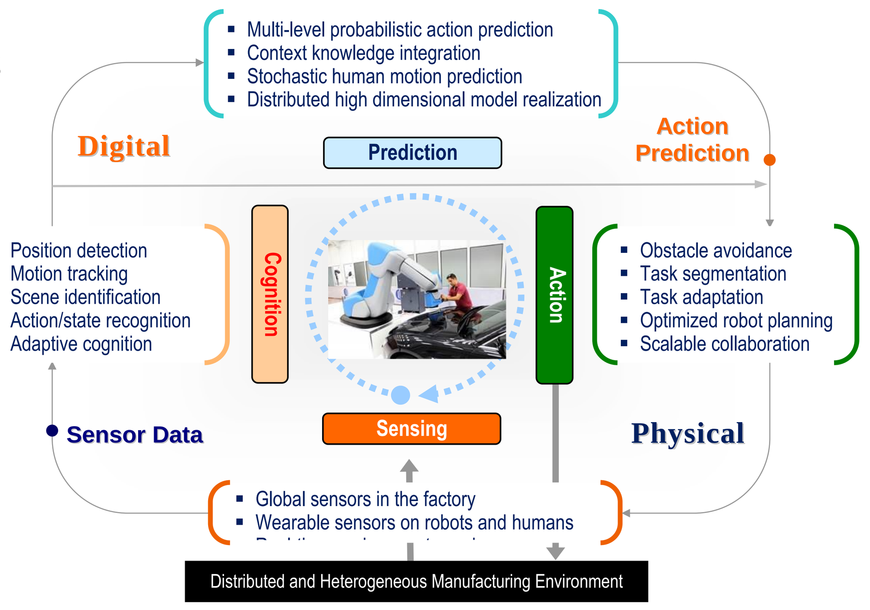
NRI: INT: COLLAB: Manufacturing USA: Intelligent Human-Robot Collaboration for Smart Factory

Robert X. Gao, Department of Mechanical & Aerospace Engineering, Case Western Reserve University
Ming Leu, Department of Mechanical & Aerospace Engineering, Missouri University of Science and Technology
Gloria Wiens, Department of Mechanical & Aerospace Engineering, University of Florida
Zhaozheng Yin, AI Institute, Department of Biomedical Informatics and Department of Computer Science, Stony Brook University

We propose four research components to realize the envisioned human-robot collaboration (HRC) for an automated HRC manufacturing cell, from data acquisition in the physical domain to data manipulation in the digital domain, and back to robot control in the physical domain: sensing, cognition, prediction, and action.



