ARIA
APPLIED RESEARCH IN ACTION

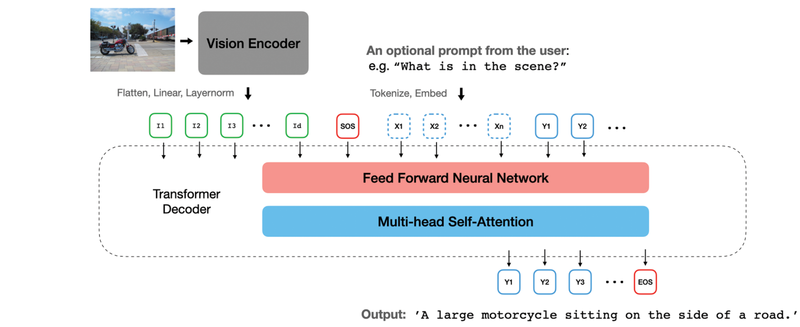
Visual Perception of Walking Environments using Deep Learning and Large Language Models

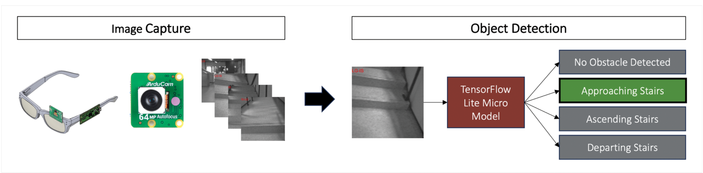
Stair obstacle detection using AI-powered smart glasses and image captioning of walking environments using generative vision-language models to provide assistance to patients with visual impairments

Haining Tan

Alex Mihailidis and Scott Schwartz
ACADEMIC SUPERVISORS

Brokoslaw Laschowski
INDUSTRY SUPERVISOR





PROJECT SUMMARY

AI-powered bionic vision systems can replicate various aspects of biological vision. These assistive technologies can help restore vision such as recognizing objects and describing scenes to patients with visual impairments. In this study, we developed models for visual perception of human walking environments, with an emphasis on stair obstacle detection and environment captioning. For stair recognition, we optimized an efficient convolutional neural network on a large image dataset of real-world stair environments (Meta Ego4D) for smart glasses. After training, the model was converted to Tensorflow Lite and deployed on our embedded device for onboard real-time predictions. For environment captioning, we used transformer-based vision encoders and text decoders for image feature extraction and caption generation, leveraging the recent advances in vision-language models. Initialized from large-scale pre-training, the model was optimized on image-text pairs from the MSCOCO dataset related to walking environments. The architecture also incorporates user prompts as an optional feature, allowing captions to be generated based on user feedback. Our visual perception models demonstrate robust performance and provide a foundation for next-generation AI-powered smart glasses that provide assistance and sensory feedback to patients with visual impairments.

UHN Toronto Rehabilitation Institute
The KITE Research Institute

Computer Science
UNIVERSITY OF TORONTO
Master of Science in Applied Computing