ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Towards Deep Continual Workspace Monitoring: Performance Evaluation of CL Strategies for Object Detection in Working Sites

Aslı Çelik[1], Oğuzhan Urhan[1], Andrea Cossu[2], Vincenzo Lomonaco[2] *

1- Kocaeli University ; 2- University of Pisa

Abstract. Object detection plays a crucial role in computer-based monitoring tasks, where the adaptability of object detection algorithms to complex and dynamic backgrounds is essential for achieving accurate and stable detection performance. Despite the effectiveness of state-of-the-art object detectors, continual object detection remains a significant challenge in real-world applications. In this study, we utilized a dataset tailored for continual object detection in diverse working environments. Using this dataset, a task-incremental and task-agnostic continual learning scenario was established in which each experience, corresponding to object detection sub-datasets collected from different work sites. Common baseline continual learning (CL) strategies were employed throughout the continual training process to evaluate their efficacy. Our findings, consistent with the CL literature, underscore replay-based strategies as the top performers, assessed across both task-aware and task-agnostic settings. Additionally, zero-shot object detection demonstrates notably lower performance compared to the best-performing CL strategies, emphasizing the critical importance of CL strategies in maintaining consistent detection performance and adapting to new environments and work sites.

1 Introduction

Computer vision-based monitoring systems have emerged as indispensable tools for enhancing safety and efficiency across diverse work environments, including construction sites, factories and production lines [1, 2, 3, 4]. Object detection is central to these systems, a fundamental component ensuring accurate and reliable monitoring. However, achieving robust detection results poses a significant challenge in environments characterized by complex backgrounds, such as warehouses and industrial sites. Moreover, the dynamic nature of workplaces demands adaptive algorithms capable of learning from a continuous data stream. Unlike traditional systems with predefined tasks, such as fixed locations or objects, an adaptive and dynamic workspace monitoring system must possess the agility to assimilate new information seamlessly. This capability ensures its efficacy in evolving environments where the introduction of new locations or objects is commonplace. Hence, the success of such systems hinges on their ability to adapt and learn continuously.

Ideally, continual learning models should progressively adapt to evolving data distributions over time without succumbing to catastrophic forgetting or

*Work supported by: EU EIC project EMERGE (Grant No. 101070918)

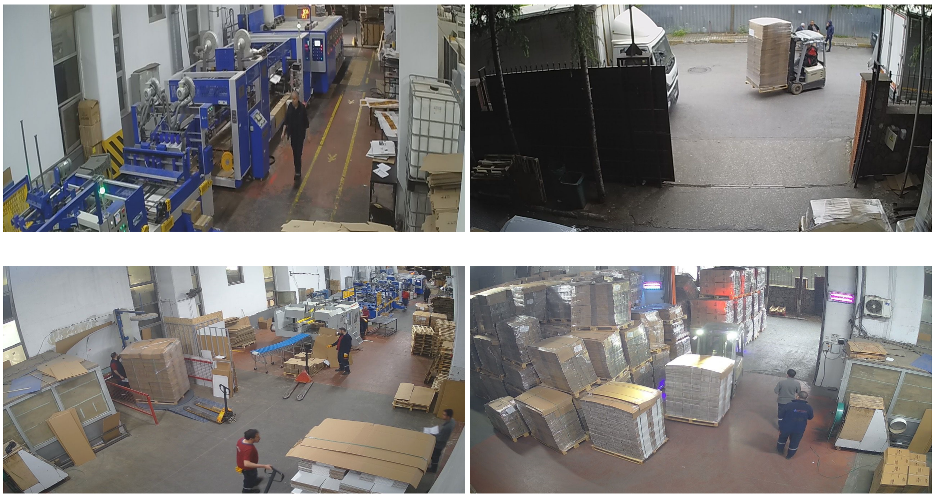

Fig. 1: Representative examples for continual object detection dataset

interference. While existing continual learning strategies promise to mitigate catastrophic forgetting, their application to real-world use cases remains relatively unexplored. Our study aims to bridge this gap by showcasing the necessity of continuous learning strategies in real-world computer vision tasks.

To this end, we curated an object recognition dataset consisting of frames taken from real-world work sites. The dataset includes two class categories: forklifts and people. Commonly used in warehouses, factories, and various work environments, forklifts play a central role in lifting and transporting loads. However, their versatility also raises safety concerns, as they are often associated with workplace accidents and injuries. To maintain and improve workplace efficiency while prioritizing safety, detecting and monitoring such vehicles, e.g., forklifts and personnel, is essential.

