Facilitating Construction Scene Understanding Knowledge Sharing and Reuse via Lifelong Site Object Detection

Carnegie Mellon University
RIT Rochester Institute of Technology
Ruoxin Xiong, Yuansheng Zhu, Yanyu Wang, Pengkun Liu, Pingbo Tang*

Overview

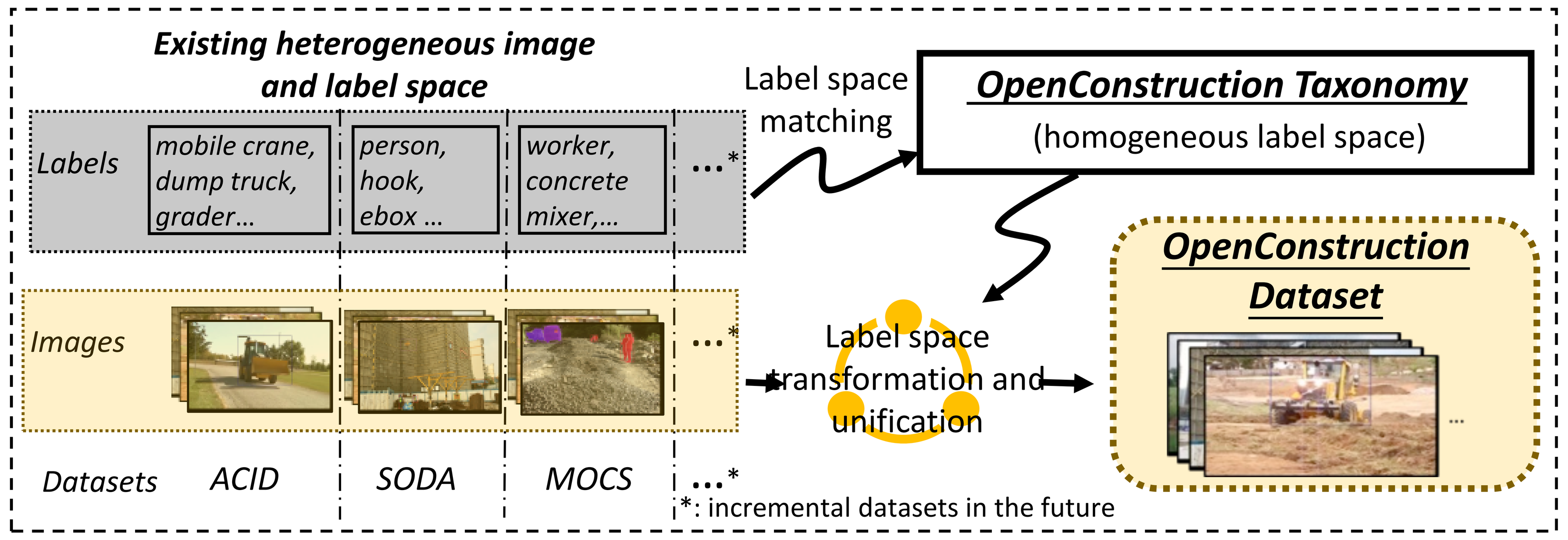
Automatically recognizing diverse construction resources (e.g., workers and equipment) supports intelligent workplace management. The existing learning systems are limited to fixed object categories and can hardly continuously learn new tasks due to the **"catastrophic forgetting"** problems.

- **Significant computational resources and storage space** are required to re-train the model each time and fully access all previous and new training data.
- As data is collected from different organizations, previous data may be unavailable due to **data privacy, cybersecurity concerns, and intellectual property rights.**

