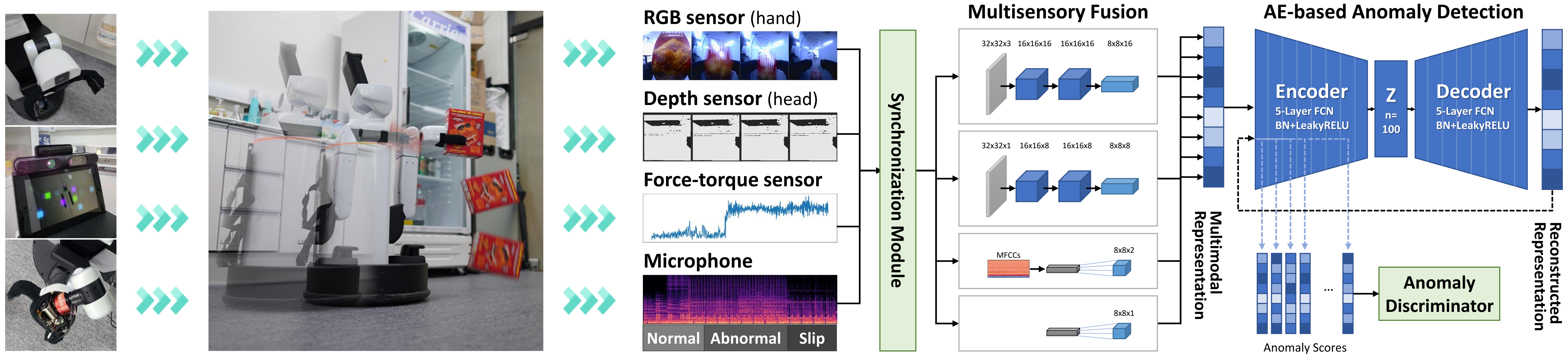
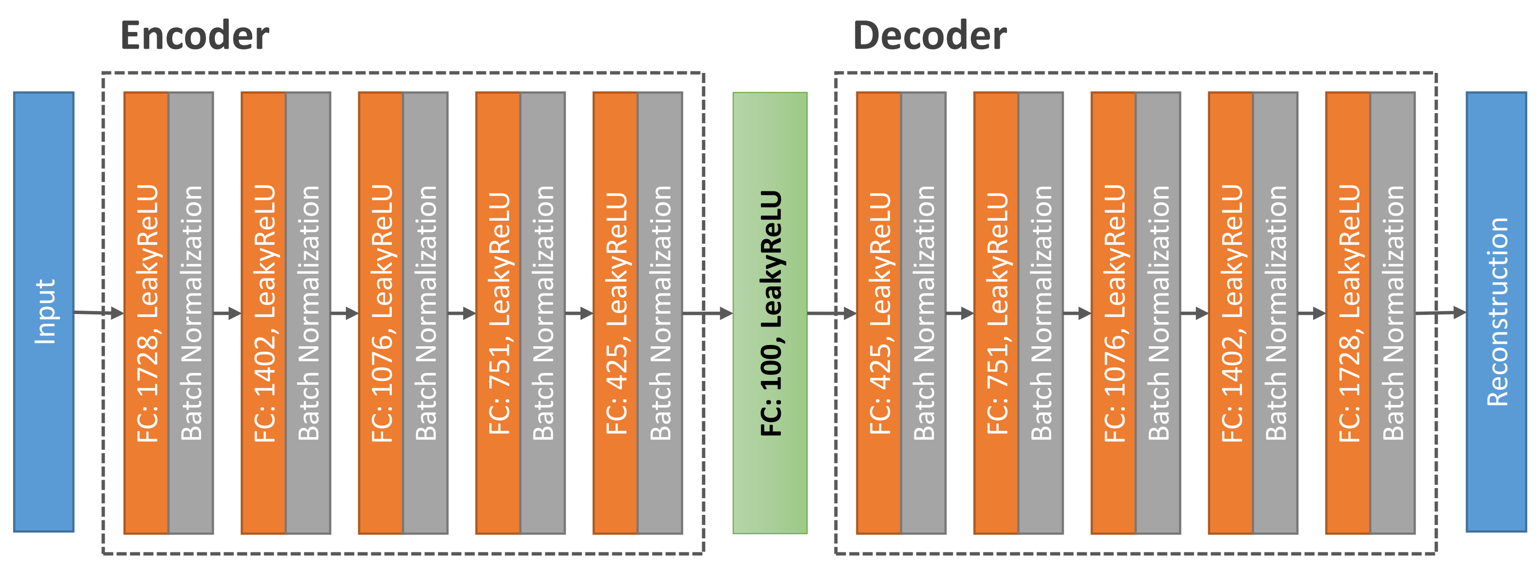
Multimodal Anomaly Detection based on Deep Auto-Encoder for Object Slip Perception of Mobile Manipulation Robots

