2020–2021
Cen Kefa Scholarship (top 1/97 in major)	2019–2020
Government Scholarship of Zhejiang Province (top 2% in Electrical Engineering Division)	2018–2019
Physics Innovation Competition of Zhejiang Province, First Prize	2019
National Mathematics Competition for College Students, Second Prize	2019

Academic Interests

Machine Learning Theory
Controllable Generative Models
Natural Language Processing

Programming Skills

python (torch, numpy, transformers(huggingface)), C++, matlab, verilog, html

Language skills

TOEFL iBT: 103

- Reading 28, Listening 25, Speaking 23, Writing 27

Hobbies

- Member of Basketball Team of Chu Kochen Honors College
- Learn some materials about Quantum Machine Learning: https://one-punch24.github.io/2021/12/22/A-Simple-Survey-of-Quantum-Machine-Learning/