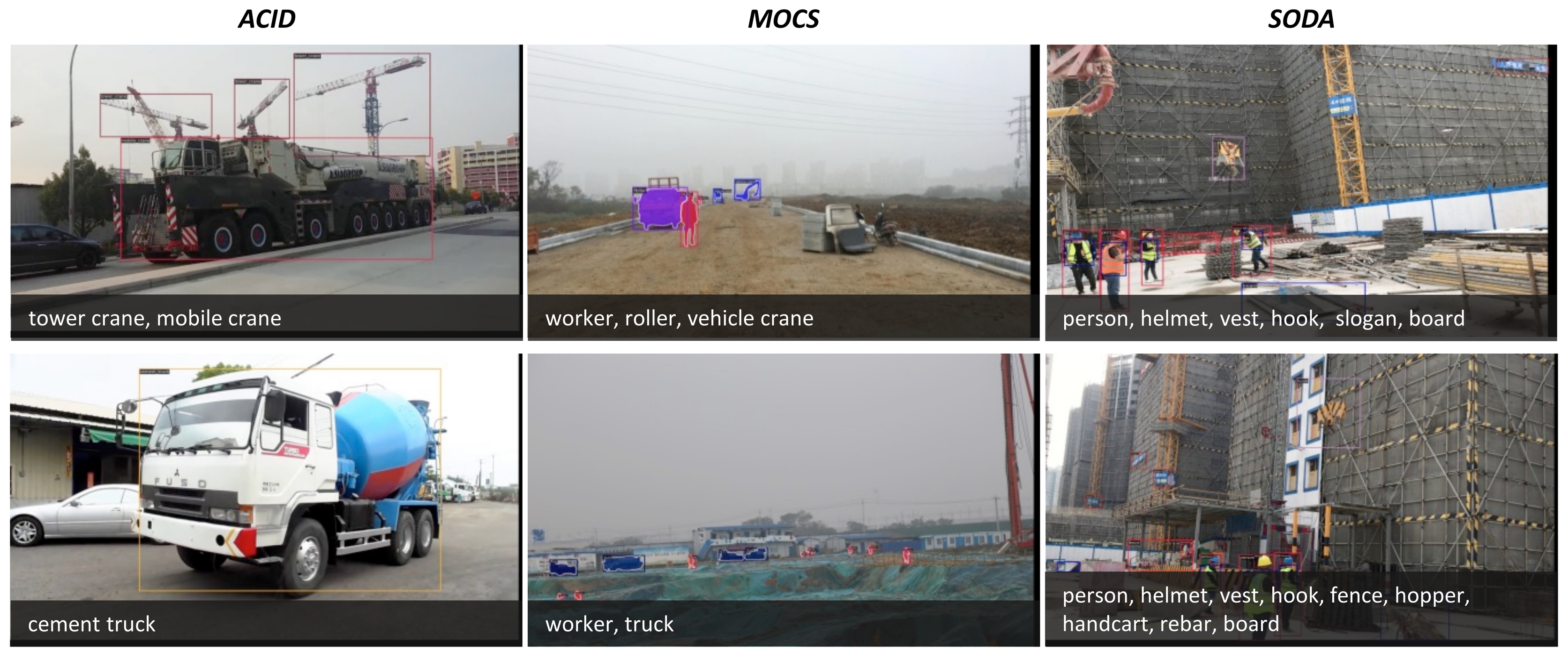
This work proposed a **novel lifelong construction resource detection framework** for continuously learning from new tasks and scenarios without catastrophically forgetting previous knowledge. **Codes and pre-trained models are at** https://github.com/YUZ128pitt/OpenConstruction

