

Multimodal Machine Learning

Yan Yan

Email: yyan34@iit.edu

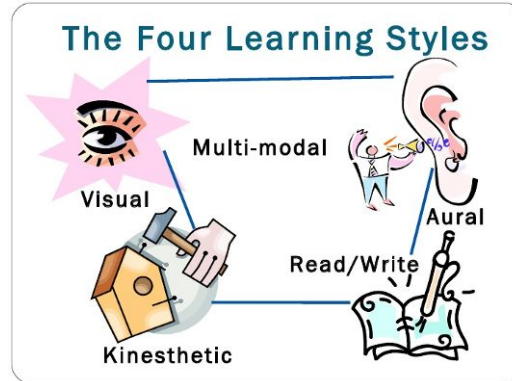
Department of Computer Science

Research Areas: Computer Vision, Multimedia, Machine Learning

Outlines:

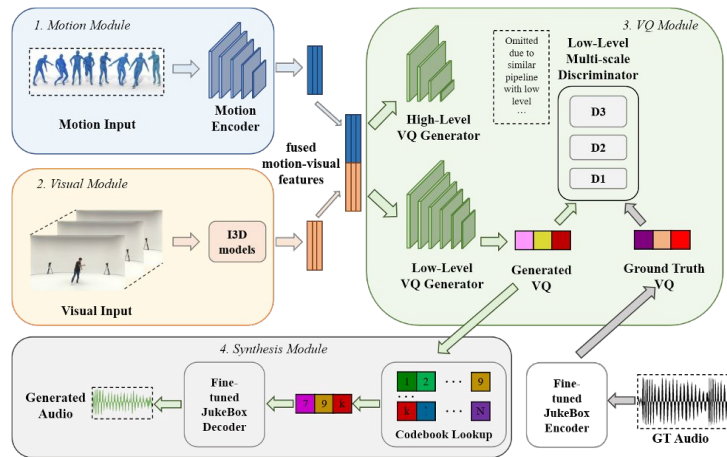
- 1. Multimodal Learning*
- 2. Typical CV problems*
- 3. New Sensors*
- 4. Interdisciplinary Research*

Computer Vision

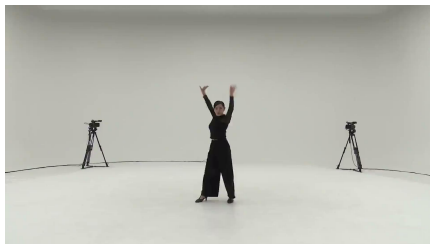


Audio / Speech Processing

Natural Language Processing

Multimodal Learning: Dance Videos to Complex Music Generation

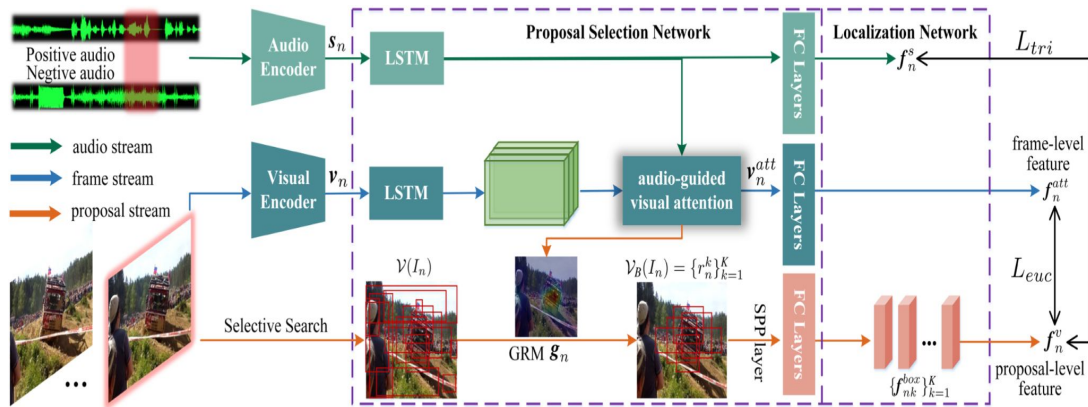
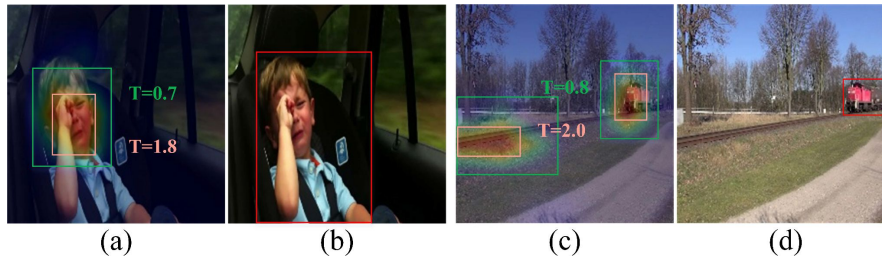
Ground Truth Sample



