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

Ruoxin Xiong[1], Yuansheng Zhu[2], Yanyu Wang[1], Pengkun Liu[1], and Pingbo Tang[1(✉)]

[1] Carnegie Mellon University, Pittsburgh PA 15213, USA
{ruoxinx, yanyuwan, pengkunl, ptang}@andrew.cmu.edu
[2] Rochester Institute of Technology, Rochester NY 14623, USA
yz7008@rit.edu

Abstract. Automatically recognizing diverse construction resources (*e.g.,* workers and equipment) from construction scenes supports efficient and intelligent workplace management. Previous studies have focused on identifying fixed object categories in specific contexts, but they have difficulty accumulating existing knowledge while extending the model for handling additional classes in changing applications. This work proposes a novel lifelong construction resource detection framework for continuously learning from dynamic changing contexts without catastrophically forgetting previous knowledge. In particular, we contribute: (1) an OpenConstruction Dataset with 31 unique object categories, integrating three large datasets for validating lifelong object detection algorithms; (2) an OpenConstruction Taxonomy, unifying heterogeneous label space from various scenarios; and (3) an informativeness-based lifelong object detector that leverages very limited examples from previous learning tasks and adds new data progressively. We train and evaluate the proposed method on the OpenConstruction Dataset in sequential data streams and show mAP improvements on the overall task. Code is available at `https://github.com/YUZ128pitt/OpenConstruction`.

Keywords: construction site, object detection, common taxonomy, object informativeness, lifelong learning

1 Introduction

Construction scene images contain rich contextual information (*e.g.,* object location, categories, and relationships) for various construction resources, such as laborers, equipment, machines, tools, materials, and the environment, across different construction stages. Comprehensively monitoring and properly managing these construction resources across many different contexts contribute to the economy, quality, safety, and productivity of construction project performance [21,32]. Establishing computational methods that can continuously accumulate capabilities of recognizing new objects in updated contexts is critical for

supporting such cross-context management of various construction resources. Such methods should be able to keep learning from new data sets from new scenarios without losing the detection capability gained from previously used training data.

Over the last decade, the availability of ubiquitous on-site cameras and advanced deep neural networks (DNNs) have enabled the automatic scene understanding of construction activities from images and videos. Many researchers have explored deep learning-based methods for localizing and identifying construction site objects, including workers [14], building machinery and equipment [46], construction materials [24,40], and personal protective equipment [47,31], from large-scale domain-specific datasets. However, these deep learning-based models are limited to a fixed number of object categories and construction activities in certain types of spatio-temporal contexts. They can hardly retain and utilize the previously learned knowledge for adapting to new tasks due to the "catastrophic forgetting" or "catastrophic inference" problems [29]. For example, if new object classes or instances are required for new tasks, the training process should be restarted from the beginning each time, and all previous and new data has to be collected during the model training.

Real-world applications in construction scenarios require the learning models' abilities to learn new objects in new scenes while keeping the concepts learned from previous scenes. However, the existing learning systems fail to satisfy such requirements. The followings discuss the two main reasons: (1) significant computational resources and storage space are required to re-train the model each time and fully access all previous and new training data. This constraint could also hinder the applications of automation and robotic technologies in construction scenarios, such as construction robotics, drones, and autonomous vehicles, which require consistent and timely interactions with the surrounding scenes with limited computational abilities [38]; (2) as the data are collected and stored from different organizations, previous training data may be unavailable due to data loss, privacy or cybersecurity concerns, and intellectual property rights [43]. Therefore, a universal construction scene understanding model demands the ability to sustain knowledge of objects and concepts learned from previously encountered scenarios while learning new tasks (*e.g.,* new object instances and classes) - a lifelong/continuous learning model.

In this work, we develop a lifelong-based scheme that leverages the common concepts existing between different contexts and objects in those scenes to keep previously learned scene and object interpretation capabilities while learning new tasks. The developed approach enables a more efficient and scalable knowledge sharing and reuse of the concepts and objects in construction domains across different contexts. The main contributions of this work are:

- This study introduces a lifelong construction resource detection benchmark consisting of 31 object classes, 31,084 images, and 64,841 instances, namely the OpenConstruction Dataset, for identifying common concepts across scenes and validating lifelong object detection algorithms. We build such a dataset

by integrating three existing large datasets, ACID [46], SODA [14], and MOCS [7], in continuous data stream settings.
- This study proposes a common taxonomy that captures common concepts critical for detecting similar objects in different contexts or tasks, namely OpenConstruction Taxonomy, providing hierarchical representations for unifying duplicate, conflicting, and new label spaces. This taxonomy can also support multi-scenario information exchanges and inference with other information sources based on the unified label space.
- This study develops a new informativeness-based lifelong learning algorithm to accumulate previous knowledge and learn new construction objects in continuous data streams that keep bringing new objects and contexts. The experimental results tested on the OpenConstruction Dataset show that our proposed method performed better than state-of-the-art methods for adapting to new scenes while keeping previously learned construction scene understanding capability.