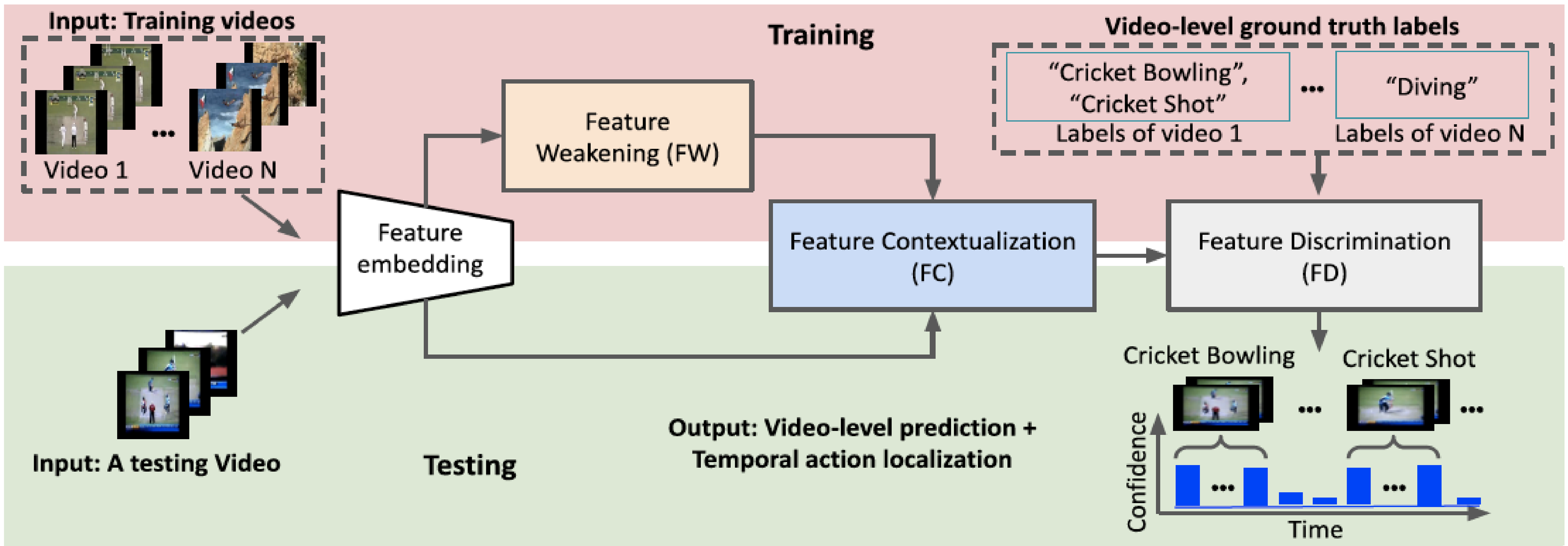
Sensing/ Cognition

Feature Weakening, Contextualization, and Discrimination for Weakly Supervised Temporal Action Localization [1]

Task Objective: Design a weakly-supervised temporal action localization network to localize all human action instances potentially from different classes in an untrimmed video.

Technical Approaches:

Feature weakening: Design a novel Feature Weakening (FW) module that can precisely localize the action instances in both discriminative and ambiguous action-related frames, without spreading to the background intervals.

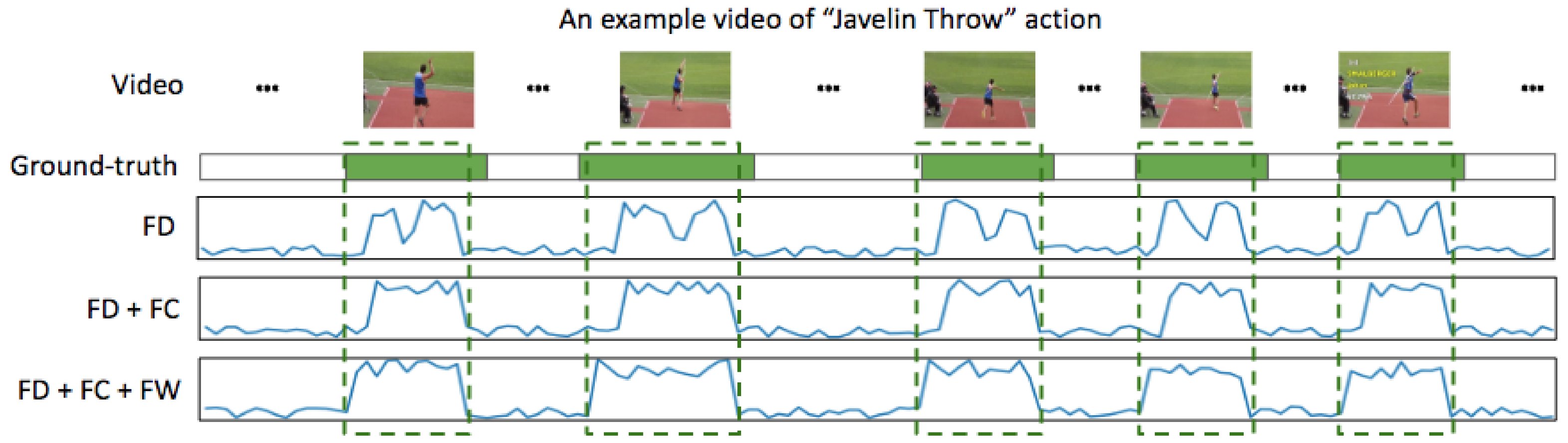


The workflow of our weakly-supervised temporal action localization.

Feature contextualization: Develop a novel Feature Contextualization module that can infer the global contexts among video segments and attentionally fuse them with the local contexts from individual video segments to generate more representative contextualized features.

Feature discrimination: Develop a new Feature Discrimination (FD) module that can highlight the most discriminative video segments/classes corresponding to each class/segment, respectively.