Youngjae Yoo[1,2†], Chung-Yeon Lee[2,3†], Byoung-Tak Zhang[1,2]

AIIS[1], Seoul National University[2], Surromind Robotics[3]





Motivation

- **Slip perception** is an essential ability for manipulation robots to perform reliable tasks in the dynamic real-world.
- General approaches → - **Tactile**: expensive, disturbing grasp
  - **RGB**: background noise, object diversity

Anomaly detection by reconstruction of multisensory data using a deep auto-encoder can handle this problem.

Multimodal Anomaly Detection

- Sensory inputs are: transformed to 10Hz → synchronized to the nearest timestamp index → normalized to [0, 1]
- Combined and compressed by convolutional computations

Autoencoder-based Anomaly Scoring

- AE encodes the input data as the latent representation with significant features less affected by signal noise.
- We use the property that an AE trained with normal data cannot effectively compress and restore abnormal data.
- The reconstruction error of the AE is used as an indicator to detect whether the input data is normal or abnormal.

- Latent value of AE with input x

$$H_l(x) = f_{1:l}(x)$$
$$\hat{H}_l(x) = f_{1:l}(\hat{x}) = f_{1:l}(g(f(x)))$$

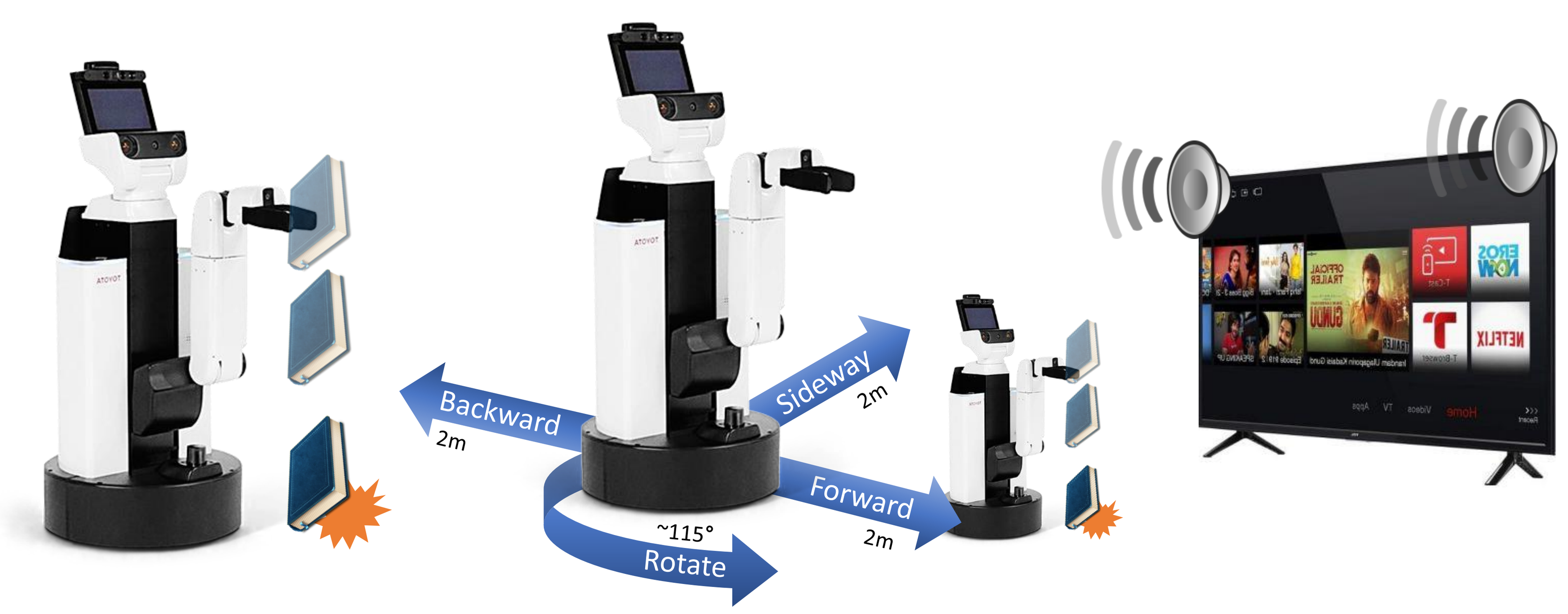
- Reconstruction error

$$d(x) = H(x) - \hat{H}(x)$$

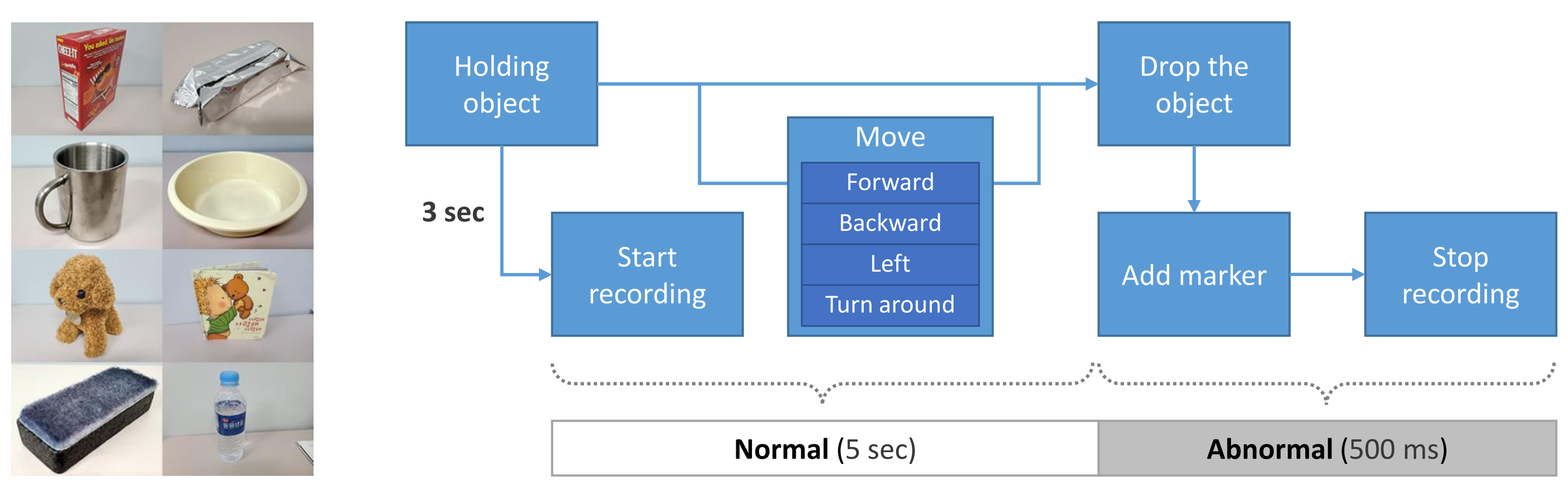
- Anomaly score (NAP)

