Contact Info.

Name: Kent(Changlong) Li **E-mail:** lclkent@qq.com **Address:** Beijing Institute of Technology, Beijing, China

Education Background

2018.09-2021.06 **Master of Science in Mechanical Engineering, Beijing Institute of technology**

- **Grade**: Top 10%; **TOEFL**: 102; **Awards**: First class scholarship for graduate students.
- **Research direction:** Deep learning, multi-modal ground terrain identification.
- **Research interest:** Robotic perception, multi-modal perception/generation, industrial design.

2014.09-2018.07 **Bachelor of Engineering in Engineering Mechanics, Hunan University**

- **Grade:** Top 15%; **Awards**: Scholarship for undergraduate students, Golden award of national 3D industrial design competition, National patent of invention.
- **Research direction**: Coupled thermo-mechanical analysis, CAD, CAE, industrial design.

Research experience

2017.12-2018.06 **Coupled thermo-mechanical analysis of marble-disk brake**

The work is to simulate the working state of the marble disk brake compression mechanism at high temperature, and to analyze the stress and strain of the structure and the change law of the contact pressure in the loading process by the sequential coupled thermal structure coupling analysis. The structure of the marble disc compression mechanism is improved and optimized based on the analysis results.

2018.10-Now **Multi-modal ground terrain identification**

A multi-modal ground terrain identification method based on vehicle body vibration signal as "tactile" information and road image information as "visual" information. A synchronized multi-modal dataset was collected using ROS on the experimental vehicle. Temporal models (RNNs, Transformer-encoder) is used as the network architecture for "tactile" perception. A texture classification model is used for "visual" perception. Two perception results were fused by a confidence decision mechanism.

2019.05-2019.09 **Multi-scale speech synthesis based on GAN and wavelet transform**

Using multiscale conditional GAN to learn the coefficients of high frequency wavelet transform on multi-scale to capture the high-frequency features of audio signals. In the speech synthesis stage, the inverse wavelet transform is used to reconstruct the signal based on the generated wavelet coefficients. The method can synthesize a large amount of realistic speech samples in parallel.

2019.10-2020.4 **Voice2Video: Imagining the talking face behind the voice**(work done as an intern at Ping An Technology)

An end-to-end speech-to-talking face generation task is proposed. A dual GAN model is proposed to restore the speaker's facial information and facial dynamics from speech. In the stage of speech information decoupling, cross modal dictionary learning is proposed. The speaker dictionary learned from face recognition network is used to supervise the speech dictionary learning network. Image and video generators are interconnected to generate realistic talking face videos.

2020.01-2020.04 **Speaker Recognition Based on multimodal knowledge distillation and self supervised learning**

The student network uses self-supervised contrastive learning model CPC and face recognition network VGGFace is used as cross modal teacher network, the teacher network helps to extract speaker info in mixed embeddings of the self-supervised model. It is more robust to different emotional tones by training with facial information. Downstream speaker recognition task using multimodal knowledge distillation reported 93.4% accuracy, better than training on labeled speech-along dataset (accuracy: 91.7%) (Dataset: CREMA-D)

Work experience

2019.10-2020.04 Deep learning algorithm researcher at **Ping An Technology**.

2020.05-2020.08 Algorithm engineer at **DeepGlint**, responsible for optimizing the plate recognition algorithm.

Skills

Proficient in CAD software UGNX/Solidworks, CAE software ANSYS/ABAQUS, 3D rendering software KeyShot.

Proficient in Python/MATLAB/C++, deep learning framework Pytorch/Tensorflow. Comfortable with Linux, ROS.

Achievements

- "Road pavement identification based on acceleration signals of off-road vehicles using the batch normalized recurrent neural networks", ICAICA2019. (First author)
- "Voice2Video: Imagining the talking face behind the voice", in submission to ICASSP2021. (First author)
- "A Combined Robust Approach based on Auto-Regressive Long-Short Term Memory Network and Moving Horizon Estimation for State-of-charge Estimation of Lithium-ion Batteries", in submission to International journal of energy research. (Second author).
- "Optimized Handling Stability Control Strategy for a Four In-wheel Motor Independent-drive Electric Vehicle", IEEE Access. (Fifth author)
- Golden award industrial design demo: https://www.youtube.com/watch?v=nkHCSxH6snw