2 Related Work

2.1 Object Detection

Automatic construction resource detection (*e.g.,* worker, materials, machines, and tools) is fundamental for various operation-level applications, such as safety management [47,26], progress monitoring [44,13], productivity analysis [17,20,10], and material tracking [24,40]. The object detectors used for localizing and identifying these construction site objects can be generally classified into two groups: two-stage based methods, such as Fast/Faster R-CNN [16,36] and Cascade R-CNN [9], and one-stage based methods, such as Single Shot Detection (SSD) [27], You Only Look Once (YOLO) [34], and its variants [35,8]. In particular, two-stage object detectors first generate the region proposals by selective search [16] or a Region Proposal Network (RPN) [36], then classify the object classes for each region proposal. One-stage object detectors simultaneously predict the object bounding boxes and estimate class probabilities. Typically, the inference speed of one-stage detectors is faster than the two-stage detectors due to the single-network computation. However, the two-stage detectors can achieve better detection performance than the one-stage detectors.

2.2 Metadata Standards in Construction Domains

The semantic representations for construction scenes are subject to individual organizations and do not employ pre-defined and common taxonomy across various construction scenarios. The inconsistent and even conflicting label systems widely exist in the current datasets (*e.g.,* mobile crane (ACID [46]) vs. vehicle crane (MOCS [7]), and worker (MOCS [7]) vs. person(SODA [14])), hindering the knowledge sharing, reuse, and exchange in the construction domains. Many existing building classification standards, such as National Building Specification [5], MasterFormat [4], and OmniClass [6], are developed for organizing and

connecting construction information and specifications, while none of them can provide unified and consistent taxonomy for organizing construction site images in dynamic changing contexts. The classes listed in these standards have not covered various construction resources in the field. Similarly, existing construction data exchange standards, such as Industry Foundation Classes (IFC), also provide the schema of different building components [23]. However, they are not intended to classify available construction resources, such as the equipment and machines used in the workplace. Other industry standards and reports contain multiple construction resources but lack the classification taxonomy. For example, the Occupational Safety and Health Administration (OSHA) construction incidents investigation engineering reports [1] contains detailed illustrations of various incidents and accidents on the construction site and thus involves a wide range of construction resources. Since OSHA lists these reports separately, they do not have a classification taxonomy for the construction resources that appeared in these incidents. However, as the new instances and classes are increased continuously, an extensible taxonomy of construction resources is desired to organize massive and information-dense image data and support flexible data exchanges and reuse across various stakeholders [45].

2.3 Lifelong Learning

The parameters of trained machine learning models are usually fixed, limiting their abilities to handle new tasks or changing scenarios in real-world applications. In contrast, human beings can continuously learn and accumulate knowledge throughout their lives by forming and reusing concepts across scenes, enabling themselves to adapt to dynamic environments and new jobs. Inspired by the mechanism of human beings in concept formulation across different scenes, lifelong learning [41] aims to equip the machine learning models with concept reusing and adaptation across different scenes. Such algorithms can learn consecutive tasks without forgetting concepts encountered in previous contexts.

The bottleneck of lifelong learning is catastrophic forgetting, which refers to the phenomenon that the trained machine learning models tend to lose their previous knowledge when learning from new data due to the distribution shift between new and old training data. Current practices for mitigating catastrophic forgetting include parameter isolation, regularization-based techniques, replay techniques, or hybrid methods [29,12]. Particularly, as the knowledge is stored in the model parameters, parameter isolation methods prevent catastrophic forgetting by specifying networks' parameters for each task. For example, progressive neural networks [37] train columns of layers to execute a single task and fix them when learning other tasks. Instead of specifying the parameters, the regularization-based techniques, such as the Elastic Weights Consolidation (EWC) [22] prevent the parameters from changing too much via a regularizer such that the model could retain the previous knowledge. Replay techniques mix the previously seen samples with new samples and use the augmented data to retrain a model. In particular, the replay examples could either be a small subset of old data [28] or virtual samples derived from generative models [39].

3 Lifelong Construction Resource Detection Benchmark

3.1 Open-source Datasets for Detecting Construction Resources

Although many researchers have developed various datasets for detecting construction site objects, these individual datasets are designed for "static" evaluation protocols with limited object categories and scenarios. Typically, the data sources are acquired incrementally in real-world applications. This study proposed a new lifelong construction resource detection benchmark, namely the OpenConstruction Dataset, by integrating and unifying available datasets in sequential settings. Specifically, two main criteria are considered in selecting these data sources from the public datasets:

- *Data diversity*: The developed OpenConstruction Dataset aims to comprehensively integrate and cover various types of construction resources from public datasets in the community. For example, the three integrated large datasets have covered diverse categories of available open-source datasets.
- *Data quality*: We implemented the following three processes to ensure the data quality: (1) objects with tiny instance size ratios (less than 1.8%) are removed from the dataset. For example, some samples of workers and helmets in SODA [14] are very small and thus are deleted from the dataset. (2) Inaccurate annotations located out of the image areas were resized to the boundaries. (3) Class categories with unclear definitions were removed from the dataset, *e.g.,* "other vehicles" in the MOCS [7], to avoid ambiguous and conflicting detection results.