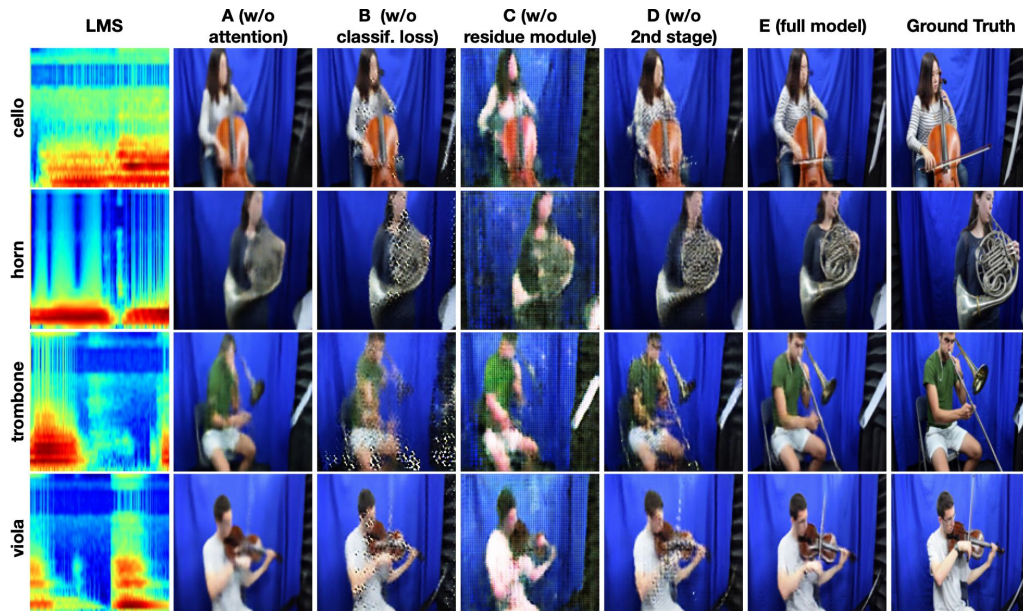
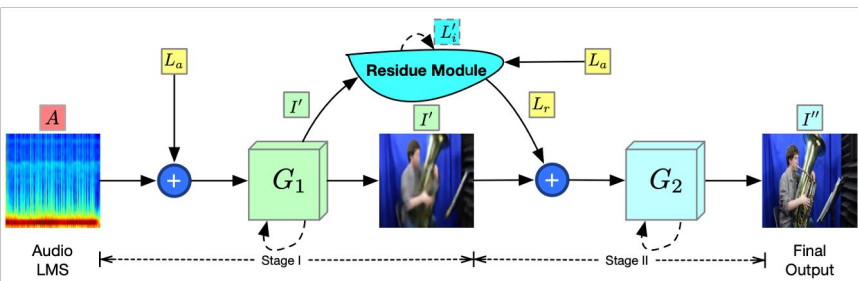
Music Generation Results