Lifelong Construction Resource Detection Benchmark

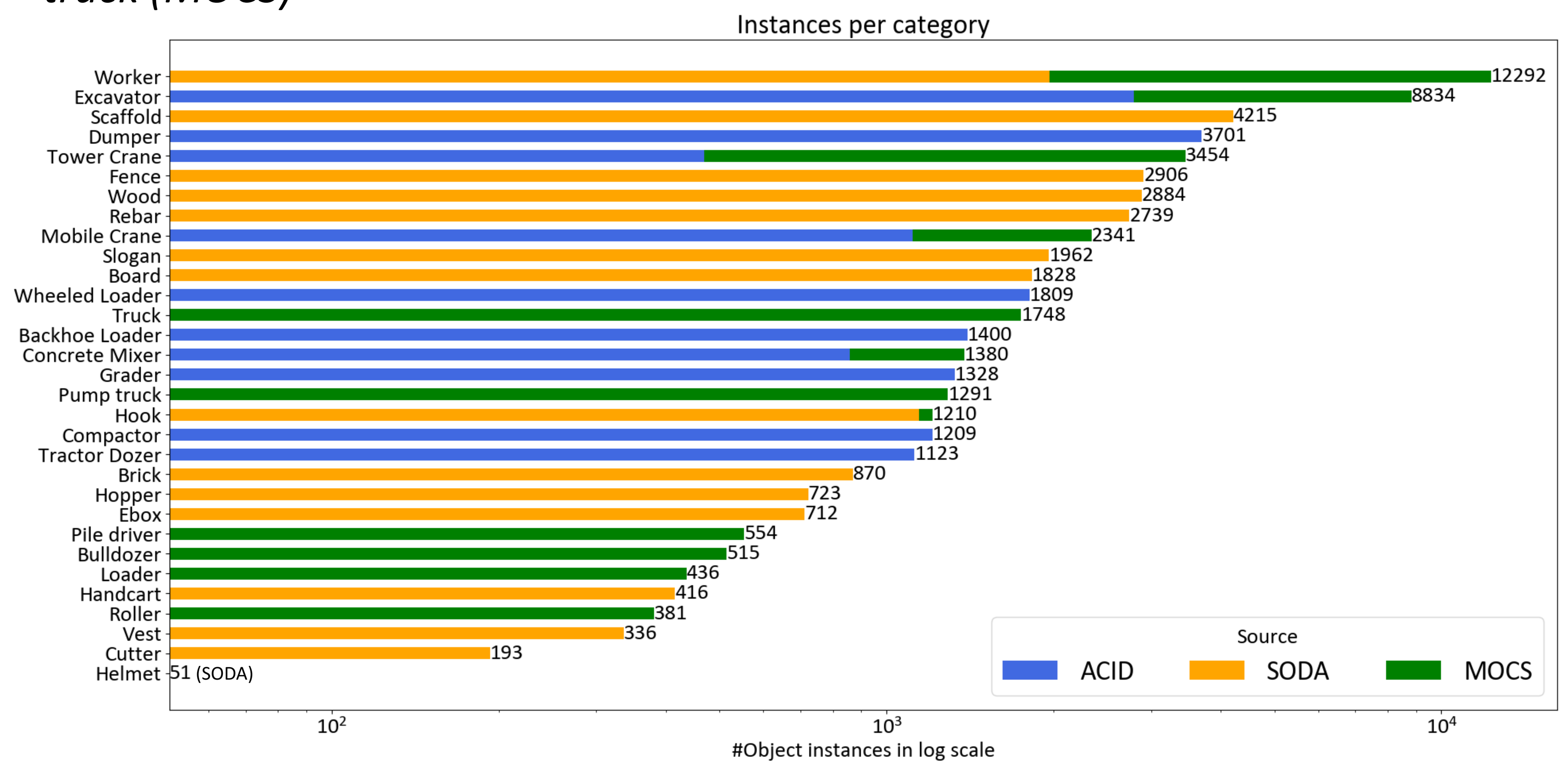
Current individual datasets are designed **for "static" evaluation protocols with limited object categories and scenarios.** Meanwhile, different datasets have **heterogeneous label space**.