For constructing the OpenConstruction Dataset, we selected three large and diverse datasets, including ACID [46], SODA [14], and MOCS Dataset [7]. Descriptions and statistics of these open-source construction datasets are shown in Table 1. These diverse and heterogeneous datasets cover different construction resources from various scenarios. However, the label systems from these datasets are often duplicated, inconsistent, and even conflicting. Fig. 1 shows some examples of images and their inconsistent annotations in these three datasets. The following section will develop a common taxonomy for transforming and unifying these heterogeneous label systems.

3.2 Label Space Transformation and Unification

Taxonomy building. This study proposed a common and hierarchical taxonomy, namely OpenConstruction Taxonomy, for unifying duplicate, conflicting, and new object classes and building the mega-scale dataset of construction resources. Fig. 2 shows the procedures for building and determining the object categories in the OpenConstruction Taxonomy. Referred to [15], we constructed the domain-specific taxonomy in an iterative process: (1) identify and extract common concepts by reviewing related international standards and industry reports, such as ISO/TR 12603 building construction machinery and equipment [3]

Table 1: Data descriptions of three integrated construction resource datasets.

Datasets	Num. of categories	Object categories	# Images
ACID [46]	10	mobile crane, tower crane, cement truck, backhoe loader, wheel loader, compactor, dozer, dump truck, excavator, grader	10,071
SODA [14]	15	person, helmet, vest, board, wood, rebar, brick, scaffold, handcart, cutter, ebox, hopper, hook, fence, slogan	19,846
MOCS [7]	13	worker, static crane, hanging head, crane, roller, bulldozer, excavator, truck, loader, pump truck, concrete mixer, pile driving, other vehicle	41,668

Fig. 1: Examples of annotated images in three construction site object datasets. The inconsistent and conflicting labels across these overlapping examples include: *mobile crane* (ACID) vs. *vehicle crane* (MOCS), *worker* (MOCS) vs. *person* (SODA), and *cement truck* (ACID) vs. *truck* (MOCS).

and OSHA Construction Incidents Investigation Engineering Reports [1], as well as existing domain-specific datasets. (2) Connect these concepts in a hierarchical structure combining top-down and bottom-up strategies. A top-down strategy first identifies major construction resources (*i.e.,* equipment, human, machine, material, and tool) and follows down to specific subclasses. This top-down strategy can help avoid integrating redundant, conflicting, and inconsistent concepts into the Taxonomy. On the other hand, a bottom-up strategy merges the specific labels into general groups based on their affiliated relationships. The bottom-up strategy enables the extensibility of the developed Taxonomy by continuously absorbing new classes from real-world applications.

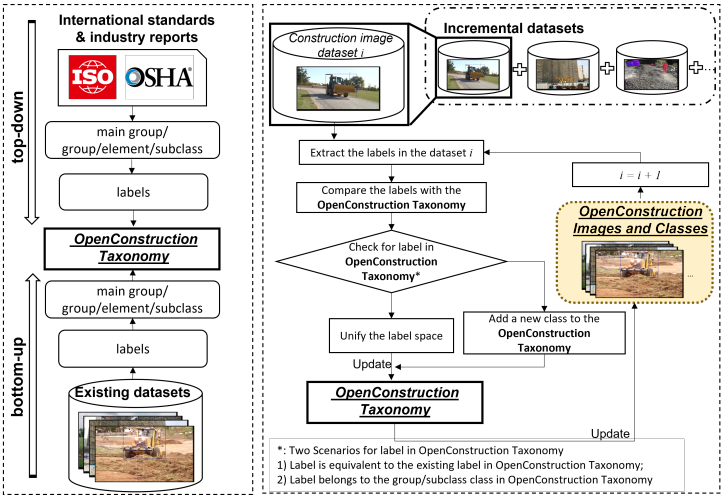

Fig. 2: Procedures for building the OpenConstruction Taxonomy.

Following the iterative top-down and bottom-up processes, we classify the diverse construction resources into four hierarchical levels: *main group* (level 1), *group* (level 2), *element* (level 3), and *subclass* (level 4), based on their specified purposes across various construction stages. The hierarchical structure of the OpenConstruction Taxonomy using the Protégé plugin [30], an ontology building and management system, is shown in Fig. 3.

Label space mapping. We manually transform and unify object labels assisted by the "string match" similarity using the spaCy tool [2]. However, as instances and classes steadily increase in new detection tasks, some new object categories may not be able to integrate into OpenConstruction Taxonomy in future applications. Therefore, we propose three extending steps that enable the continued growth of the OpenConstruction Taxonomy by analyzing their affiliated relationships with existing object categories:

- *Step 1: Link new subclasses.* The new concept is first examined as a new "subclass" using its affiliated relationship with the existing taxonomy. For example, for the new concept "wheeled backhoe loader", we add it as the new subclass of the "backhoe loader".
- *Step 2: Link new elements.* If a given concept cannot be linked through the "subclass", we query this element as new instances to its closest groups.
- *Step 3: Link new groups.* If no existing subclass and element are related to the new instance, we add it to the "group". For example, if a new concept is "personal protective equipment", we will integrate this class as the new group of "helmet" and "vest".