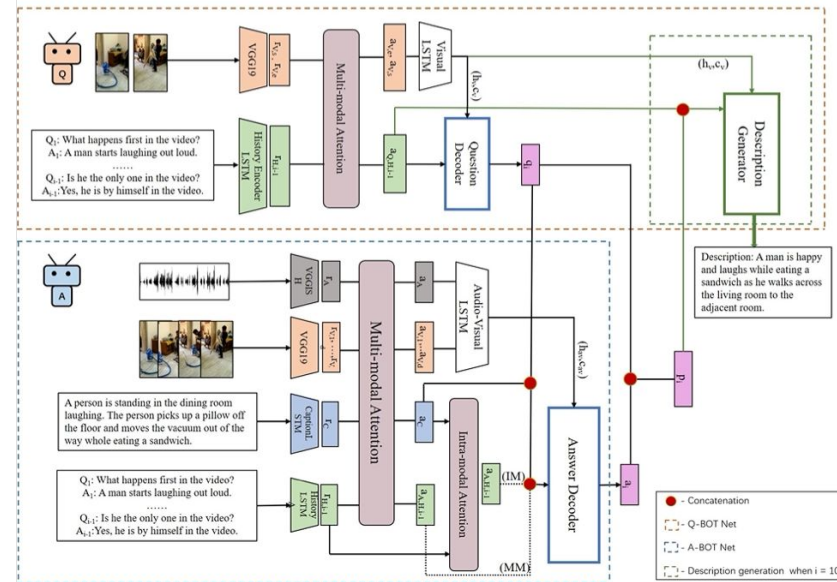
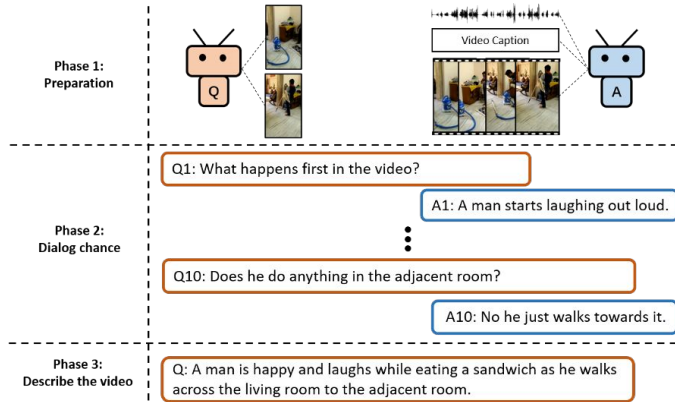
Seeing the voice: Sound Source Localization in Videos



Sound to Video Generation



Unseen Video Description via QA-Cooperative Network

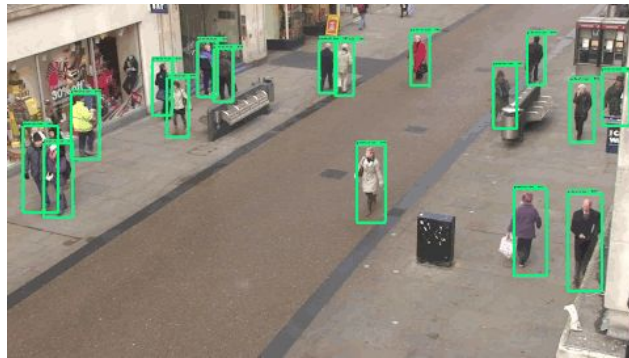


A step forward to more secure and reliable AI systems

- No direct access to original data
- Natural language dialog as the implicit information source
- Balance between the performance and security concerns

Typical CV problems:

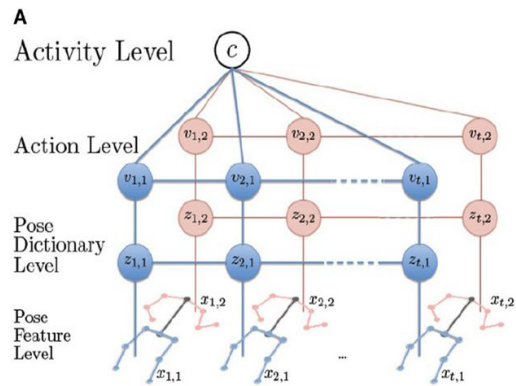
Surveillance



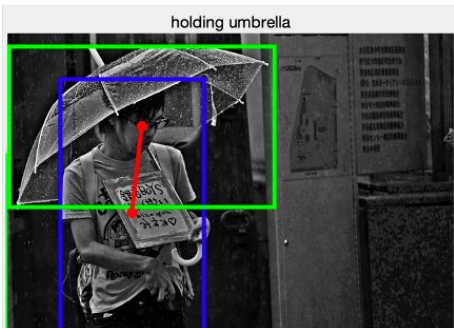
Person Re-ID



Activity Recognition and Human Pose estimation



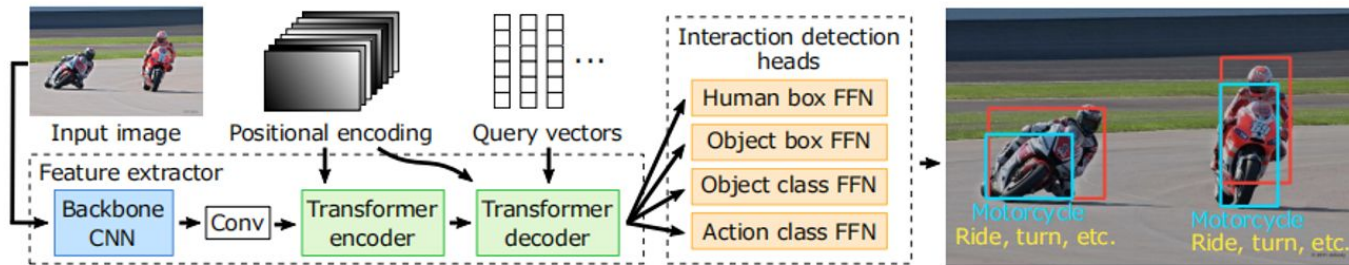
Human-Object Interaction Detection



Human Holding Umbrella

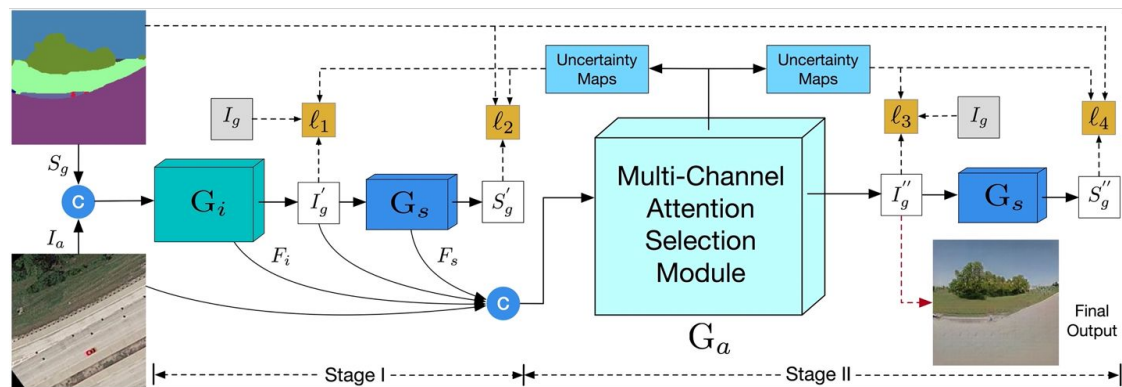
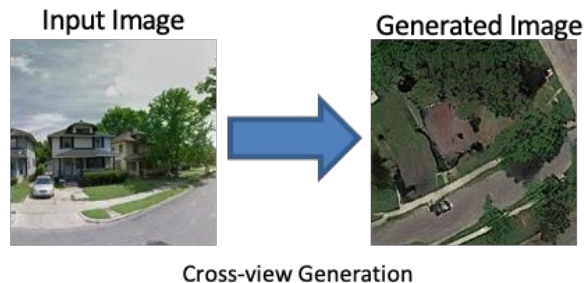
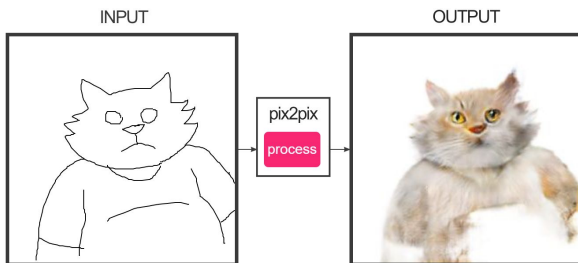
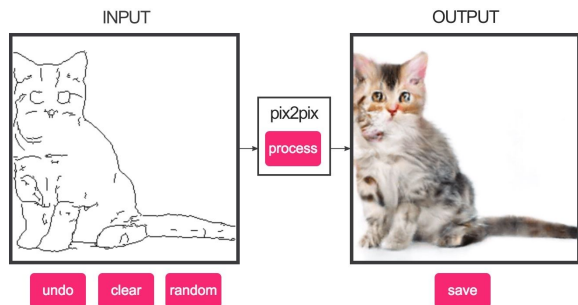