An example video of "Javelin Throw" action



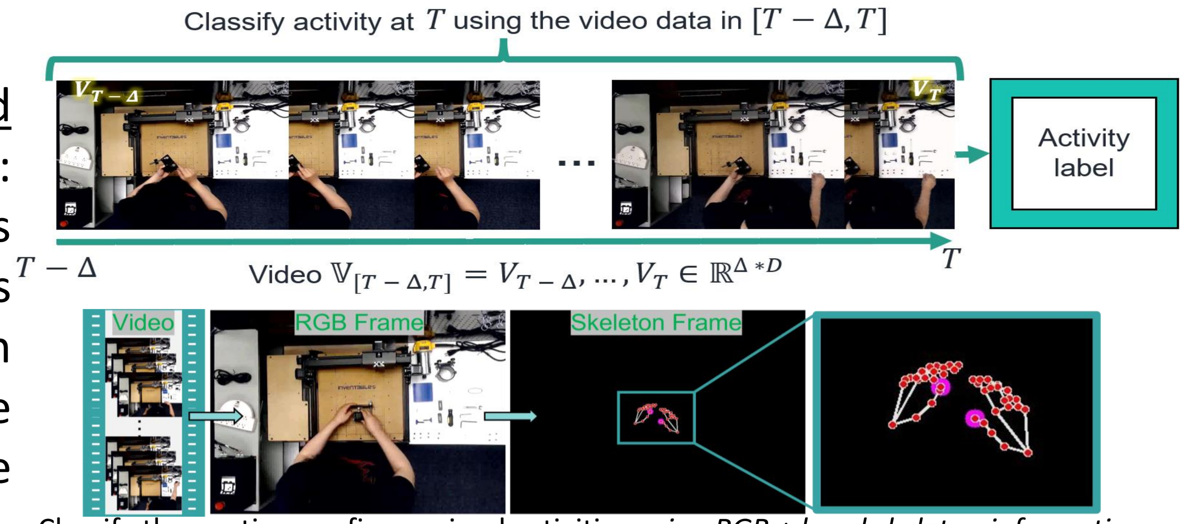
Qualitative results. With many background frames, the video contains multiple instances of "Javelin Throw" action. The FD coarsely localizes all the action instances, while the FC helps us get smooth localization. Finally, the FW further improves the performance by localizing the action instances in ambiguous frames.

Fine-Grained Activity Classification in Assembly

Task Objective: Design a multi-visual-modality-based system to sense, recognize, and predict a worker's continuous fine-grained assembly activities in a manufacturing platform using convolutional and recurrent neural networks.

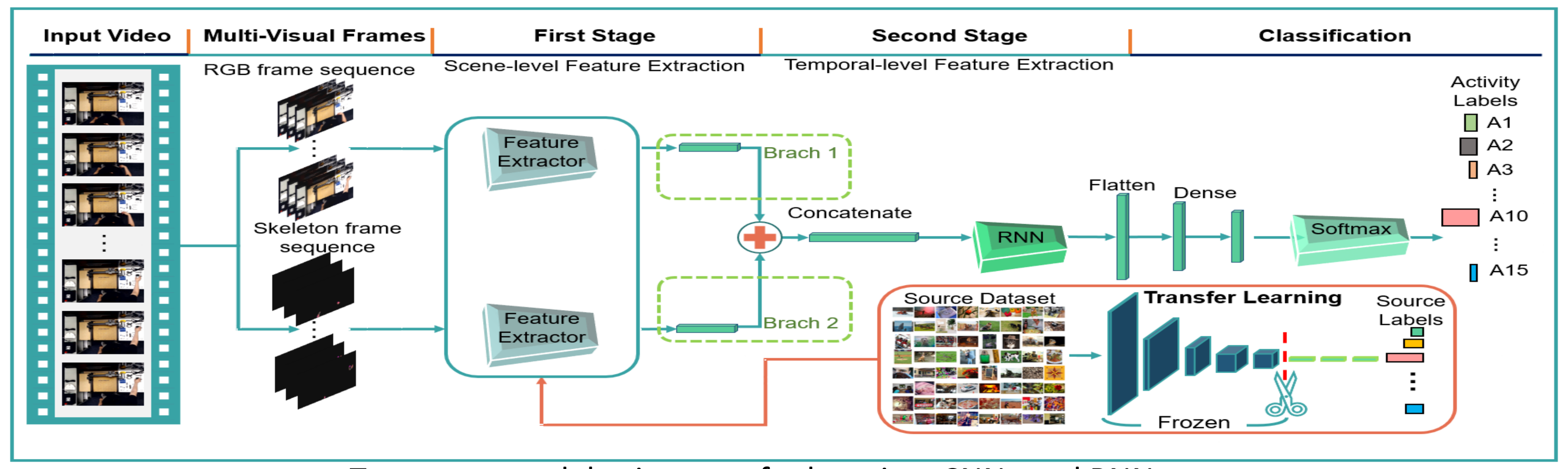
Technical Approaches:

Assembly Fine-grained activity recognition: Classify the continuous fine-grained activities in assembly at each time step T using the video data in the time interval $[T-\Delta, \ldots, T]$.



