HUANG JEN-TSE

Ph.D, The Chinese University of Hong Kong jthuang@cse.cuhk.edu.hk

Education

- **The Chinese University of Hong Kong** Hong Kong SAR, P.R.China
  Ph.D. in Computer Science and Engineering *2020.08 - Current*
  - ARISE Lab, Comp. Sci. and Eng. Dept.
  - Supervised by Prof. Michael R. Lyu
  - Fields of interest: Robust & Interpretable AI; Adversarial Attack & Defense;
- **Peking University** Beijing, P.R.China
  B.Sc. in Computer Science *2015.09 - 2019.07*
  - Yuanpei College
  - Major in Computer Science and Technology; Minor in Economics
  - Supervised by Prof. Jiaying Liu
  - Codes of selected course projects are available at https://github.com/penguinnnnn

Adwards

- **Scholarship for Hong Kong, Macao, Taiwan and Overseas Chinese** 2018
  Peking University

Research Projects

1. (Ongoing) Explore the boundary of bugs found by adversarial examples
   - Supervised by Prof. Michael R. Lyu.
   - Current methods for adversarial attack start from a seed datum that is successfully classified by the model and add perturbation until the model no longer classified it into the correct class. In this work, we start from a seed that is not classified correctly. We aim to answer two questions: 1. What is the boundary of those bugs? 2. Can a series of bugs be categorized into a same root cause?
2. (Ongoing) Textual adversarial attack using a general synonym set
   - Supervised by Prof. Michael R. Lyu.
   - Many existing adversarial attacks for NLP are based on synonym substitution on word level. In this work, we aim to extend the synonym granularity to phrase level and sentence level, leveraging the technique of aspect category sentiment analysis (ACSA).
3. (Under review in ISSTA'22) Improve the quality of test cases for NLP Software
   - Supervised by Prof. Michael R. Lyu.

- We find that most of the test cases produced by existing methods are in fact false alarms, which cannot discover DNN bugs and may harm model performance while re-training. We propose a method to score test cases and filter out bad ones. Results show that our method can select high-quality test cases and with the prioritization of test cases in re-training, we can obtain a more robust model with less resources/time consuming.

4. (Bachelor Thesis) Understanding and Analysis of Human Face Attributes

   - Supervised by Prof. Jiaying Liu and Prof. Bolei Zhou.
   - This work explores the interpretability of GANs on human face generation task, focusing on how face attributes are represented in latent space. This work provides a tool to modify desired attributes on a given synthesized image.

Research Internship

1. (2020.02 - 2020.07) Research Assistant in CSE Dept., The Chinese University of Hong Kong

   - Automated Reliable Intelligent Software Engineering (ARISE) Lab
   - Supervised by Michael R. Lyu
   - Research Topic: Interpretability and Robustness in Deep Learning Models;

2. (2018.11 - 2019.07) Internship in School of EECS, Peking University

   - Lab: Spatial and Temporal Restoration, Understanding and Compression Team (STRUCT)
   - Supervised by Jiaying Liu
   - Research Topic: Generative algorithms; Image Processing;

3. (2018.02 - 2019.06) Internship in Research Dept, SenseTime, Beijing

   - Team: Human Face Analysis and 3D Reconstruction
   - Supervised by Chen Qian and Bolei Zhou
   - Research Topics: Pose Estimation; Interpretability in Generative algorithms;

Languages and Skills

1. Languages

   - Chinese (Mandarin, Native)
   - English (Advanced) with certificates of CET4, CET6 and TOEFL

2. Programming

   - Unix Programming, C/C++, python, MATLAB, stata
   - Deep learning frameworks: pytorch, tensorflow, caffe