This work presents a task-incremental and task-agnostic continual learning scenario, in which each distinct experience corresponds to data collected from different working sites and warehouses. Several common baseline CL strategies were leveraged during continual training in order to perform an evaluation for this use case of object detection in working sites. As expected, the results obtained show that replay-based strategies perform best. We also study zero-shot generalization to unseen, novel sites and show that continual learning is helpful, clearly outperforming zero-shot prediction models trained on previous sites.

2 Related work

Object detection. Object detection is an integral part of computer vision-based monitoring systems. There is a trade-off between accuracy and efficiency in choos-

Table 1: Performance Evaluation of Task-Agnostic Continual Learning Strategies for Object Detection

Strategy	**Average Precision**
Naive	0.371 ± 0.24
EWC	0.409 ± 0.23
Replay	0.712 ± 0.12
Joint Training	0.817 ± 0.11

ing models and algorithms for low-resource use cases. One-step object detection algorithms coupled with lightweight, computationally inexpensive backbones are preferred for embedded platforms. MobileNetV3 [5] is a lightweight convolutional neural network designed for mobile and embedded devices. It balances model size, computational efficiency, and performance, making it suitable for resource-constrained environments. Single Shot Multibox Detector [6] is a single-stage detection algorithm that provides low computational cost, making it also a good candidate for embedded use.

Continual Learning. CL strategies are grouped into four categories: regularisation strategies, replay strategies, architectural strategies, and hybrid Strategies. Regularization strategies focus on preventing catastrophic forgetting by imposing constraints or penalties on model parameters during training. EWC [7] estimates the importance of parameters during continuous training and, with a penalty term, forces the model to change important parameters less for previous examples. Replay-based approaches store and replay past experiences to mitigate forgetting [8]. A common baseline, random replay, uses a predefined memory of size S to store samples, where T denotes the number of tasks; for each task, S/T number of examples are stored in memory for future training. Architectural strategies modify the model architecture to facilitate continual learning, examples include modular architectures or dynamically expandable networks. Multi-head architecture, another common baseline for Architectural Strategies, uses separate output heads for each task when task labels are provided.[9]

3 A dataset for continual object detection in working sites

To create a CL benchmark, we used real-world data from 20 different sites, including workplaces, warehouses, industrial sites, and factories. Cameras are placed in a fixed position to monitor a specific part of the site. Depending on the sites' size and changing monitoring needs, some sites are monitored with a single camera, while others are monitored with 2, 3, or even more cameras. Both human and forklift class instances were present in all subsets corresponding to different sites. For each site, 900 and 100 480×480 frames were used for training and testing, respectively. The benchmark presented consists of a total of 18000

Table 2: Performance Evaluation of Task-Aware Continual Learning Strategies for Object Detection

Strategy	Average Precision
Multi-Head Naive	0.366 ± 0.19
Multi-Head EWC	0.497 ± 0.20
Multi-Head Replay	0.714 ± 0.12
Joint Training	0.817 ± 0.11

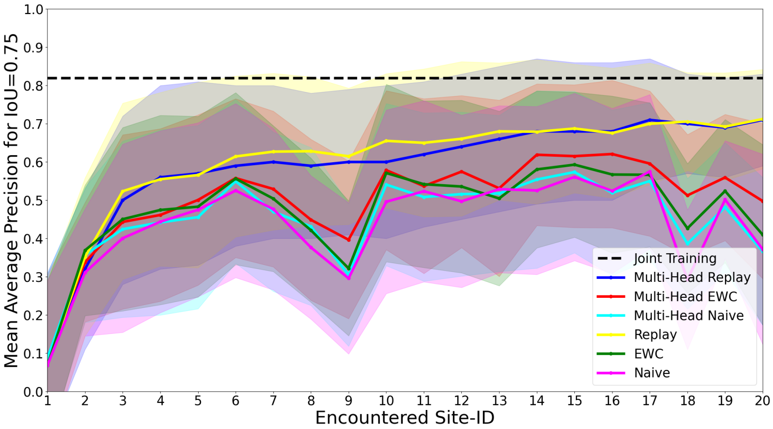

Fig. 2: mAP of CL Strategies Across All Sites During Continual Learning

frames for training and 2000 frames for testing. Sharing the dataset has not been possible because the necessary permissions have not been obtained. However, Figure 1 shows representative examples of the used dataset.