Classify the continuous fine-grained activities *using RGB + hand skeleton information*

Two-stage model: Multi-visual modality: RGB + Skeleton frame sequences;
First stage-> using CNN -> scene-level features of RGB and skeleton frames;
Second stage -> using RNN -> temporal-level features of frame sequences;



Two-stage model using transfer learning, CNNs and RNNs.

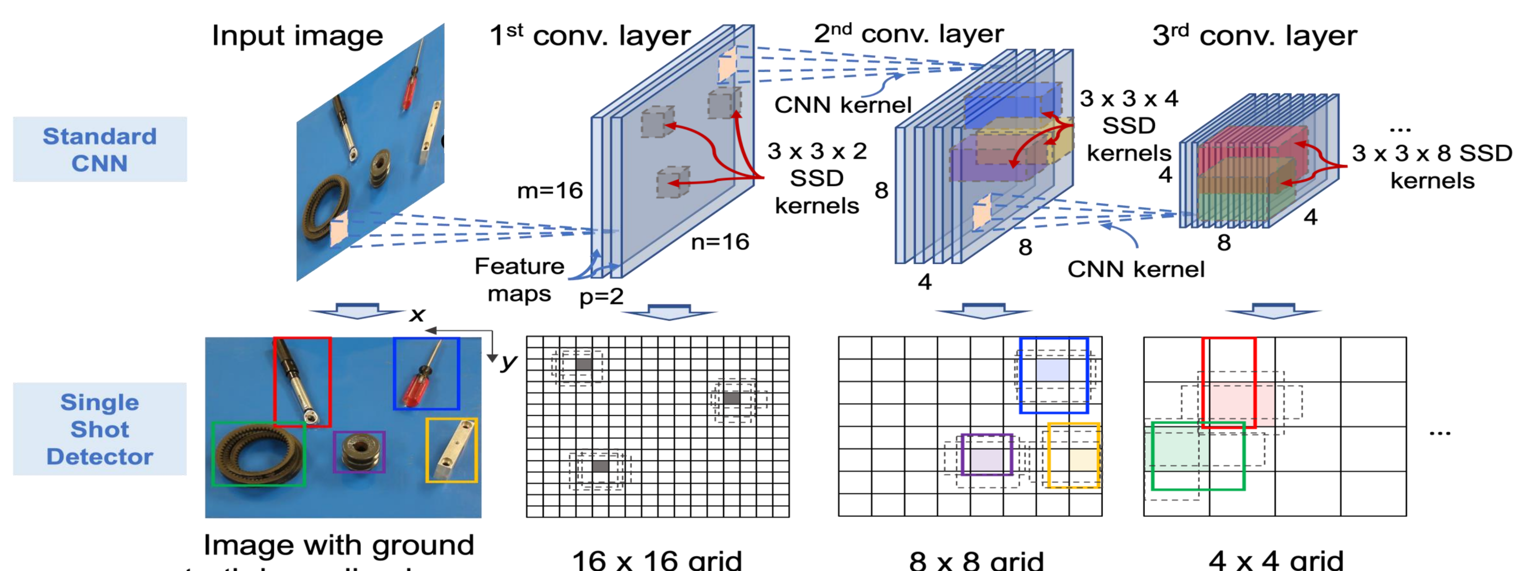
Prediction

Object Detection and Robot Joint Angle Prediction to Enable Robot Action for Collaboration [2][3]

Task Objective: To determine location and orientation of part/tool needed for subsequent assembly operation given predicted human action and trajectory, and to predict required robot joint angles for successful part/tool grasping.