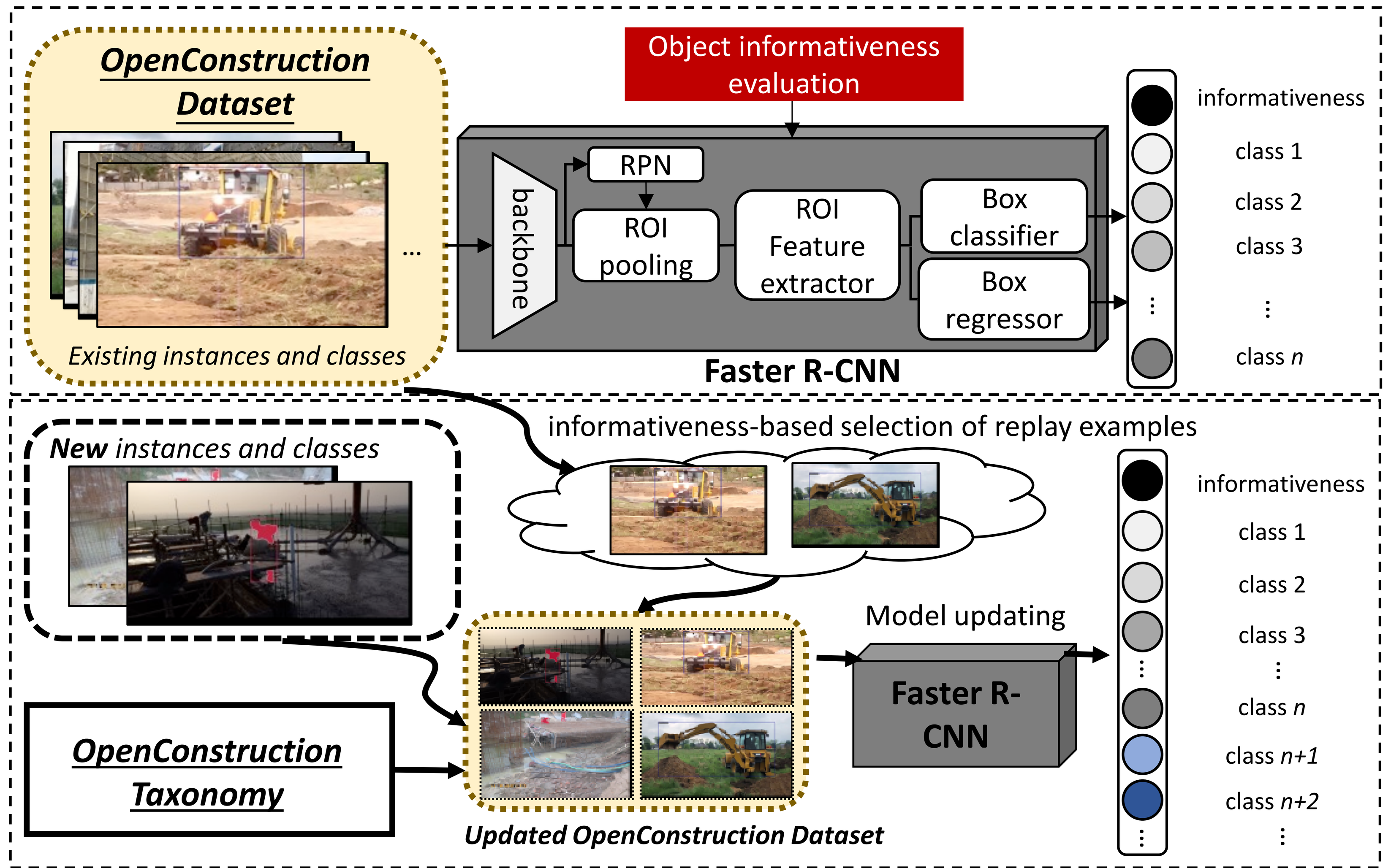
- This study proposed a **lifelong construction resource detection benchmark** by unifying three available datasets (ACID, SODA, and MOCS) in sequential settings.
- **OpenConstruction dataset** comprises **31,084 images and 64,841 object instances**, covering **31 diverse construction object categories** in the workspace.



The inconsistent and conflicting labels include *mobile crane (ACID) vs. vehicle crane (MOCS), worker (MOCS) vs. person (SODA), and cement truck (ACID) vs. truck (MOCS)*



Informativeness-based Lifelong Learning for Construction Resource Detection



Object informativeness evaluation

When learning a new task, we sort the known objects and retain the **top 30 images based on their informativeness.**

