Dongjie Cheng

E-Mail: chengdongjiedev@outlook.com
Homepage: Dongjie Cheng (dongjie-cheng.github.io)

Education

SICHUAN UNIVERSITY — Chengdu, China
Artificial Intelligence — Undergraduate, Senior Year

GPA (Compulsory/Overall) — **3.88**/4, **3.78**/4
Rank — **4**/48
CET-6 — **636**

Honors

National Scholarship (1/48 that year) — 2023.11
National Third Prize, "China Software Cup" — 2023.08
Regional Second Prize, "RoboMaster North Region Competition" — 2023.06
Provincial Second Prize, "National College Mathematics Competition" — 2021.12
School Outstanding Student — 2022.11
School First-Class Scholarship — 2023.11

Research Publications

TV-SAM: Increasing Zero-Shot Segmentation Performance on Multimodal Medical Images Using GPT-4 Generated Descriptive Prompts Without Human Annotation

Co-First Author, Accepted, Big Data Mining and Analytics (JCR Q1, IF=7.7)
ArXiv, abs/2402.15759

Calibrated Self-Rewarding Vision Language Models

Co-First Author, Submitting to NeurIPS-2024 Main Track
(the short version is presented in ICML 2024 FM-Wild Workshop)
ArXiv, abs/2405.14622

Evaluating Hallucination in Text-to-Image Diffusion Models with Scene-Graph based Question-Answering Agent

Co-First Author, Submitting to NeurIPS-2024 D&B Track

SAM on Medical Images: A Comprehensive Study on Three Prompt Modes

Co-First Author, Cited by:78
ArXiv, abs/2305.00035

Experience

WEST CHINA HOSPITAL – BIG DATA CENTER — Chengdu, Sichuan
Research Assitant of Dr.Kang Li's Lab — February, 2023 – March, 2024

UNC-CHAPEL HILL — Remote
Remote Intern of Dr.Huaxiu Yao's Lab — March, 2024 – Present

VLM project — Key Project Member

Our work addresses these challenges by proposing the Calibrated Self-Rewarding (CSR) approach, which enables the model to self-improve by iteratively generating candidate responses, evaluating the reward for each response, and curating preference data for fine-tuning. In the reward modeling, we employ a step-wise strategy and incorporate visual constraints into the self-rewarding process to place greater emphasis on visual input. Empirical results demonstrate that CSR enhances performance and reduces hallucinations across ten benchmarks and tasks, achieving substantial improvements over existing methods by 7.62%. Our empirical results are further supported by rigorous theoretical analysis, under mild assumptions, verifying the effectiveness of introducing visual constraints into the self-rewarding paradigm.

I was responsible for the specific implementation and optimization of the CSR method, as well as core tasks such as DPO training and SFT training for VLM.

- *Calibrated Self-Rewarding Vision Language Models*

SAM project Key Project Member

In the SAM project, We proposed using large models to generate descriptions for segmentation targets, feeding theses descriptions to the detection model to produce bounding boxes for SAM, thereby achieving zero-shot segmentation.

I was responsible for conceiving and implementing specific experiments. Firstly, I completed the evaluation of the SAM model on multiple modalities medical datasets. Then I verified the effectiveness of the improvement method driven by LLM (Large Language Models).

The results show that the improved method performs well under zero-shot conditions, outperforming the GSAM (GLIP+SAM) baseline on most datasets. The project ultimately resulted in two papers, of which I am a co-first author.

- *SAM on Medical Images: A Comprehensive Study on Three Prompt Modes.*
- *TV-SAM: Increasing Zero-Shot Segmentation Performance on Multimodal Medical Images Using GPT-4 Generated Descriptive Prompts Without Human Annotation*

T2i-Eval project Key Project Member

In the T2i-Eval project, we proposed a method combining Scene Graph and Graph QA to score the quality of generated images, conducting a comprehensive evaluation of images from perspectives such as object omission, attribute inaccuracies, relational errors, and hallucinations.

I was responsible for generating evaluation dataset images, the specific design and experimentation of the Scene Graph part, achieving the construction of Scene Graphs through the use of GroundingDINO+BLIP VQA.

We constructed a human-evaluated dataset containing 12,000 images from 1,000 prompts and validated the effectiveness of our method. Compared with human evaluations, our Pearson and Kendall correlation coefficients surpassed those of T2ICompbench(Neurips 2023). This project ultimately resulted in one paper, for which I am a co-first author.

- *Evaluating Hallucination in Text-to-Image Diffusion Models with Scene-Graph based Question-Answering Agent*

Representative Courses

Math Courses:

Matrix Analysis **100**
Theory of Optimization (Convex Optimization) **94**
Discrete Mathematics **93**

Programming Courses:

Programming Language-I (C) **95**
Programming Language-II (C++) **99**
Python Programming for Artificial Intelligence **96**
Data Structure and Algorithm Analysis **94**
Robotics Programming with ROS **98**

Skills

- Proficient in programming languages including Python, C++, and C.
- Experienced with deep learning libraries such as Hugging Face and PyTorch.
- Skilled in version control with Git and proficient in using the Linux shell.
- Equipped with basic theoretical knowledge about machine learning