$$S(x) = \|(d(x) - \mu)^T V \Sigma^{-1}\|_2^2$$

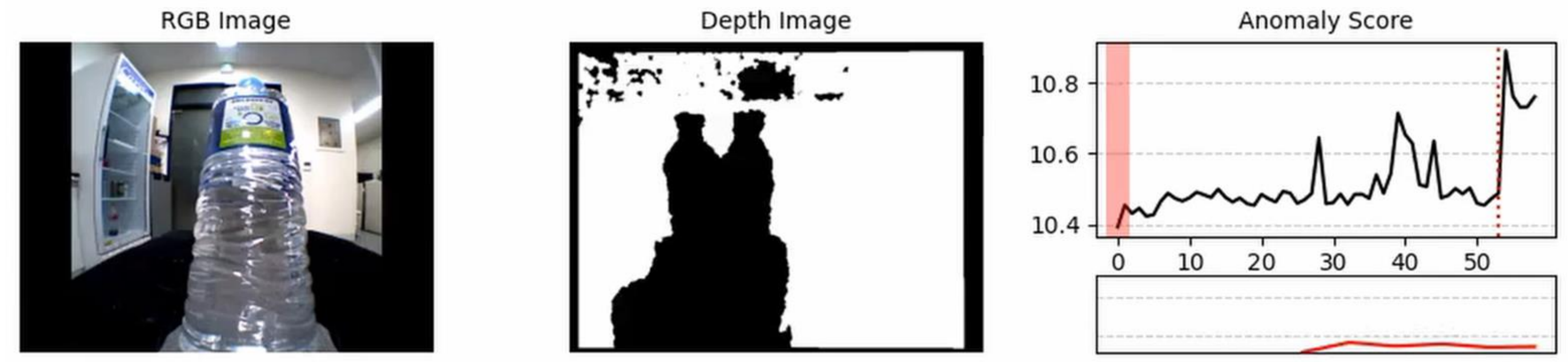
Experimental Setup



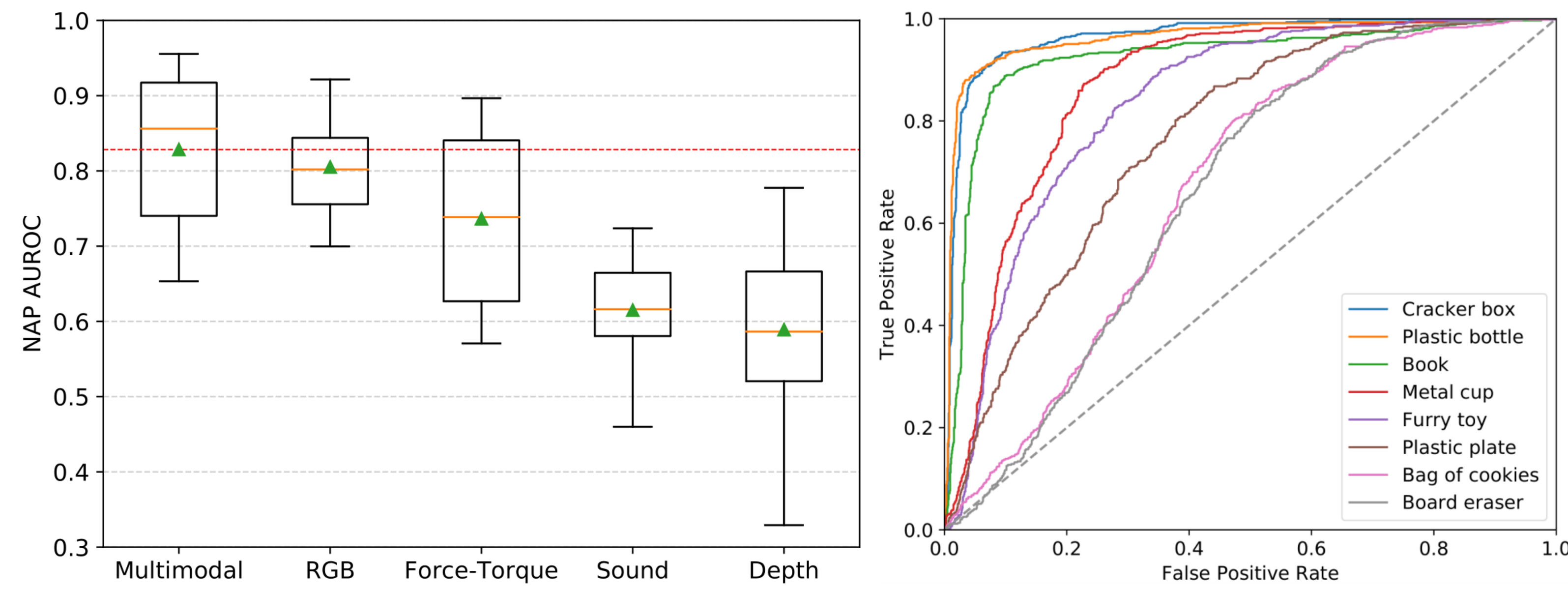
(a) Slip while standing (b) Slip while moving (c) Visual-auditory disturbances



Experimental Results



- Performance of anomaly detection is evaluated by comparing AUROC of the anomaly scores of different sensors or data.
- AUROC of multimodal data's scores outperformed scores of any other unimodal data, and it is varied for each object type.