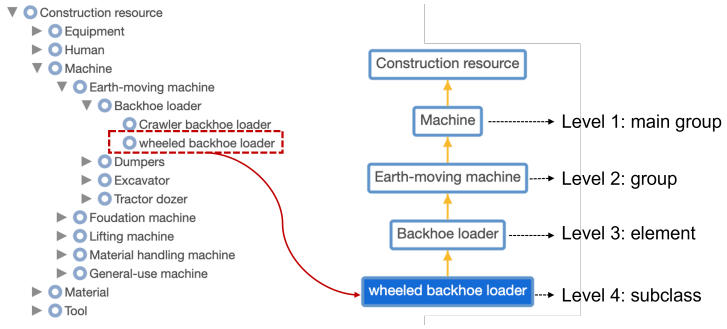

Fig. 3: Hierarchical structure of the OpenConstruction Taxonomy using the Protégé plugin [30]. The main groups consist of five categories: equipment, human, machine, material, and tool.

Dataset statistics. After a label space transformation and unification process, this study introduces a lifelong construction resource detection benchmark, namely the OpenConstruction Dataset. This dataset comprises 31,084 images and 64,841 object instances, covering 31 diverse construction object categories in the workspace, including workers, tools, machines, materials, and equipment. As these three datasets are collected individually, the unified OpenCoinstruction Dataset can serve as a benchmark for validating and testing lifelong object detection algorithms. Fig. 4 shows the data distributions and categories of the dataset. The results indicate that the developed dataset has a long tail distribution among object instances, categories, and instance sizes per image.

4 Informativeness-based Lifelong Learning for Construction Resource Detection

4.1 Preliminaries

Lifelong object detection setup. Lifelong object detection setting requires the model's abilities to accommodate streaming data with new labels and avoid catastrophic forgetting. We define the continuous learning task at time timestamp i as T_i. The developed approach can continuously learn from a sequence of tasks $\mathcal{T} = \{T_1, T_2, T_3, \cdots\}$. In the developed comprehensive OpenConstruction benchmark, the streaming data come from three data sources that different organizations collect at various times (ACID, SODA, MOCS). The developed approach for constructing the OpenConstruction benchmark $\mathcal{T}_{Open}$ continuously learns from the different datasets. The mathematical representation of the developing process is $\mathcal{T}_{Open} = \{T_{ACID}, T_{SODA}, T_{MOCS}\}$. The developed lifelong object detection approach can label both known and unknown objects. For example, when learning the T_{MOCS} after learning T_{ACID}, the model will encounter both the *worker* (new instances) and the *concrete mixer* (new class). We use the

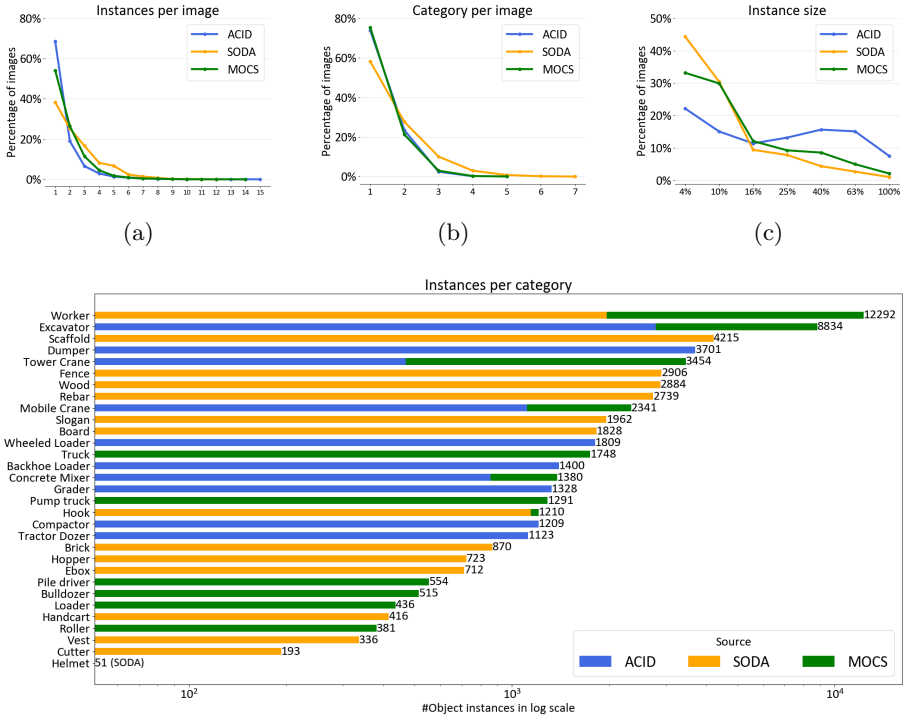

Fig. 4: Data distributions of the OpenConstruction Dataset. (a, b, c) distributions of the number of instances, categories, and instance size per image for ACID, SODA, and MOCS. (d) distributions of object instances in the OpenConstruction Dataset and their sources.

informativeness-based approach to avoid catastrophic forgetting in lifelong object detection. The details are illustrated in 4.2.

