EgoPAT3D is a comprehensive dataset created and maintained by the AI4CE group at the New York University to enhance action target location prediction during human manipulation tasks.

In recent years, egocentric datasets were created to address three types of tasks: (1) Action Anticipation (2) Region Prediction (3) Trajectory Prediction. Meanwhile, the aforementioned datasets were created mainly for 2D, with only a few were created in 3D. With this in mind, Prediction of Action Target in 3D (PAT3D) is a special case of trajectory prediction, considering it is usually hard to predict trajectories when human hands remain mostly out of scope in egocentric views during manipulation tasks. Therefore, we limit our focus to predicting the endpoints (target) for human manipulation tasks, with a goal to compensate for action latencies and support timely reactions in human-robotics interactions.

To that end, our dataset contains 15 scenes, and each scene contains (1) 10 RGB-D (visual)/IMU (motion) recordings (2) 1 point cloud of the scene. Along with the dataset, we design a classification loss function by using the prediction confidence to weigh the classification importance. In addition, we propose a baseline model, where the visual information and the motion information are extracted separately by two MLP’s before being concatenated and passed to another RNN-based prediction module for 3D target localization.

Overall, our main contributions are:

1. We initiate the first study of egocentric target prediction in 3D.
2. We build a novel dataset as well as evaluation metrics.
3. We propose a baseline model for continuous prediction of action target.