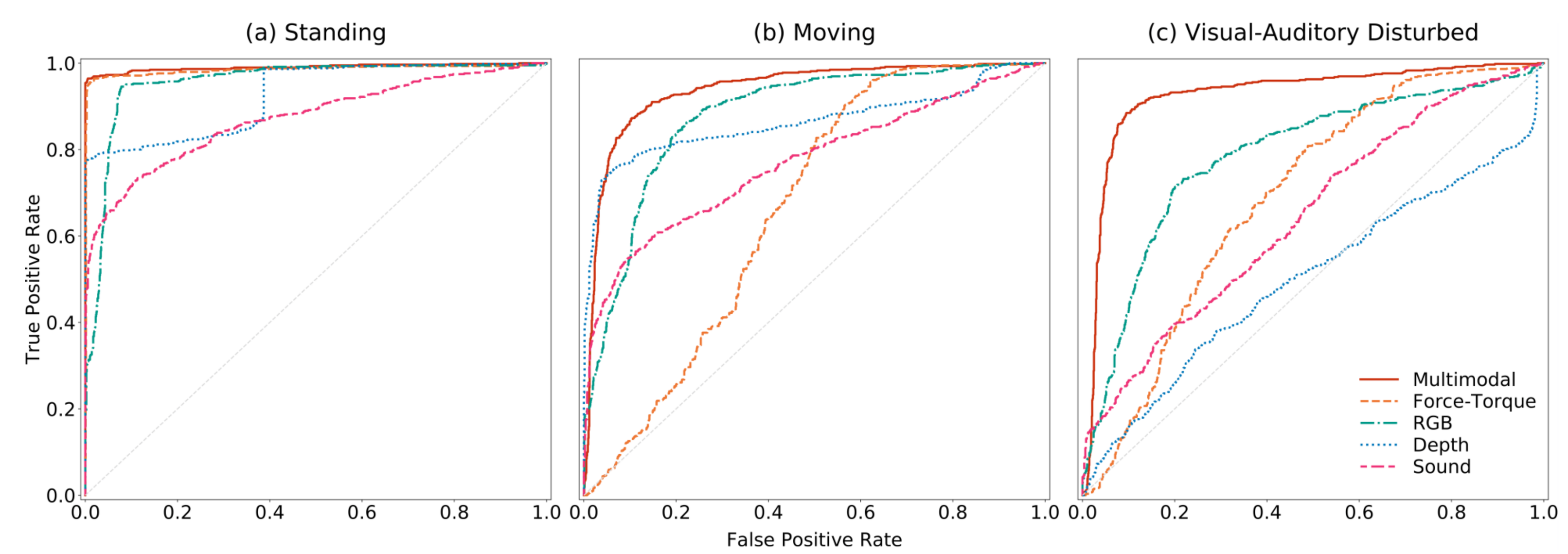
Comparison under different noise conditions

- The proposed multimodal approach showed more robust performance with a relatively small decrease in motion and audio-visual noise conditions compared to the unimodal data.

TABLE II

Comparison of model performance by multimodal and unimodal data under different experimental conditions and metrics

Sensors	AUROC			AUPRC			F1 Score		
	Standing	Moving	V.A.D.*	Standing	Moving	V.A.D.	Standing	Moving	V.A.D.
Multimodal	**0.9904**	**0.9323**	**0.9199**	**0.9883**	**0.8276**	**0.7865**	**0.8940**	**0.8188**	**0.8342**
Force-Torque	0.9867	0.6589	0.6681	0.9832	0.4006	0.4107	0.8891	0.2032	0.2173
RGB	0.9580	0.8762	0.7826	0.9096	0.7236	0.6616	0.8729	0.6559	0.5826
Depth	0.9309	0.8456	0.5207	0.9105	0.7747	0.3565	0.7747	0.7571	0.2227
MIC	0.9188	0.7264	0.6490	0.8884	0.6508	0.4884	0.7970	0.5486	0.3662



Spring Retreat of AIIS | Seoul, Korea | 30 April 2021