Model family. Fig. 5 shows the proposed informativeness-based lifelong learning framework for detecting construction resources incrementally. We use the Faster R-CNN [36], a two-stage object detector that consists of a backbone, an RPN, a region of interest (ROI) pooling, an ROI feature extractor, a bounding box regressor, and a classification head, as the basic detector. The box classifier and box regressor take the ROI features as the input and output the coordinates and class posterior probabilities of the bounding box.

4.2 Informativeness-based Lifelong Object Detector

We proposed a new training approach that enables lifelong object detection based on the Faster R-CNN [36] using streaming data. Previous approaches to training

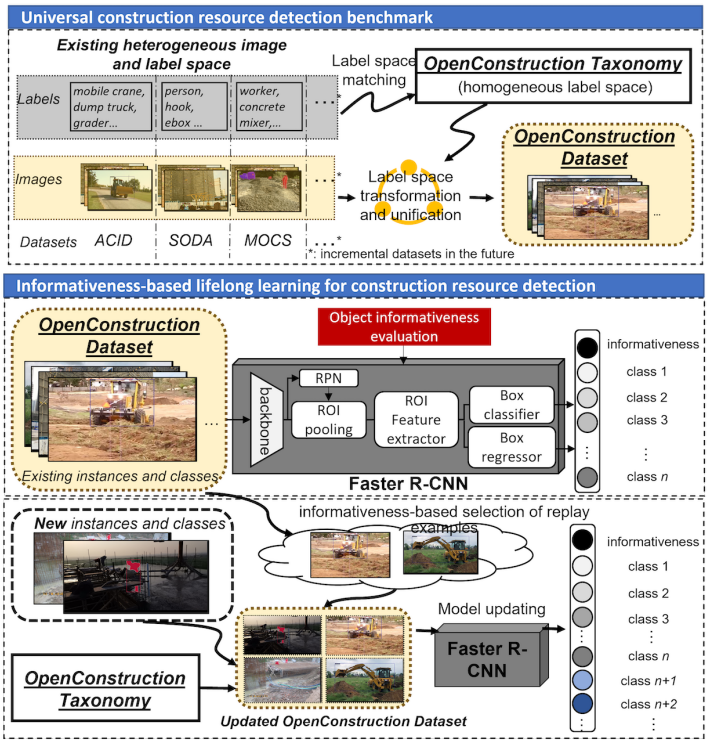

Fig. 5: Overview of the proposed informativeness-based lifelong learning framework for construction resource detection.

an object detection model require the availability of all training data simultaneously. The proposed training approach enables continuous learning based on Faster R-CNN through the following settings: 1) for the first task, we train the detector regularly; 2) for the subsequent tasks, we freeze the first three blocks of the Encoder and train the fourth block, the ROI, RPN, classification head, and box regressor. Using such settings, the adapted Faster-RCNN can learn new tasks while retaining knowledge from previous tasks. Furthermore, we design principle criteria to identify the desired relay examples to mitigate the catastrophic forgetting issue, which is a key challenge in lifelong learning tasks.

Replay examples selection. Fine-tuning the trained model in the new data inevitably impedes the model's ability to classify previously learned tasks, *i.e.*, catastrophic forgetting issues. Wang *et al.* [42]'s empirical results show that replay example-based methods outperform incremental fine-tuning and EWC methods in continuous object detection settings. The proposed informativeness-based lifelong learning approach uses the balanced memory bank [42,19] to keep instances from previous learning tasks and provides a principle to select the re-

playing instances. Previous studies select the old instances into the memory bank randomly [42,19]. The developed selection principle can better help the detector retains the knowledge of the corresponding task. Randomly selected replay examples may include non-representative instances (*i.e.,* objects that are partially hidden or objects in a complex background), resulting in less effectiveness for mitigating catastrophic forgetting.

We derive a simple yet efficient sampling function for evaluating the object informativeness, denoted as S^k, to address the inefficient instance selection challenge, as shown in the following: $S^k = m(f_i, X^j)/N^j$, where k and j is the index of object instances and image, respectively. f_i denotes the object detector which has learned task $\{T_1, T_2, \cdots, T_i\}$. m is a metric for evaluating the object detection performance of a model f on the image X^j, and mean average precision (mAP) is used in the implementation. N^j denotes the number of object instances that image X^j contains. This function encourages a selection of the simple background (*i.e.,* fewer objects), which will help reduce the replay examples where the objects not selected in the images are regarded as "background". When learning a new task, we sort the known objects and retain the top-30 objects based on their informativeness values.

