

IROS 2019 Lifelong Robotic Vision: Object Recognition Challenge

By Heechul Bae, Eoin Brophy, Rosa H.M. Chan, Baoquan Chen, Fan Feng, Gabriele Graffieti, Vidit Goel, Xinyue Hao, Hyonyoung Han, Sathursan Kanagarajah, Somesh Kumar, Siew-Kei Lam, Tin Lun Lam, Chuanlin Lan, Qi Liu, Vincenzo Lomonaco, Liang Ma, Davide Maltoni, German I. Parisi, Lorenzo Pellegrini, Duvidu Piyasena, Shiliang Pu, Qi She, Debdoot Sheet, Soonyong Song, Youngsung Son, Zhengwei Wang, Tomas E. Ward, Jianwen Wu, Meiqing Wu, Di Xie, Yangsheng Xu, Lin Yang, Qihan Yang, Qiaoyong Zhong, and Liguang Zhou

Humans have a remarkable ability to learn continuously from the external environment and inner experience. One of the grand goals of robots is to build an artificial “lifelong learning” agent that can shape a cultivated understanding of the world from the current scene and previous knowledge via an autonomous lifelong development. It is challenging for the robot learning process to retain earlier knowledge when robots encounter new tasks or information. Recent advances in computer vision and deep-learning methods have been impressive due to large-scale data sets, such as ImageNet [1] and COCO [2]. However, robotic vision poses unique new challenges for applying visual algorithms developed from these computer vision data sets because they implicitly assume a fixed set of categories and time-invariant task distributions [3].

Semantic concepts change dynamically over time [4]–[6]. For bridging the gap between robotic vision and stationary computer vision fields, we utilize a real robot mounted with multiple high-resolution sensors [e.g., monocular/red-green-blue-depth (RGB-D) from RealSense D435i, dual fisheye images from RealSense T265, and lidar; see Figure 1] to actively collect the data from the real-world objects in several kinds of typical scenarios, such as homes, offices, campuses, and malls.

Lifelong learning approaches can be divided into

- 1) regularization methods, e.g., Learning without Forgetting (LwF) [7], elastic weight consolidation (EWC) [8], and synaptic intelligence (SI) [9]
- 2) network expansion methods, e.g., context-dependent gating [10] and Dynamic Expandable Network [11]
- 3) rehearsal approaches with a sampling replay or generative mechanism to fit distribution from prior tasks [12]–[14], e.g., incremental classifier and representation learning [15], Deep Generative Replay (DGR) [16], and DGR with dual memory [17] and feedback [18].

This report summarizes the IEEE/RSJ International Conference on Intelli-

gent Robots and Systems (IROS) 2019 Lifelong Robotic Vision Competition (Lifelong Object Recognition Challenge) with the data set, rules, methods, and results from the top eight finalists (of over 150 teams) (Figure 2). Individual reports, data set information, rules, and released source codes can be found at the project home page [19].

Challenge Data Set and Rules

This challenge aimed to explore how to leverage the knowledge summarized from previous tasks for learning a new task efficiently as well as how previously learned tasks could be efficiently memorized in lifelong robotic vision. The goal of this competition was to test a model’s capability to continuously learn objects

