Curriculum Vitae

Lingyu Zhang
aileenlingyu@gmail.com
9179458460

EDUCATION:

Rensselaer Polytechnic Institute, Troy, NY
Ph.D. in Electrical Engineering, expected Aug 2021, GPA: 3.93/4.00

Columbia University, New York, NY
Master of Science in Electrical Engineering, Dec 2017, GPA: 3.71/4.00

Huazhong University of Science and Technology, Wuhan, China
Bachelor of Engineering in Optoelectronic Information Engineering, June 2016, GPA: 3.83/4.00

RESEARCH TOPIC:

Multimodal deep learning (combining vision, audio and language), Computer vision, Information retrieval, Visual question answering, Video moment retrieval, Semantic video analysis, Vision and Language fusion, Natural language understanding.

SKILLS:

- Python, C++, MATLAB
- TensorFlow, Keras, PyTorch, scikit-learn, OpenCV
- NumPy
- Training neural networks on a cluster-based distributed computing platform such as AiMOS supercomputer
- CUDA, multiprocessing, Linux environment
- Blender, Meshlab
- Hadoop, Spark

PUBLICATIONS:

- **L. Zhang** and R.J. Radke, Temporal Attention and Consistency Measuring for Video Question Answering, ACM International Conference on Multimodal Interaction (**ICMI**), October 2020.
- **L. Zhang** and R.J. Radke, A Multi-Stream Recurrent Neural Network for Social Role Detection in Multiparty Interactions. IEEE Journal of Selected Topics in Signal Processing (2020).
- **L. Zhang**, I. Bhattacharya, M. Morgan, M. Foley, C. Riedl, B.F. Welles, and R.J. Radke, Multiparty Visual Co-Occurrences for Estimating Personality Traits in Group Meetings, 2020 IEEE Winter Conference on Applications of Computer Vision (**WACV**). IEEE, 2020.
- **L. Zhang**, M. Morgan, I. Bhattacharya, M. Foley, J. Braasch, C. Riedl, B.F. Welles, and R.J. Radke. Improved Visual Focus of Attention Estimation and Prosodic Features for Analyzing Group Interactions. ACM International Conference on Multimodal Interaction (**ICMI**), October 2019.
- M. Li, **L. Zhang**, R.J. Radke, and H. Ji, Keep Meeting Summaries on Topic: Abstractive Multi-Modal Meeting Summarization, Annual Meeting of the Association for Computational Linguistics (**ACL**), July 2019

PROFESSIONAL RESEARCH EXPERIENCE:

Rensselaer Polytechnic Institute, Troy, NY May 2018 – Now
Multimodal deep learning for human interaction dynamics analysis from visual, language and audio signals.
Position: Ph.D. research assistant
Research Supervisor: Prof. Richard Radke

- Visual focus of attention estimation and emergent leader prediction in a group meeting scenario.
  - **Collected** an **audio-visual dataset** for estimation social signals in **multi-people interaction**.
  - Annotated social signals including Big-Five **personality traits**, emergent **leadership** rating within a group.
  - Designed a **neural network**-based visual focus of attention estimation algorithm from un-calibrated 2D visual recordings based on **eye gaze and head pose** to detect the **dynamic visual target** (i.e., **where the person is looking at**) at each frame for each participant in the meeting, predicted **emergent leader** within a group.

Achievements: Got state-of-the-art result for visual focus of attention estimation from frontal videos. Published a paper in ICMI 2019.

- **Detecting** co-occurrences of **actions across multiple people** in the interaction for automatic personality traits screening.
  - Detected frame-wise **hand positions** from 2D video using CNN.

- Estimated human **motion intensity** level and **number of moving body parts**.
- Constructed **co-occurrence** patterns that capture the **action/ behaviors** of the participant herself and her interactions with others**,** predicted personality traits (Openness, Consciousness, Extraversion, Agreeableness, Neuroticism) scores based on the **interactive behavior cues**.
- Analyzed the **correlation** between behavior cues and personality traits thus making an interpretable model more **interpretable**.

Achievements: Designed a novel approach that integrates multimodal co-occurrence patterns of human behavior cues in spatial and temporal dimensions. Published a paper in WACV 2020.

- A **multi-stream recurrent neural network** for **dynamic** multiparty **human** interactive **behavior** analysis from **visual, audio and language.**
  - Extracted frame-wise visual features including **head pose and eye gaze behavior**, body movements, **facial expression (facial action units)**.
  - Extracted **natural language embedding** for transcripts in the video using **GloVe** model.
  - Extracted **acoustic features** such as zero-crossing rate, energy, spectral entropy and MFCCs.
  - Performed **multimodal fusion** for **visual, audio and language** sequences on a frame-basis.
  - Designed a novel **LSTM-based deep neural network** model that takes **multimodal temporal sequences** from multiple people as input and captures intra-modality, inter-modality and inter-personal temporal dependencies to predict **dynamic social roles** (Protagonist, Neutral, Supporter, or Gatekeeper) in the group meeting.
  - Analyzed the importance of various behavior cues that are relevant for dynamic social roles, making the model **interpretable**.

Achievements: Designed a novel approach for comprehensive human **multimodal interactive behavior** modelling. Got state-of-the-art result for dynamic social role prediction. Published a journal paper in IEEE JSTSP.

- Temporal Attention and Consistency Measuring for **Video Question Answering**.
  - Constructed multimodal temporal sequences for video context and the question based on visual features extracted by **DenseNet161**, **audio** features extracted by COVAREP and natural language features by **BERT** model.
  - Designed a **temporal attention** mechanism that **highlights** the **keywords** in the question (**person subjects**, **objects**, **actions**, **environmental** constraints), **key sentences** in the transcript and **critical moments** in the long video.
  - Designed a multi-level **consistency measuring** module and a reasoning module to infer which candidate answer semantically matches the question most.

Achievements: Designed a novel VQA model, got state-of-the-art result on Social-IQ dataset. Published a paper in ICMI 2020.

- **Video Moment Retrieval from Natural Language Query**.

Achievements: Designed a novel algorithm for retrieving video moment from **natural language query**, got state-of-the-art result on TVR and Charades-STA dataset. Submitted a paper to ICML and is currently under review.

DVMM lab at Columbia University, New York, NY September 2016 – December 2017
Refined RANSAC for 3D primitive shape fitting based on multi-scale local features
Research Supervisor: Prof. Shih-Fu Chang

- Extracted multi-scale local features for scanned **3D object** using C++.
- Developed machine-learning based classification algorithm to predict primitive shape class using scikit-learn, Python.
- Developed Refined RANSAC fitting algorithm to represent whole 3D object with a set of geon shapes using C++.

Achievements: The large-scale scanned LiDAR data lacks structure and semantic interpretation. Without further processing, the information directly gathered from raw point cloud data can be used for maximum distance detection which is far from sufficient for automatic car driving systems or interactive robot sensing. In this project, we developed an algorithm which combines machine learning methods with RANSAC to fit the LiDAR data with a set of basic geon shapes including cone, cube, cylinder, ellipsoid, pyramid, sphere and torus. It can help us evaluate the objects' outlines and manage the scale of object model efficiently by modifying basic parameters.

American International Group (AIG), New York, NY September 2017 – December 2017
3D reconstruction for car accident damage estimation from single 2D image
Position: Research Assistant
Achievements: Trained a deep learning model to detected anchor points of the damaged car from 2D image and reconstructed 3D shape of the vehicle. Estimated the 3D area of the damage for the vehicle in an accident.