Additionally, the object class distributions vary significantly, which could cause imbalanced issues for training object detectors. To balance the data distributions of the training dataset, we augment the stored and new data using the resampling strategy introduced by Gupta *et al.* [18] in the training process.

5 Experiments

5.1 Model Performance

To mimic the lifelong object detection task in real-world applications, we train and test all models on three individual datasets, *i.e.,* ACID [46], SODA [14], and MOCS [7], sequentially in three stages. Particularly, the new dataset is considered the new task, and the previous dataset is regarded as the old task for each training stage. Once a model is re-trained for learning a new task, the model weights stored in the previous task are used for subsequent training of the new task. This sequential training strategy can help evaluate the models' performance for continuously learning old and new tasks.

We use the standard mAP averaged for intersection over union (IoU)$\in$ [0.5: 0.05: 0.95], denoted as mAP@[.5, .95], to evaluate the proposed method. Four scenario settings are considered for model comparisons, all using the same object detector with different training strategies:

- *Joint training*: All instances and classes were added simultaneously for training an object detector in a typical training strategy, where all components of the objects are trainable.
- *Non-adaptation*: The model was trained in the same way as the *joint training* strategy, but the three datasets are sequentially exposed, and only the new data are used at each stage.

- *Fine-tuning*: Following the incremental fine-tuning in Wang *et al.* [42], we freeze the feature encoder after the first task and fine-tune the rest components of the detector (*i.e.,* RPN, box classifier, regressor).
- *iCaRL* [33]: *iCaRL* is proposed for incremental object classification. Following Wang *et al.* [42], we adapt *iCaRL* for lifelong object detection as a baseline for this study. On top of the *fine-tuning* strategy, this setting randomly selects the replay examples over each class and jointly trains the object detector with replay examples and new data.

Implementation details. We use the same object detector for ours and all baselines methods, a Faster-RCNN model that is pre-trained on the MS COCO dataset [25]. We use the MMDetection toolbox [11] to build the test platform, running it on two NVIDIA RTX A6000 GPUs.

Lifelong object detection results. Table 2 shows the mAP results on three individual tasks and the overall one, under off-line (*joint training*) and sequential training settings (*fine-tuning*, *non-adaptation*, and *iCaRL*). While the OpenConstruction dataset contains diverse construction scenes (31 object categories) and is highly imbalanced, the *joint training* strategy achieves more than 60% mAP, indicating that this unified dataset could also serve as a high-quality benchmark for detecting diverse construction resources in the civil engineering area.

Under the sequential learning settings, both *non-adaptation* and *fine-tuning* strategies get relatively low mAP on the first two tasks (T_{ACID} and T_{SODA}), particularly for the later one, meaning that they forget previously learned abilities to a large extend. On the other hand, because the T_{ACID} has more overlapped object categories with the last task than T_{SODA}, both methods retain some knowledge for learning this task. In terms of the final task, *non-adaptation* strategy achieves the highest mAP on the last task (T_{MOCS}), which even outperforms the *joint training* strategy by more than 10%. It could be explained that the *joint training* strategy learns the three tasks simultaneously, which is more complex than solely learning a single one. Both *non-adaptation* and *iCaRL* [33] strategies freeze the encoder after they learn the first task. Thus, they cannot learn new tasks well compared to trainable encoders (*i.e.,* Ours and *non-adaptation* strategy).

In particular, *iCaRL* [33] and ours store replay examples from previous tasks. The results of the first two tasks show that such a replaying technique can help mitigate the catastrophic problem. However, while the budget is set the same, ours achieves higher overall performance, especially for the T_{SODA}, indicating that our selected examples are more effective than randomly selected examples for retaining previous knowledge. Moreover, our method achieves the second best performance on the last task, owing to the trainable encoder. Finally, the overall performance shows the advantage of our approach, which could learn the new task well while retaining the ability to detect previous tasks. To summarize, our proposed method achieved a good balance of maintaining previous knowledge and the ability to learn a new task owning to the proposed selection strategy