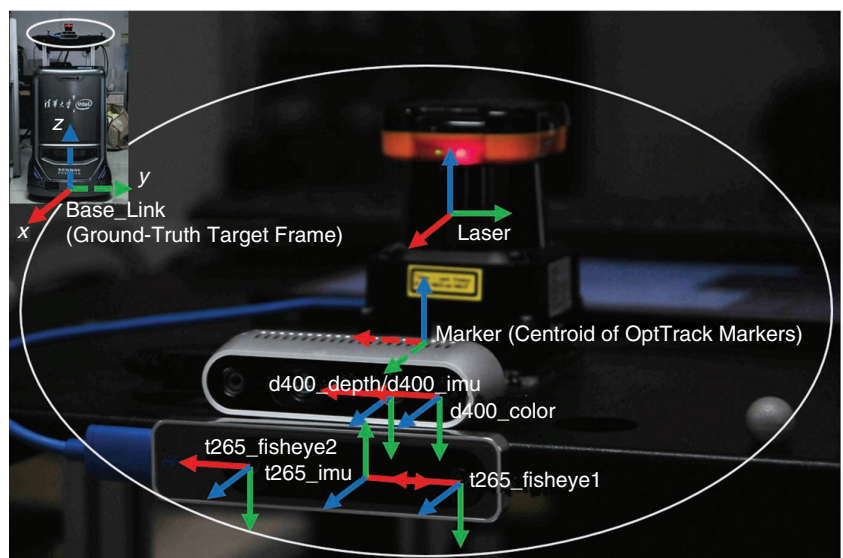


Figure 1. The OpenLORIS robotic platform (left) mounted with multiple sensors (right). In the OpenLORIS-Object data set, the RGB-D data are collected from the depth camera.

in a service robot scenario. Over 150 registered participants representing eight teams competed in the final testing phase. The work paved the way for robots to behave like humans in terms of knowledge transfer, association, and combination capabilities.

Challenge Data Set

The data set Lifelong Robotic Vision (OpenLORIS)-Object Recognition (OpenLORIS-Object) is designed to drive lifelong learning research and potential applications in the robotic vision domain, with everyday objects that exist in home, office, campus, and mall scenarios. The data set explicitly quantifies the variants of illumination, object occlusion, object size, camera-object distance/angles, and clutter information. The IROS 2019 competition organizers provided the first version of the OpenLORIS-Object data set for the participants.

Note that our data set has been updated with twice the size in content available at the project home page [20], including data set visualization, download instructions, and more benchmarks on state-of-the-art lifelong learning methods [21].

The competition data set is a collection of 69 instances, including 19 categories of daily necessities objects under seven scenes (see Table 1). For each instance, a 17-s video (at 30 frames per second) was recorded with a depth camera delivering 260 distinguishable chosen RGB-D frames. Four environmental factors, each with three level changes, are considered explicitly (Table 1). The data were divided into 12 sequential tasks by randomly sampling from different factors and levels. The organizers also provided a more challenging bonus test set that was recorded under different context backgrounds with some deformation and extreme view angles.

Challenge Rules

Rules are designed to quantify the learning capability of the robotic vision system when faced with the objects appearing in the dynamic environments. Different from a standard computer vision challenge, not only was the overall accuracy on all tasks evaluated; the model efficiency, including model size, memory cost, and replay size (the number of old task samples used for learning new tasks; smaller is better), was also considered (Table 2). Meanwhile, instead of directly asking the participants to submit the prediction results on the test data set as in standard deep learning challenges [1], [2], the organizers received either source or binary codes to evaluate their whole lifelong learning process to make a fair comparison. The finalists' methods were tested by the organizers on an Intel Core i9 CPU and a Nvidia RTX 1080 Ti graphics processing unit.

Challenge Methods and Results

The finalists and their results are summarized in Table 3, with the top result(s) in each category designated in bold. Details such as the report, slide, and poster of each solution can be found on the project home page [19]. With excellent participants and the solutions they presented, we anticipate that the resulting solutions can help robots perform well under dynamic environments.

HIK_LIG Team (Champion)

- **Title:** Dynamic Neural Network for Incremental Learning
- **Members:** Liang Ma, Jianwen Wu, Qiaoyong Zhong, Di Xie, and Shiliang Pu
- **Affiliation:** Hikvision Research Institute, Hangzhou, China.