Human Drinking with Bottle



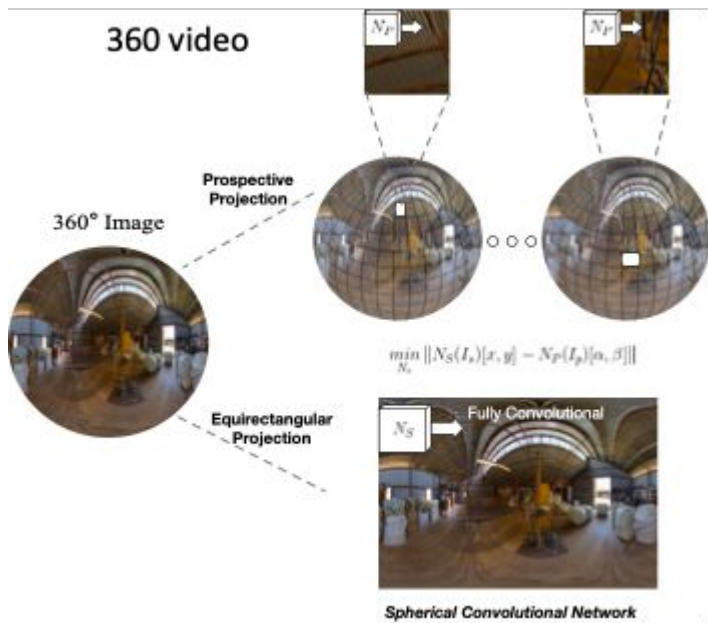
End-to-end HOI detection: Transformer-based query embedding learning and multi-branch detection network

Cross-view Image/Video Generation

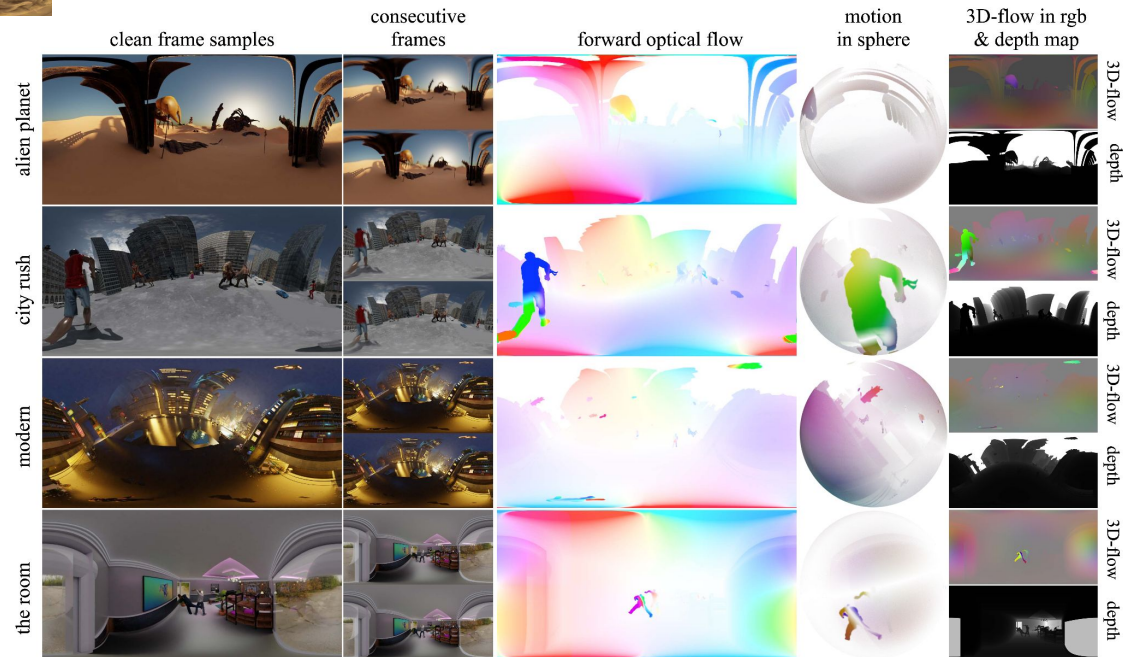


New Sensors:

360 Video Analysis



Spherical Convolutional Network



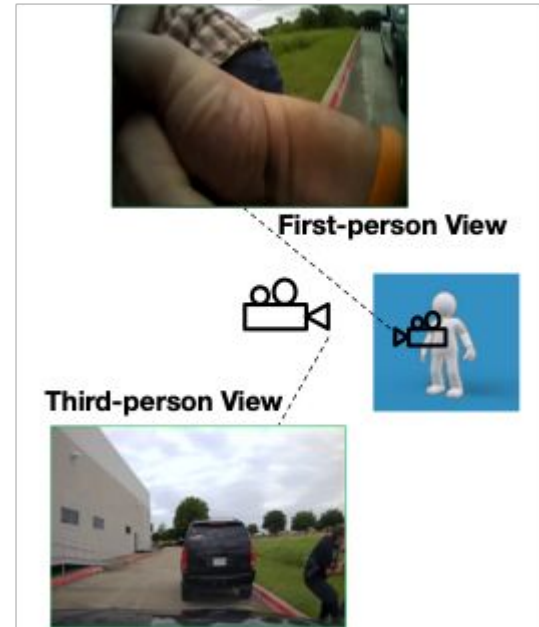
Optical Flow Estimation in 360 Video

First Person Video Analysis (Wearable Devices)



