Yuhan Wei

281-710-2383 | yuhanwei10@gmail.com | Nirvana1010 | Yuhan Wei | Houston, TX, 77005 | GHC 23'

Education

Rice University *May 2024 (Exp.)*
Master of Data Science; ***GPA: 4.0/4.0*** *Houston, TX*
Nankai University *Jun 2021*
B.Eng. in Computer Science; ***GPA: 3.7/4.0 (Top 10%)*** *Tianjin, China*

Skills

Programming: Java, Python, Javascript, C, C++, SQL
Tools & Technologies: Docker, Shell, Git, Linux, HTML, CSS, Jupyter, Microsoft Office
Frameworks & Libraries: Express.js, React, Node.js, Flask, Spring, PyTorch

Publications

[1] **Yuhan Wei**, Zhifei Yang, Han Fang, Xianghao Zang, Chao Ban, Zerun Feng, Zhongjiang He, Yongxiang Li, and Hao Sun. Alignment and generation adapter for efficient video-related multitask learning. ***ICCV Workshop***, 2023.

Professional Experiences

Research Engineer Intern @ Tencent, China *May 2023 – Aug 2023*
Topic: Designed an efficient multi-task model for video-text retrieval and video captioning [1].

- **Cross-modal Adaption:** Proposed the **AGAdapter**, an alignment-generation adapter that integrates CLIP and large language models (e.g., LLaMA), which effectively bridges the gap between visual representations and text generation, resulting in enhanced cross-modal understanding and video-to-text generation capabilities.
- **Cooperative Learning:** Developed and implemented a novel approach in video captioning by incorporating visual representations learned from video-text retrieval, to enhance the alignment between different modalities.
- **Instructed Captioning:** Augmented the video-text pairs with instructions-following data to extract specific and informative video features, thereby providing better control over the process of generating video captions.
- **Performance Evaluation:** Demonstrated the superiority in 4 public datasets, achieving over **3%** enhancement in video captioning and **3%** improvement in video-text retrieval performance compared to SOTA baselines.

Relevant Projects

SpotsPedia - Fullstack | MongoDB, Express, React, Node, Javascript, Python, HTML/CSS
Developed and deployed a website for users to search and share information about their preferred travel spots.

- **Database:** Optimized **MongoDB** database with efficient data schema, reducing the query response times by **25%**, while ensuring data integrity through advanced features, leading to a **15%** improvement in website stability.
- **Security & Scalability:** Enhanced website security by implementing **JWT**-based authentication and role-based access control, descreasing the security incidents by **35%**. Improved API performance and scalability by leveraging **Express.js**, yielding a **45%** enhancement in response times and enabling support for higher user traffic.
- **Server Support:** Demonstrated proficiency in **Node.js** by creating scalable and efficient server-side applications, utilizing its asynchronous and event-driven nature, leadng to a **20%** increase in server performance.
- **AI-Enhanced Rec.:** Integrated AI to recommend customized travel spots, increasing **30%** satisfaction ratings.

CLIP-based models for Image Captioning | Python, PyTorch, Shell, Docker, Git
Designed 2 efficient CLIP-based image captioning models that achieved superior performance with less training time.

- **Zero-shot Prediction:** Incorporated CLIP-extracted object descriptions through a novel approach as input and then generated image captions with FlanT5. Achieved high-quality captions in a **training-free** manner.
- **Fine-tuned GPT2:** Utilized the CLIP image encoder to extract image features and combined them with text embeddings to fine-tune a pretrained GPT-2 model, reducing the training time by **4-fold** compared to baselines.

Music Genre Classification | Python, Javascript, PyTorch, Flask
Proposed a method for music genre classification, achieving accuracy improvement and efficiency enhancement.

- **Classification:** Proposed a **multimodal**-based method for music genre classification, incorporating a fusion of Mel-spectrograms (visual) and augmented MFCCs (textual) as input, resulting in a **10%** increase in accuracy.
- **Compression:** Introduced pruning and knowledge distillation to reduce model sizes, imporved efficiency by **20%**.
- **Web Demo:** Designed and built a website using **Flask** and **JavaScript**, allowing user-model interactions.