

# Lifelong Learning via Progressive Distillation and Retrospection

Saihui Hou<sup>1</sup> [0000–0003–4689–2860], Xinyu Pan<sup>2</sup> [0000–0001–9951–4842],  
Chen Change Loy<sup>3</sup> [0000–0001–5345–1591], Zilei Wang<sup>1</sup> [0000–0003–1822–3731], and  
Dahua Lin<sup>2</sup> [0000–0002–8865–7896]

<sup>1</sup> Department of Automation, University of Science and Technology of China

<sup>2</sup> Department of Information Engineering, The Chinese University of Hong Kong

<sup>3</sup> Nanyang Technological University

**Abstract.** Lifelong learning aims at adapting a learned model to new tasks while retaining the knowledge gained earlier. A key challenge for lifelong learning is how to strike a balance between the preservation of old tasks and the adaptation to a new one within a given model. Approaches that combine both objectives in training have been explored in previous works. Yet the performance still suffers from considerable degradation in a long sequence of tasks. In this work, we propose a novel approach to lifelong learning, which tries to seek a better balance between preservation and adaptation via two techniques: Distillation and Retrospection. Specifically, the target model adapts to the new task by knowledge distillation from an intermediate expert, while the previous knowledge is more effectively preserved by caching a small subset of data for old tasks. The combination of Distillation and Retrospection leads to a more gentle learning curve for the target model, and extensive experiments demonstrate that our approach can bring consistent improvements on both old and new tasks<sup>4</sup>.

**Keywords:** Lifelong Learning, Knowledge Distillation, Retrospection.

## 1 Introduction

Lifelong learning aims at adapting a learned model to new tasks while retaining the knowledge acquired in the past. With the wide adoption of computer vision in real-world applications, there is an increasing demand for learning systems that are able to carry out lifelong learning over a series of tasks in a continual fashion. For example, a real-world object classification system is often required to be upgraded constantly by absorbing the knowledge from fresh domains. Directly repeating the training process with both previous and new data is often infeasible, due to various issues such as computation cost, storage budget, and privacy. For lifelong learning, a key challenge is to overcome the risk of catastrophic forgetting [9], namely a learned model usually suffers from accuracy degradation on old tasks when it adapts to a new one.

<sup>4</sup> Project page: <http://mmlab.ie.cuhk.edu.hk/projects/lifelong/>  
indicates joint first authorship.