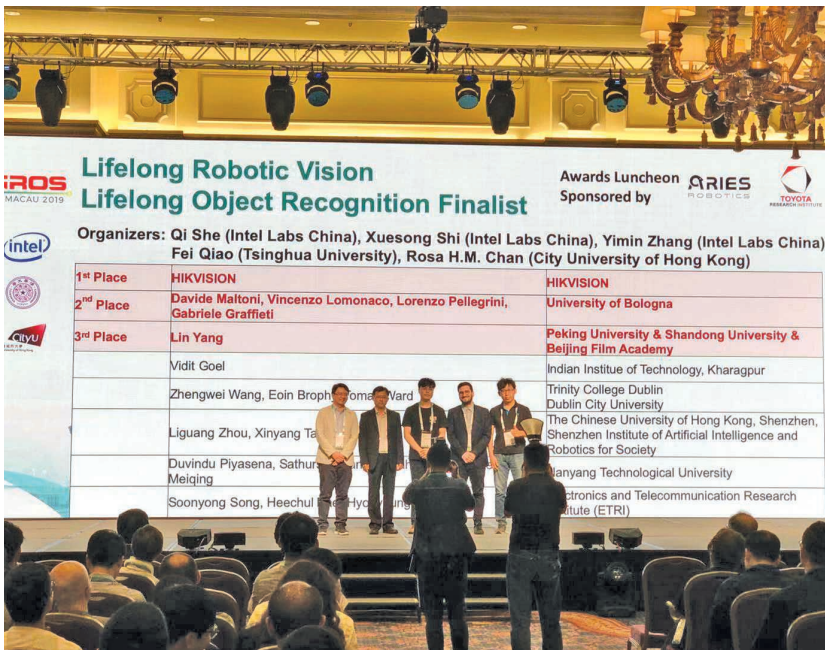


Figure 2. The Lifelong Robotic Vision Challenge finalists at IROS 2019.

Table 1. The details for each of the three levels of four real-life robotic vision challenges.

Level	Illumination	Occlusion (%)	Object Pixel Size	Clutter	Context	Classes	Instances
1	Strong	0	> 200 × 200	Simple	Home/office/ campus/mall	19	69
2	Normal	25	30 × 30 – 200 × 200	Normal			
3	Weak	50	< 30 × 30	Complex			

Table 2. The metrics and grading criteria.

Metric	Accuracy	Model Size	Inference Time	Replay Size	Oral Presentation	Accuracy on the Bonus Data Set
Weight	50%	8%	8%	8%	10%	16%

Table 3. The IROS 2019 Lifelong Robotic Vision Challenge final results.

Teams	Final Accuracy (%)	Model Size (MB)	Inference Time (s)	Replay Size (Number of Samples)	Bonus-Set Accuracy (%)
HIK_ILG	96.86	16.3	25.42	0	21.86
Unibo	97.68	5.9	22.41	1,500	8.5
Guinness	72.9	9.4	346	0	10.96
Neverforget	92.93	342.9	467.1	0	1.52
SDU_BFA_PKU	99.56	171.4	2,444	28,500	19.54
Vidit98	96.16	9.4	112.2	1,300	1.39
HYDRA-DI-ETRI	10.42	13.4	1,323	21,312	7.1
NTU_LL	93.56	467.1	4,213	0	2.1

ROBOTIC END-EFFECTORS

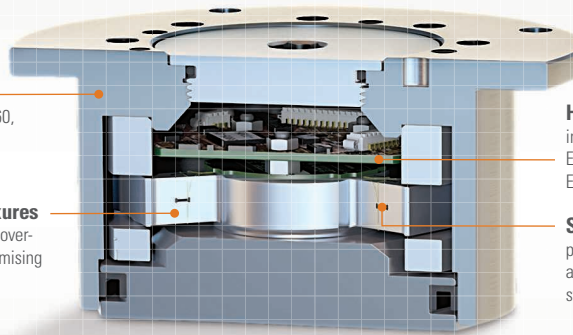
Measure all six components of force and torque in a compact, rugged sensor.

Interface Structure

high-strength alloy provides IP60, IP65, and IP68 environmental protection as needed

Sensing Beams and Flexures

designed for high stiffness and overload protection without compromising resolution



High-Speed Electronics

interfaces for Ethernet, PROFINET, EtherNet/IP, Analog, USB, CAN, EtherCAT, Wireless, and more

Silicon Strain Gages

provide high noise immunity, accuracy, and high factor-of-safety, standard on all F/T models

Engineered for high-performance and maximum stiffness, with the highest resolution and accuracy available, it's the ultimate force/torque sensor. Only from ATI.



www.ati-ia.com
919.772.0115

- **Method:** The team developed a dynamic neural network comprising two parts: dynamic network expansion for data across dissimilar domains and knowledge distillation for data in similar domains [Figure 3(a)]. The domain similarity was determined by the accuracy of the previous model before training on the current task.