Technical Approach:

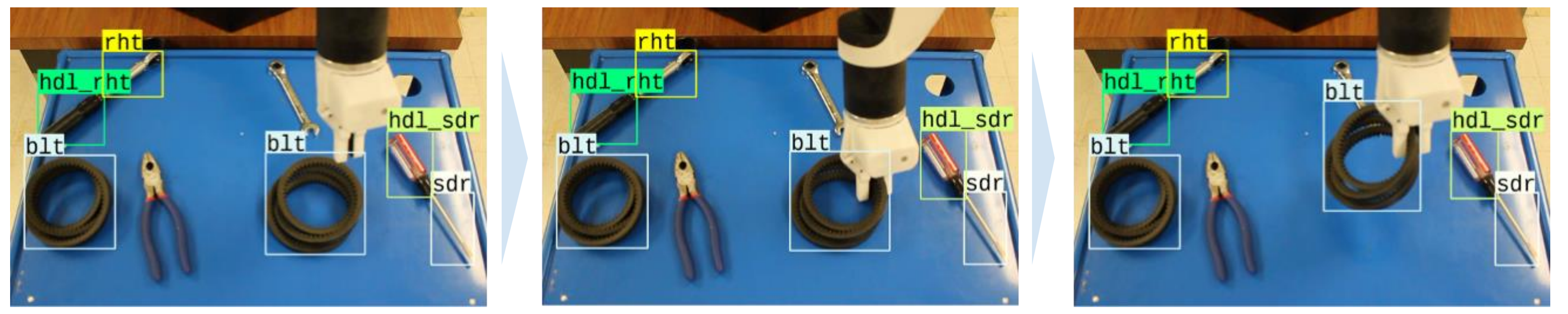
Single-shot object detector (SSD): utilize feature maps in convolutional neural network (CNN) as grids to predict bounding boxes around objects. Each grid cell predicts a bounding box through convolution with SSD kernels. Feature maps of different dimensions detect and localize objects of different scales.



Self-supervised joint angle prediction: combine data-driven approach and analytical forward kinematics to determine joint angles needed for end-effector to arrive at desired location and orientation as determined by SSD for part/tool grasping.

Evaluation Outcome:

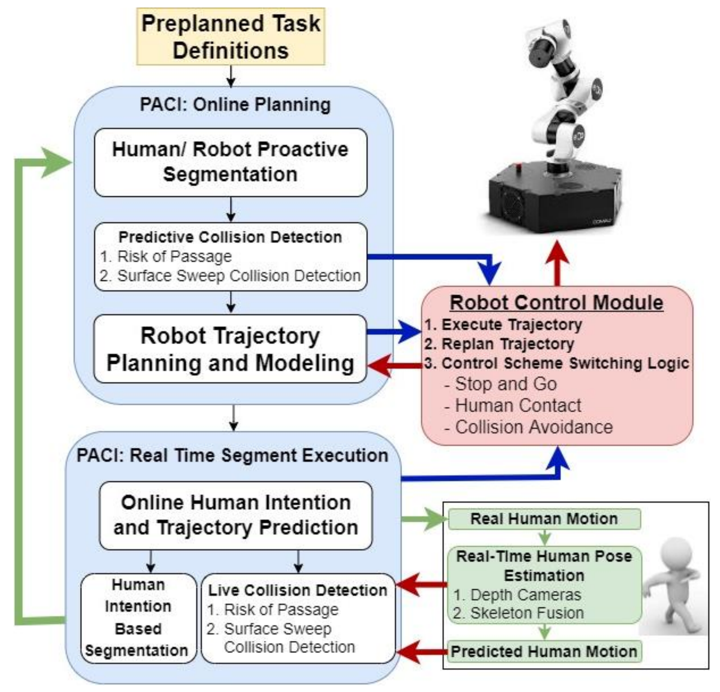
Object detection: 97.1% mean accuracy among 7 different assembly parts/tools
Joint angle prediction: 2° mean error in robot end-effector orientation during grasping



blt: belt **cap**: bearing cap **shv**: sheave **hdl_sdr**: screwdriver **sdr**: screwdriver (shank) **hdl_rht**: ratchet **rht**: ratchet (shank)