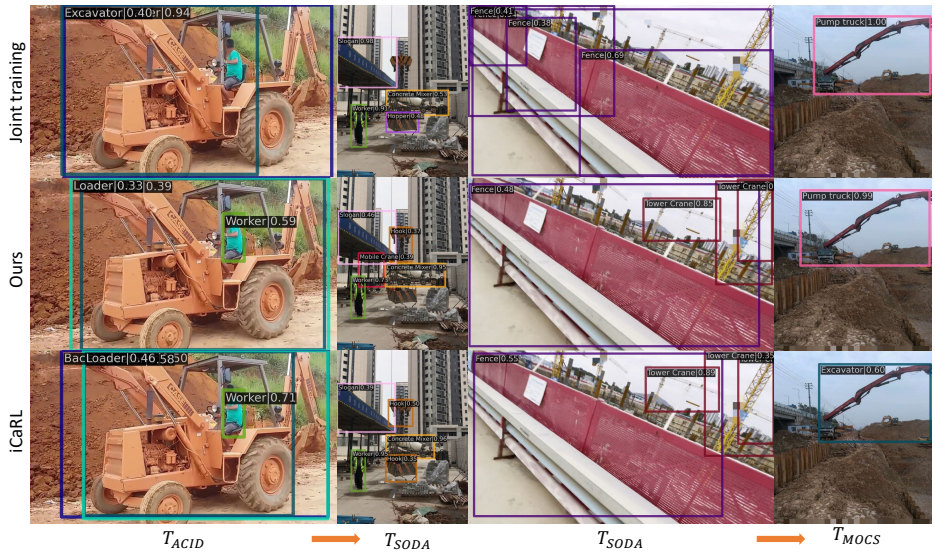

Fig. 6: Prediction examples on ACID, SODA, and MOCS for Joint training (top), Ours (middle), and iCaRL (bottom).

for replay examples and learnable encoder. Fig. 6 visualizes three test examples using the joint training strategy, our method, and iCaRL.

6 Conclusions

This paper proposed a novel scalable construction resource detection framework based on the lifelong learning scheme. Specifically, we construct a benchmark dataset with 31 unique object categories, namely the OpenConstruction Dataset, by integrating three large and heterogeneous datasets. To facilitate transforming and unifying heterogeneous label spaces, we develop a common taxonomy, namely the OpenConstruction Taxonomy. This taxonomy provides hierarchical annotations for linking inconsistent concepts actively. Finally, this study proposes a novel informativeness-based lifelong learning algorithm for continuous learning of new detection tasks by selecting the most informative examples in previous tasks. We train and evaluate the proposed method on three diverse datasets in continuous data streams, *i.e.,* ACID, SODA, and MOCS. The experimental results show that our proposed method achieved 0.373 mAP on the overall task, outperforming the other three training strategies, *i.e.,* non-adaptation, fine-tuning, and iCaRL, for continuous construction scene understanding. The proposed lifelong construction resource detection framework can support efficient and scalable construction scene understanding knowledge sharing and reuse in real-world applications.

Table 2: Lifelong object detection results on the OpenConstruction Dataset after sequential learning $\mathcal{T}_{Open} = \{T_{ACID}, T_{SODA}, T_{MOCS}\}$ (Metric = mAP@[.5, .95]).

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: joint training is not trained in a continuous setting, thus only serving as a reference model here.

This study also has some limitations and will be improved in the future. First, we will integrate more datasets to increase the image samples of existing object categories such as helmets and vests and add new classes. Second, the taxonomy development and extensions are mostly completed by manual transformations. Future work will examine automatic methods to build such a taxonomy. Finally, this study did not fully handle the situations where object classes contained in the previous training dataset while not included in the subsequent learning tasks and thus considered as "background". We will model this issue as an open-world detection problem [19] to improve the model performance in the future.

Acknowledgments. We would like to thank the anonymous reviewers for their constructive comments. This material is based on work supported by Carnegie Mellon University's Manufacturing Futures Institute, the Nuclear Engineering University Program (NEUP) of the U.S. Department of Energy (DOE) under Award No. DE-NE0008864, and Bradford and Diane Smith Graduate Fellowship. The supports are gratefully acknowledged.

References

1. Construction incidents investigation engineering reports. https://www.osha.gov/construction/engineering, (Accessed on 07/12/2022)
2. Industrial-strength natural language processing in python. https://spacy.io/, (Accessed on 07/12/2022)
3. ISO/TR 12603:2010(en), building construction machinery and equipment — classification. https://www.iso.org/standard/50886.html, (Accessed on 07/15/2022)
4. Masterformat - construction specifications institute. https://www.csiresources.org/standards/masterformat, (Accessed on 07/15/2022)
5. National building specification: Connected construction information. https://www.thenbs.com/, (Accessed on 07/15/2022)
6. Omniclass - construction specifications institute. https://www.csiresources.org/standards/omniclass, (Accessed on 07/15/2022)
7. An, X., Zhou, L., Liu, Z., Wang, C., Li, P., Li, Z.: Dataset and benchmark for detecting moving objects in construction sites. Automation in Construction **122**, 103482 (2021)
8. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
9. Cai, Z., Vasconcelos, N.: Cascade R-CNN: High quality object detection and instance segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(5), 1483–1498 (2019)
10. Chen, C., Zhu, Z., Hammad, A.: Automated excavators activity recognition and productivity analysis from construction site surveillance videos. Automation in construction **110**, 103045 (2020)
11. Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, J., Shi, J., Ouyang, W., Loy, C.C., Lin, D.: MMDetection: Open MMLab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155 (2019)
12. De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., Slabaugh, G., Tuytelaars, T.: A continual learning survey: Defying forgetting in classification tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(7), 3366–3385 (2022)
13. Dimitrov, A., Golparvar-Fard, M.: Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections. Advanced Engineering Informatics **28**(1), 37–49 (2014)
14. Duan, R., Deng, H., Tian, M., Deng, Y., Lin, J.: SODA: Site object detection dataset for deep learning in construction. arXiv preprint arXiv:2202.09554 (2022)
15. El-Gohary, N.M., El-Diraby, T.E.: Domain ontology for processes in infrastructure and construction. Journal of Construction Engineering and Management **136**(7), 730–744 (2010)
16. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1440–1448 (2015)
17. Gong, J., Caldas, C.H.: An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations. Automation in Construction **20**(8), 1211–1226 (2011)
18. Gupta, A., Dollar, P., Girshick, R.: LVIS: A dataset for large vocabulary instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5356–5364 (2019)