Unibo Team (First Runners Up)

- **Title:** Efficient Continual Learning with Latent Rehearsal
- **Members:** Gabriele Graffieti, Lorenzo Pellegrini, Vincenzo Lomonaco, and Davide Maltoni
- **Affiliation:** University of Bologna, Italy
- **Method:** The team proposed a new lifelong learning approach based on latent rehearsal, namely, the replay of latent neural network activation

instead of raw images at the input level [see the architecture and corresponding Android application in Figure 3(b)]. The algorithm can be deployed on the edge with low latency. Details can be found in [22].

Guinness Team

- **Title:** Learning Without Forgetting Approaches for Lifelong Robotic Vision

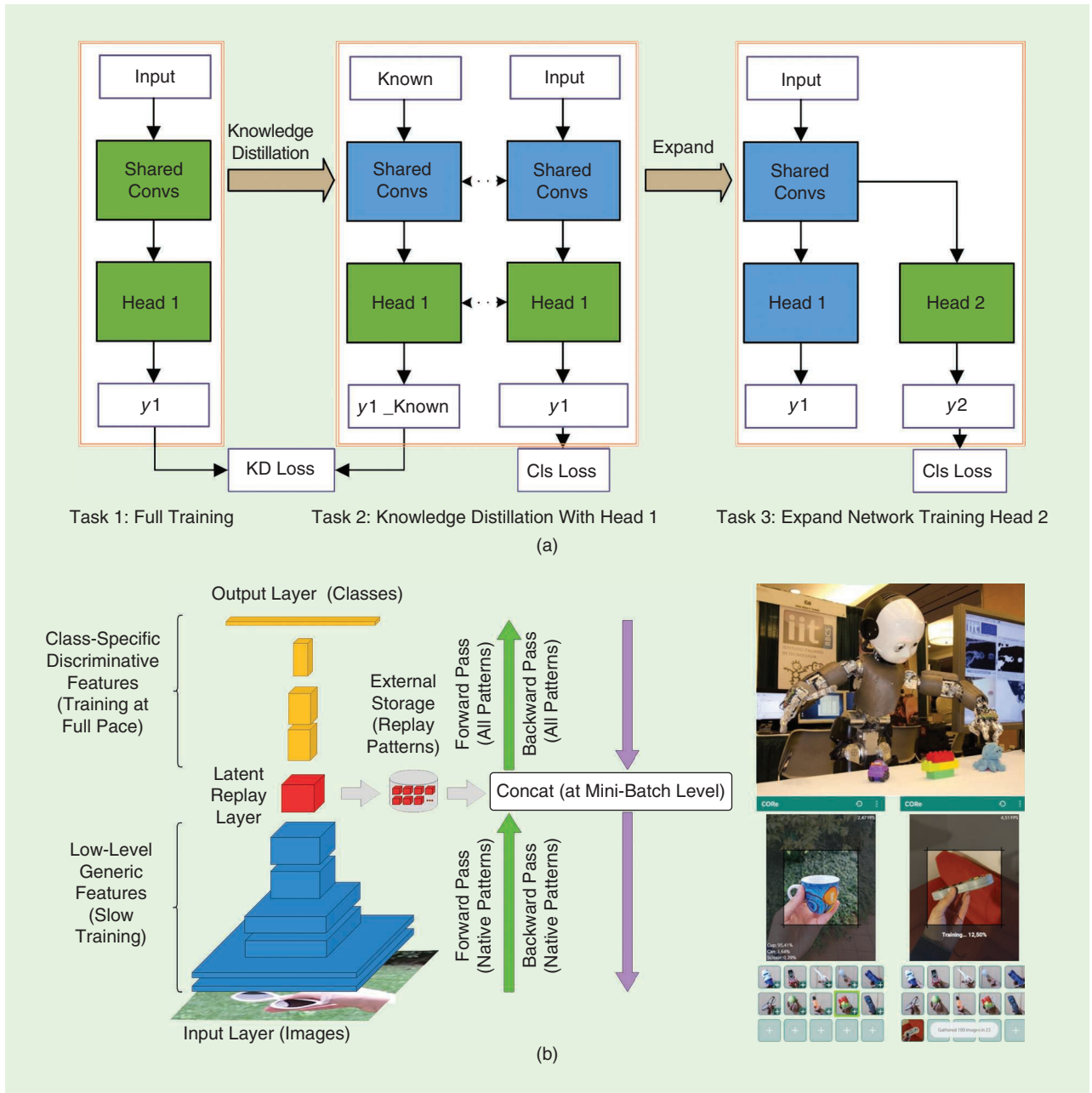


Figure 3. The challenge solutions. (a) HIK_LIG Team: dynamic network expansion for data across dissimilar domains and knowledge distillation for data in similar domains. (b) Unibo Team: replay of latent neural network activation instead of raw images at the input level (architecture and corresponding Android application). Cls: class; Convs: convolutions; KD: knowledge distillation. [(a) Source: HIK_LIG Team; used with permission; (b) source: Unibo Team; used with permission.]

- **Members:** Zhengwei Wang, Eoin Brophy, and Tomás E. Ward
- **Affiliations:** Wang: V-SENSE, School of Computer Science and Statistics, Trinity College, Dublin, Ireland; Brophy and Ward: Insight Center for Data Analytics, School of Computing, Dublin City University, Ireland
- **Method:** The core backend of the method was LwF [7]. There was no replay of previous task images in this structure.

Neverforget Team

- **Title:** A Small Step to Remember: Study of Single Model Versus Dynamic Model
- **Members:** Liguang Zhou, Tin Lun Lam, and Yangsheng Xu
- **Affiliation:** The Chinese University of Hong Kong, Shenzhen, China, and Shenzhen Institute of Artificial Intelligence and Robotics for Society, China
- **Method:** This approach was based on EWC [8] without a replay mechanism. The team also found the fact that the estimation of the Fisher information matrix might be biasedly estimated.

SDU_BFA_PKU Team

- **Title:** SDKD: Saliency Detection with Knowledge Distillation
- **Members:** Lin Yang and Baoquan Chen
- **Affiliation:** Peking University, Beijing, China; Shandong University, Qingdao, China; and Beijing Film Academy, Beijing, China
