About Me

I'm **Sen Fang**, a bachelor of computer science with double degree. I am an intern at **Apache** and have also interned at universities such as **NTU**, **UTD**, and **UCF**, etc. My goal is to develop a general embodied intelligent agent that can assist ordinary people in various domains. In the early 2020s, I was drawn to topics like VR/AR in human and virtual interactions and NLP topics related to the interaction and unified representation across different modalities. These two areas were considered to be at the forefront of creating future technologies. However, the current understanding of artificial intelligence, especially in terms of achieving Artificial General Intelligence (AGI), has a significant gap from the initial concepts.

Initially, discussions about artificial intelligence focused on developing embodied intelligent agents with physical bodies. These agents were expected to possess three main capabilities: task planning, information gathering, and task execution, to assist people in achieving their goals in various domains. I believe that without sufficient intelligence, it is challenging to acquire the task planning capability. Fortunately, the advancements in large language models like OpenAI GPT-4 have largely addressed this issue. The remaining capabilities that need to be achieved are information gathering and task execution, which are closely related to my research areas.

My main research field is **Multimodal**. My research interests covers **Audio-Visual** (talking-face and representation of text/audio, Audio Generated Image), **AIGC** (AI-generated content, Multi-view learning, NeRF), **Self-Supervised** Learning (Pose recognition and modeling, object & action detection/recognition in videos, Medical Image Analysis) and **VR/AR/DCG** and **Visual Perception** (Enables the agent to make plan and navigate), as they are all relevant to developing the information gathering and task execution capabilities of a general embodied intelligent agent.

How to enhance the information gathering capability of AI ?

1. I believe that fusing different modalities of perception can provide a more comprehensive understanding and intelligence to computer systems. Therefore, I have focused on multimodal learning, particularly in the audio-visual domain, including talking-face and representation of text/audio. I have been developing new models and algorithms to better understand and process multimodal data and apply them in practical scenarios such as human-computer interaction, virtual reality, and audio-generated images. As a result, I completed a paper titled *"Exploring Efficient-Tuned Learning Audio Representation Method from BriVL."* which was accepted by ICONIP 2023, the flagship conference of APNNS, with the assistance of a professor at Nanyang Technological University in Singapore.

2. I consider self-supervised learning as the foundation for training models to gather information and execute tasks. Self-supervised learning is a method for automatically learning representations from unlabeled data, enabling us to leverage large-scale data for training. I have been exploring the applications of self-supervised learning and multimodal analysis in various areas. Through self-supervised learning, I hope to provide more accurate and reliable solutions in these fields and contribute to the advancement of related technologies. Based on my previous work, I completed a paper titled *"UniBriVL: Robust Universal Representation and Generation of Audio Driven Diffusion Models."* which was accepted by EMNLP 2023

MRL. I also submitted a paper titled *"Bridging the Gap between Text, Audio, Image, and Any Sequence: A Novel Approach using Gloss-based Annotation"* to ICASSP 2024.

How to develop the embodied and task execution capabilities necessary for AGI ?

1. I have focused on areas such as human pose recognition and modeling, object and action detection/recognition in videos, and applications related to sign language. I completed a paper titled *"SignDiff: Learning Diffusion Models for American Sign Language Production"* under the guidance of a prominent professor (over 8,000 citations) in the multimodal field at the University of Texas at Dallas. Additionally, I worked on *"SignLLM: Sign Languages Production Large Language Models"* under the guidance of a renowned CV professor (over 15,000 citations) at the University of Central Florida. Both papers, which are state-of-the-art works in the fields of sign language rendering and generation, have been submitted to CVPR 2024. These works have not only provided assistance to underrepresented communities, such as the deaf and mute, but also advanced my understanding and progress in human pose recognition, modeling, and object and action detection in videos.

2. Based on the above work, I can extract some fundamental methods that can significantly contribute to the advancement of "recognizing everything and learning everything." Some of my work can also advance the field of "modeling all behaviors." By combining these two aspects with the large language model as the brain, we can approach the realization of embodied intelligent agents, enabling AI to contribute to the well-being of a wider range of people.

In addition to research, I actively participate in academic activities and serve as a reviewer for multiple international conferences, including ACL, EMNLP, and ICONIP. I have also been a core member in various projects (such as Apache) to enhance my engineering skills and have been involved with my alma mater's virtual reality lab to explore more interaction modes. I have published papers at several international conferences through collaborations with researchers from around the world.

I have chosen to apply for the Ph.D. program in Computer Science at your esteemed university because it boasts an excellent faculty team and abundant research resources in the field of computer science. I am highly interested in the research projects and laboratories at your university and hope to collaborate and learn from outstanding researchers. I have already made contact with some professors at your university (i.e. DIMITRIS N. METAXAS; Xintong Wang), so I don't anticipate any issues in terms of adaptation. My future work will be determined based on the guidance and recommendations of the professors.

During my graduate studies, I intend to delve deeper into areas such as multimodal learning, self-supervised learning, human-computer interaction, and virtual reality. Particularly, I aim to achieve my ultimate goal of implementing embodied intelligent agents and contribute to the advancement of related fields. I believe that your university's Ph.D. program will provide me with an excellent platform and opportunities to fulfill my personal goals and aspirations in research. I am eagerly looking forward to joining your Computer Science Ph.D. program and making my contributions to the field's development.