19. Joseph, K., Khan, S., Khan, F.S., Balasubramanian, V.N.: Towards open world object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5830–5840 (2021)
20. Kim, H., Bang, S., Jeong, H., Ham, Y., Kim, H.: Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation. Automation in Construction **92**, 188–198 (2018)
21. Kim, J.: Visual analytics for operation-level construction monitoring and documentation: State-of-the-art technologies, research challenges, and future directions. Frontiers in Built Environment **6**, 575738 (2020)
22. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences **114**(13), 3521–3526 (2017)
23. Laakso, M., Kiviniemi, A.: The IFC standard - a review of history, development, and standardization. Journal of Information Technology in Construction **17**, 134–161 (2012)
24. Li, Y., Lu, Y., Chen, J.: A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector. Automation in Construction **124**, 103602 (2021)
25. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. In: European Conference on Computer Vision. pp. 740–755 (2014)
26. Liu, J., Luo, H., Liu, H.: Deep learning-based data analytics for safety in construction. Automation in Construction **140**, 104302 (2022)
27. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: SSD: Single shot multibox detector. In: European Conference on Computer Vision. pp. 21–37 (2016)
28. Lopez-Paz, D., Ranzato, M.A.: Gradient episodic memory for continual learning. In: Advances in Neural Information Processing Systems. vol. 30 (2017)
29. Menezes, A.G., de Moura, G., Alves, C., de Carvalho, A.C.: Continual object detection: A review of definitions, strategies, and challenges. arXiv preprint arXiv:2205.15445 (2022)
30. Musen, M.A.: The protégé project: A look back and a look forward. AI Matters **1**(4), 4–12 (2015)
31. Nath, N.D., Behzadan, A.H., Paal, S.G.: Deep learning for site safety: Real-time detection of personal protective equipment. Automation in Construction **112**, 103085 (2020)
32. Pham, H.T., Rafieizonooz, M., Han, S., Lee, D.E.: Current status and future directions of deep learning applications for safety management in construction. Sustainability **13**(24), 13579 (2021)
33. Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H.: iCaRL: Incremental classifier and representation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2001–2010 (2017)
34. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 779–788 (2016)
35. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
36. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems. vol. 28 (2015)

37. Rusu, A.A., Rabinowitz, N.C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., Pascanu, R., Hadsell, R.: Progressive neural networks. arXiv preprint arXiv:1606.04671 (2016)
38. Shaheen, K., Hanif, M.A., Hasan, O., Shafique, M.: Continual learning for real-world autonomous systems: Algorithms, challenges and frameworks. arXiv preprint arXiv: 2105.12374 (2021)
39. Shin, H., Lee, J.K., Kim, J., Kim, J.: Continual learning with deep generative replay. In: Advances in Neural Information Processing Systems. vol. 30 (2017)
40. Son, H., Kim, C., Hwang, N., Kim, C., Kang, Y.: Classification of major construction materials in construction environments using ensemble classifiers. Advanced Engineering Informatics **28**(1), 1–10 (2014)
41. Thrun, S.: Lifelong learning: A case study. Tech. rep., Dept of Computer Science Carnegie Mellon University Pittsburgh PA (1995)
42. Wang, J., Wang, X., Shang-Guan, Y., Gupta, A.: Wanderlust: Online continual object detection in the real world. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10829–10838 (2021)
43. Wang, Y., Tang, P., Liu, K., Cai, J., Ren, R., Lin, J.J., Cai, H., Zhang, J., El-Gohary, N., Berges, M., Fard, M.G.: Characterizing perceived data sharing barriers and promotion strategies in civil engineering. In: Computing in Civil Engineering 2021, pp. 42–49 (2021)
44. Wang, Z., Zhang, Q., Yang, B., Wu, T., Lei, K., Zhang, B., Fang, T.: Vision-based framework for automatic progress monitoring of precast walls by using surveillance videos during the construction phase. Journal of Computing in Civil Engineering **35**(1), 04020056 (2021)
45. Wei, Y., Akinci, B.: Construction scene parsing (CSP): Structured annotations of image segmentation for construction semantic understanding. In: International Conference on Computing in Civil and Building Engineering. pp. 1152–1161. Springer (2020)
46. Xiao, B., Kang, S.C.: Development of an image data set of construction machines for deep learning object detection. Journal of Computing in Civil Engineering **35**(2) (2021)
47. Xiong, R., Tang, P.: Pose guided anchoring for detecting proper use of personal protective equipment. Automation in Construction **130**, 103828 (2021)