- **Method:** The approach disentangled this problem with two aspects: background removal problem and classification problem. The entrant used saliency maps to implement background removal and knowledge distillation to address catastrophic forgetting.

Vidit98 Team

- **Title:** Intelligent Replay Sampling for Lifelong Object Recognition
- **Members:** Vidit Goel, Debdoot Sheet, and Somesh Kumar
- **Affiliation:** Indian Institute of Technology, Kharagpur, India
- **Method:** This approach sampled validation data from the buffer and used them as replay data. It intelligently

created the replay memory for a task. The replay memory was an efficient representation of previous task data, whose information was lost and sampled from the validation set.

HYDRA-DI-ETRI Team

- **Title:** Selective Feature Learning with Filtering Out Noisy Objects in Background Images
- **Members:** Soonyong Song, Heechul Bae, Hyounyoung Han, and Young-sung Son
- **Affiliation:** Electronics and Telecommunications Research Institute, Korea
- **Method:** The team proposed a selective feature learning method to eliminate irrelevant objects in target images. A single-shot multibox detection (SSD) algorithm selected the desired objects [23]. The SSD algorithm alleviated performance degradation by noisy objects. Then SSD

weights were trained with annotated images in task 1 and the refined data were fed into a classification module.


NTU_LL Team

- **Title:** Lifelong Learning with Regularization and Data Augmentation
- **Members:** Duvinu Piyasena, Sathursan Kanagarajah, Siew-Kei Lam, and Meiqing Wu
- **Affiliation:** Nanyang Technological University, Singapore
- **Method:** The team utilized a combination of an SI-based regularization method [9] and data augmentation for each task.


Acknowledgments

This work was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China, under project CityU

Butterfly Haptics



Magnetic Levitation Haptic Interfaces



Highest fidelity interaction
for teleoperation and virtual
environments

<http://butterflyhaptics.com>

11215618. The authors would like to thank Hong Pong Ho from the Intel RealSense Team for the technical support of RealSense cameras for recording the high-quality RGB-D data sequences. The author list is in alphabetical order: R.H.M. Chan, F. Feng, X. Hao, C. Lan, Q. Liu, V. Lomonaco, G.I. Parisi, Q. She, and Q. Yang prepared the report. The other co-authors were competition finalists. The corresponding author is Qi She (qi.she@intel.com).