Action

Human-Robot Proactive-n-Reactive Behavior Intelligence

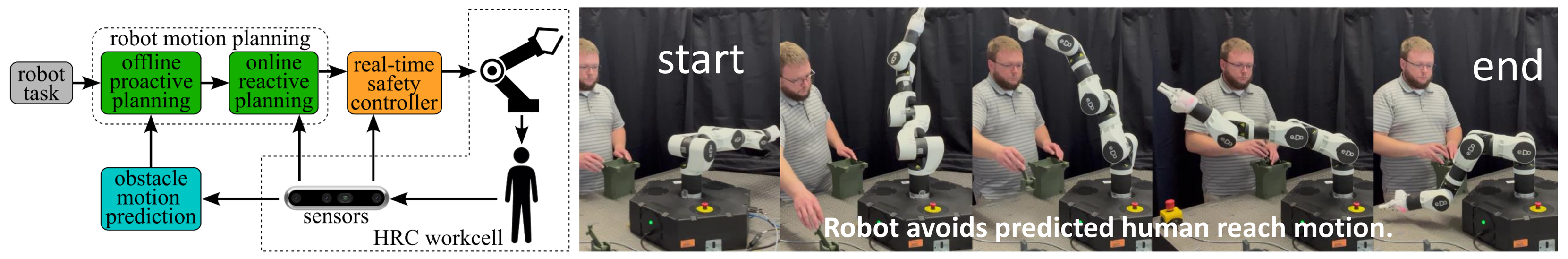


PACI Control Architecture

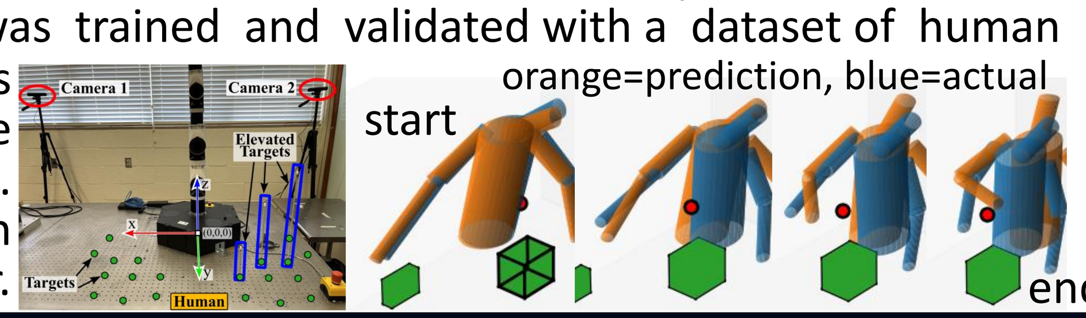
Task Objective: Seamless integration of sensing, cognition, and prediction into robot controller yielding efficient Proactive Adaptive Collaboration Intelligence (PACI) to ensure safe interactions with humans and mitigate production disruptions.

Segmentation and Augmented Intelligence: Control provides 'proactive-n-reactive' robot response using a segmentation framework of predicted, real time, and behavioral classifications to allow for modularity and flexibility within the control architecture.

Proactive Path Planning: Developed the Spatio-Temporal Avoidance of Predictions (STAP) planning method [4]. STAP integrates spatio-temporal human occupancy maps, time-avoidance cost function, and an RRT* variant. These features allow STAP to anticipate human movements and the effect of ISO1566 *Speed and Separation Monitoring* in finding optimal robot trajectories. STAP trajectories minimize robot delay and time spent very close to humans in tight human-robot collaboration. The images below show a sequence of robot motion generated by STAP.



Prediction of Human Reaching Motions: A generative neural network (GNN) predicts a multi-step sequence of human poses for tabletop reaching motions. The input is the human's reaching target (red dot in images) relative to current pelvis location combined with current human pose [5]. The predicted sequence is mapped to a time-series based on a learned human speed versus reach distance model. Network was trained and validated with a dataset of human motions to reach various positions on or above the table in front of the human. Method does not suffer an exponential growth of error.



2023 FRR & NRI Principal Investigators' Meeting
May 2-3, 2023

NSF/NRI 2.0: ENG/CMMI and CISE/IIS
Award ID#: 1954548/1830479/1830383/1830295

Related Publications
[1] M. Moniruzzaman, and Z. Yin, "Feature Weakening, Contextualization, and Discrimination for Weakly Supervised Temporal Action Localization", IEEE-TMM, 2023.
[2] Zhang *et al.*, "Machine Learning-based robotic object detection and grasping for collaborative assembly," Int. Symp. Flexible Autom. 2022.
[3] Wang *et al.*, "Data-driven process characterization and adaptive control in robotic arc welding," CIRP Annals, 2022.
[4] J. Flowers, M. Faroni, G. Wiens, and N. Pedrocchi "Spatio-Temporal Avoidance of Predicted Occupancy in Human-Robot Collaboration", submitted to IEEE ROMAN, 2023.
[5] J. Flowers, and G. Wiens, "Prediction of Human Reaching Pose Sequences in Human-Robot Collaboration", submitted to ASME IDETC, 2023.