First Person Video Summarization
First Person Activity Recognition

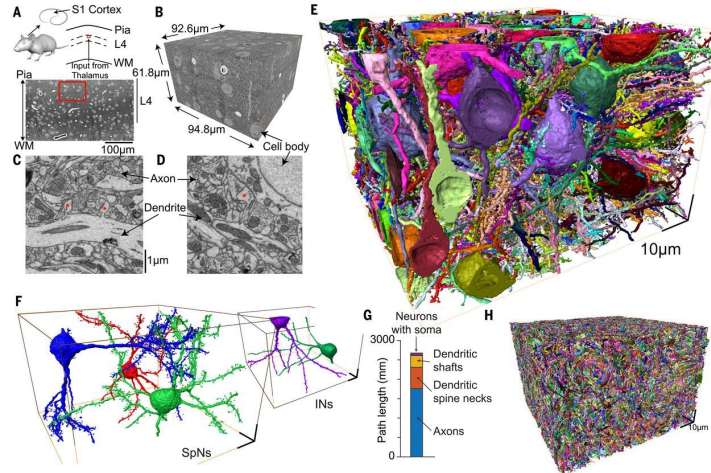
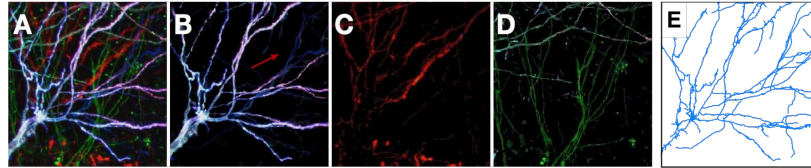
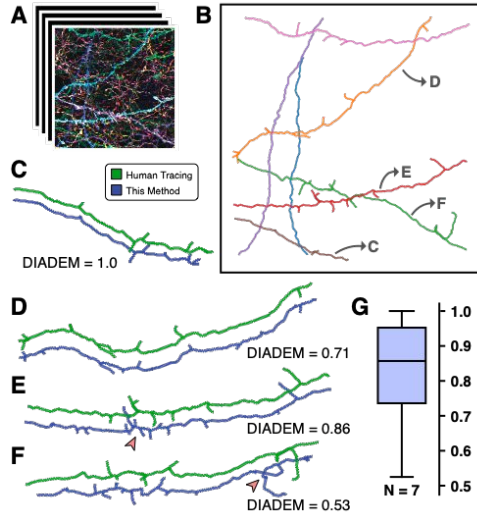
First-person (egocentric) video



Interdisciplinary Research:

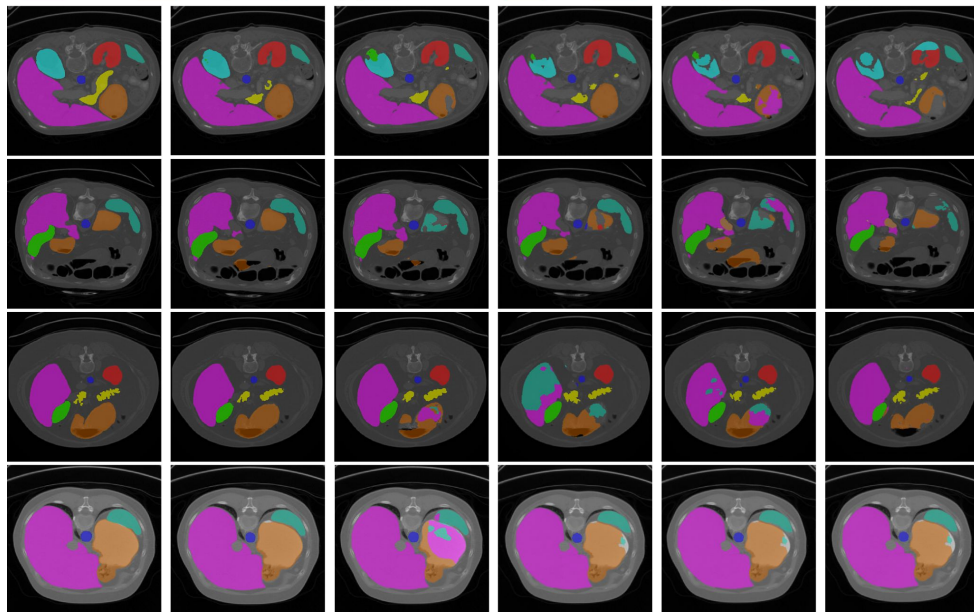
1. Neuroscience

Neural Tracing and Reconstruction in Densely Labeled Multispectral Images



2. Medical Image Analysis

■ Spleen
 ■ right kidney
 ■ left kidney
 ■ gallbladder
 ■ liver
 ■ stomach
 ■ aorta
 ■ pancreas



GT

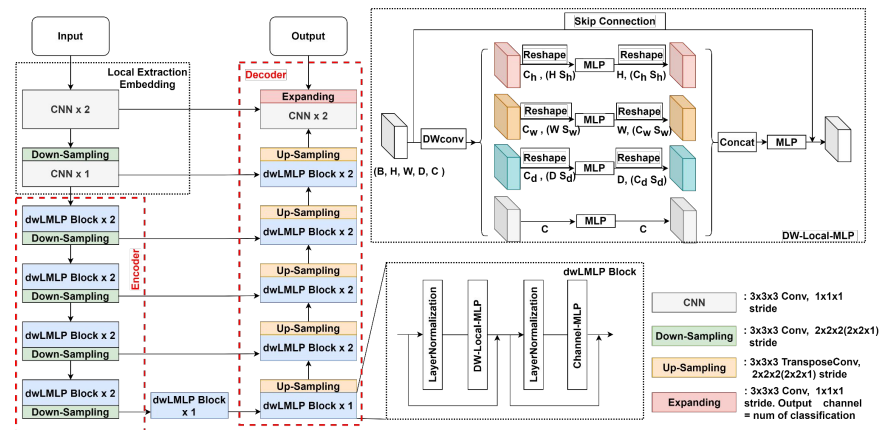
U-MLP

nnUNet

SwinUNet

TransUNet

UNet



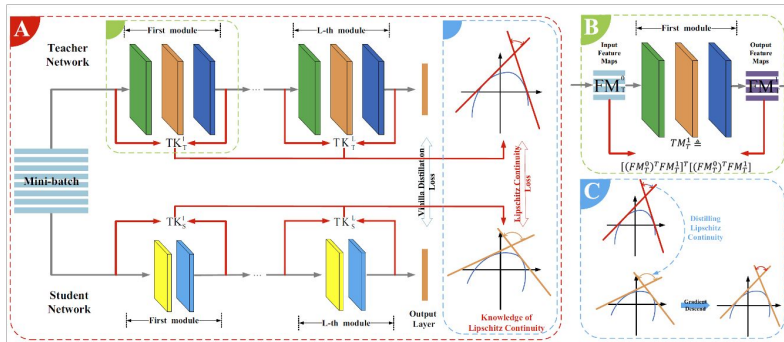
3. System - Efficient AI

Large Neural Network



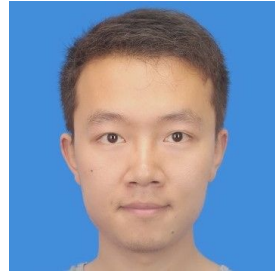
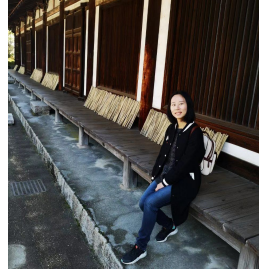
G
A
P

Resource-constrained devices



**Frequencies Guided Neural Network Pruning,
Lipschitz Continuity Guided Knowledge Distillation,
Neural Network Binarization via Contrastive Learning,
etc.**

Thanks for all my PhD students and funding agencies.



ILLINOIS TECH

College of Computing



Thanks!

Q&A

I am actively seeking potential collaborations with Argonne National Laboratory.

Yan Yan

Department of Computer Science

Research Areas: Computer Vision, Multimedia, Machine Learning

yyan34@iit.edu