References

- [1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.
- [2] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. European Conf. Computer Vision (ECCV)*, 2014, pp. 740–755.
- [3] F. Feng, R. H. M. Chan, X. Shi, Y. Zhang, and Q. She, "Challenges in task incremental learning for assistive robotics," *IEEE Access*, vol. 8, pp. 3434–3441, Nov. 25, 2019. doi: 10.1109/ACCESS.2019.2955480.
- [4] Q. She and A. Wu, "Neural dynamics discovery via Gaussian process recurrent neural networks," in *Proc 35th Conf. Uncertainty Artificial Intelligence (UAI)*, 2019, p. 159.
- [5] Q. She, and R.H.M. Chan, "Stochastic dynamical systems based latent structure discovery in high-dimensional time series," in *Proc. 2018 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 886–890. doi: 10.1109/ICASSP.2018.8461755.
- [6] Q. She, Y. Gao, X. Kai, and R. H. M. Chan, "Reduced-rank linear dynamical systems," in *Proc. 32nd AAAI Conf. Artificial Intelligence (AAAI)*, 2018, pp. 4050–4057.
- [7] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, 2017. doi: 10.1109/TPAMI.2017.2773081.
- [8] J. Kirkpatrick et al., "Overcoming catastrophic forgetting in neural networks," *Proc. Natl. Acad. Sci. (PNAS)*, vol. 114, no. 13, pp. 3521–3526, 2017. doi: 10.1073/pnas.1611835114.
- [9] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *Proc. 34th Int. Conf. Machine Learning (ICML)*, 2017, pp. 3987–3995.
- [10] N. Y. Masse, G. D. Grant, and D. J. Freedman, "Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization," *Proc. Nat. Academy Sci. (PNAS)*, vol. 115, no. 44, pp. 467–475, 2018. doi: 10.1073/pnas.1803839115.
- [11] J. Yoon, E. Yang, J. Lee, and S. J. Hwang, "Life-long learning with dynamically expandable networks." 2017. [Online]. Available: arXiv:1708.01547
- [12] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks: A survey and taxonomy." 2019. [Online]. Available: arXiv:1906.01529
- [13] Z. Wang, Q. She, A. F. Smeaton, T. E. Ward, and G. Healy, "A Neuro-AI interface for evaluating generative adversarial networks." 2020. [Online]. Available: arXiv:2003.03193
- [14] Z. Wang, Q. She, A. F. Smeaton, T. E. Ward, and G. Healy, "Neuroscore: A brain-inspired evaluation metric for generative adversarial networks." 2019. [Online]. Available: arXiv:1905.04243
- [15] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "iCaRL: Incremental classifier and representation learning," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2001–2010.
- [16] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual learning with deep generative replay," in *Proc. Advances Neural Information Processing Systems (NIPS)*, 2017, pp. 2990–2999.
- [17] N. Kamra, U. Gupta, and Y. Liu, "Deep generative dual memory network for continual learning." 2017. [Online]. Available: arXiv:1710.10368
- [18] G. M. van de Ven and A. S. Tolias, "Generative replay with feedback connections as a general strategy for continual learning." 2018. [Online]. Available: arXiv:1809.10635
- [19] Q. She et al., "iROS 2019 Lifelong Robotic Vision Challenge—Lifelong object recognition report," Nov. 2019. Accessed on: Apr. 2020. [Online]. Available: <https://arxiv.org/abs/2004.14774>
- [20] Q. She, "OpenLORIS-object dataset and Benchmark," Nov. 2019. Accessed on: Apr. 2020. [Online]. Available: https://lifelong-robotic-vision.github.io/dataset/Data_Object-Recognition
- [21] Q. She et al., "OpenLORIS-Object: A robotic vision dataset and benchmark for life-long deep learning." 2020. [Online]. Available: arXiv:1911.06487v2
- [22] L. Pellegrini, G. Graffieti, V. Lomonaco, and D. Maltoni, "Latent replay for real-time continual learning." 2019. [Online]. Available: arXiv:1912.01100v2
- [23] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. European Conf. Computer Vision (ECCV)*, 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.



We want to hear from you!

Do you like what you're reading?
Your feedback is important.
Let us know—send the editor-in-chief an e-mail!

IEEE