4 Experiments

We conducted experiments for the following CL strategies: EWC, random replay, and multi-head architecture. We employed Pytorch SSDlite model architecture with a MobileNetV3 Large backbone. We used the Avalanche library [10] to run all the experiments. The model is initialized with MS COCO [11] pre-trained weights available in Pytorch for all experiments. We run our experiments both in task-agnostic and task-aware settings. For multi-head architecture, a total of 20 classification and regression head pairs are used, and each head is initialized with the same COCO pre-trained weights. The same hyperparameters (lr=0.0005, momentum=0.9, weight decay=0.0005) were used throughout the continual training with SGD. Lambda parameter and memory size for EWC

Table 3: Zero-Shot Object Detection: Joint vs. Continual Training with Replay

Strategy	**Average Precision**
Joint	0.487 ± 0.17
CT with Replay	0.380 ± 0.18

and replay strategies were 100 and 200, respectively. In replay strategies, a mini-batch consists of current samples in addition to samples randomly selected from the memory buffer, such that there are the same number of samples for each experience in that mini-batch. Mean Average Precision (mAP) is used as the evaluation metric and it is calculated based on COCO [11] evaluation metrics and IoU (Intersection over Union) threshold is selected as 0.75.

Results. Tables I and II present the performance evaluation of common baseline CL strategies for task-agnostic and task-aware settings, respectively. The average precision for each strategy is reported as the mean final average precision of all test sets. The best performance is obtained with replay-based strategies for both settings. The results demonstrate that when the multi-head strategy is not coupled with other strategies, there is no performance gain over task-agnostic naive training. However, when coupled with EWC strategy, there is a performance gain compared to task-agnostic EWC and multi-head naive approaches. Figure 2 shows mAP performance metric during continual training. The dotted black line represents joint training, which represents the upper bound for any CL strategy. Naive and multi-head Naive refer to training without any strategy for task-agnostic and task-aware cases, respectively. The shaded area represents the standard deviation across runs.

Table 3 presents the results of zero-shot object detection when the model is trained on sites with ID between 1 and 15 and tested on sites with ID between 16 and 20 for joint training and continual training with replay strategy. The results indicate that while the zero-shot performance is comparable with the EWC and multi-head strategies, a significant performance gap exists compared to the best-performing replay-based strategies. Relying solely on zero-shot detection results in underperformance compared to fine-tuned CL models. These findings highlight the significance of CL strategies in preserving consistent detection performance while adapting to novel environments and extending detection capabilities to new work sites.

5 Conclusion

In this study, we used data collected from real-world working sites to create a CL scenario, focusing on task-incremental and task-agnostic setups where each experience corresponds to a different work site. Utilizing data collected from these diverse environments, our CL framework aimed to capture the complexity and variability inherent in practical applications. We compared the performance of common CL baseline strategies under both task-agnostic and task-aware settings. Our findings revealed that the replay-based strategy emerged as the most effective among the evaluated methods. Consistent with the CL literature, our findings suggest that replay strategies are simple and effective in mitigating forgetting. Furthermore, we compared zero-shot-detection performance on unseen sites with continual training strategies. Zero-shot detection results in inferior performance compared to fine-tuned continual learning models. Our study highlights the importance of CL in improving adaptive and robust object detection systems and in applying these strategies to real-world problems, even in cases where zero-shot prediction underperforms.

References

[1] Xiaochun Luo, Heng Li, Hao Wang, Zezhou Wu, Fei Dai, and Dongping Cao. Vision-based detection and visualization of dynamic workspaces. *Automation in Construction*, 104:1–13, 2019.

[2] Linhua Ye and Songhang Chen. Gbforkdet: A lightweight object detector for forklift safety driving. *IEEE Access*, 2023.

[3] Taofeek D Akinosho, Lukumon O Oyedele, Muhammad Bilal, Anuoluwapo O Ajayi, Manuel Davila Delgado, Olugbenga O Akinade, and Ashraf A Ahmed. Deep learning in the construction industry: A review of present status and future innovations. *Journal of Building Engineering*, 32:101827, 2020.

[4] Weili Fang, Peter ED Love, Hanbin Luo, and Lieyun Ding. Computer vision for behaviour-based safety in construction: A review and future directions. *Advanced Engineering Informatics*, 43:100980, 2020.

[5] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019.

[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016.

[7] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

[8] G. Merlin, V. Lomonaco, A. Cossu, A. Carta, and D. Bacciu. Practical Recommendations for Replay-Based Continual Learning Methods. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13374 LNCS:548–559, 2022.

[9] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3366–3385, 2021.

[10] Antonio Carta, Lorenzo Pellegrini, Andrea Cossu, Hamed Hemati, and Vincenzo Lomonaco. Avalanche: A PyTorch Library for Deep Continual Learning. *Journal of Machine Learning Research*, 24(363):1–6, 2023.

[11] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.