$$S^k = m(f_i, X^j)/N^j$$

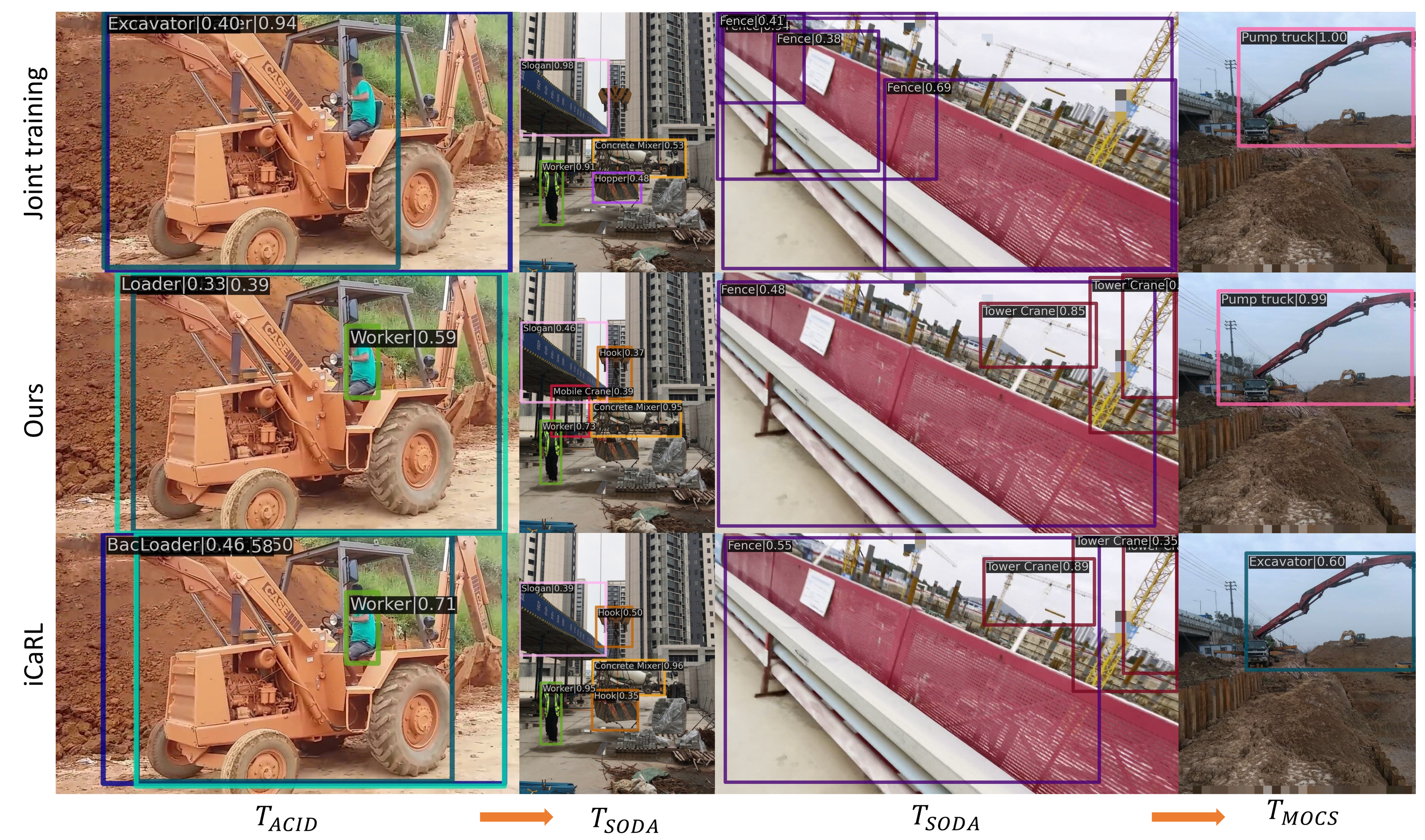
where k and j are the index of object instances and images, respectively. $m(f_i, X^j)$ is the performance evaluation metric for the model f_i on the image X^j. N^j is the number of object instances in image X^j.

Results

Our proposed method achieved a **good balance of maintaining previous knowledge and the ability to learn a new task.**

Methods	ACID	SODA	MOCS	Overall
Joint training	0.729	0.538	0.538	0.620
Non-adaptation	0.213	0.047	**0.650**	0.167
Fine-tuning	0.237	0.043	0.512	0.127
iCaRL [33]	**0.625**	0.204	0.502	0.358
Ours	0.555	**0.268**	0.566	**0.373**

Note: (1) Sequential settings: ACID→SODA→MOCS; (2) metric = mAP@[.5, .95]



Acknowledgements

We would like to thank the anonymous reviewers for their constructive comments. This material is based on work supported by Manufacturing Futures Institute and the Nuclear Engineering University Program (NEUP) of the U.S. Department of Energy (DOE) under Award No. DE-NE0008864. Funding to attend this conference was provided by the CMU GSA/Provost Conference Funding. The supports are gratefully acknowledged.