Dynamically Managing a Prompt Pool via Self-Enhancement in Continual Learning

Hayun Lee[1*] Kiseong Hong[2*]
Hwanhee Lee[2] Sungho Suh[3] Eunwoo Kim[1,2]

[1]School of Computer Science and Engineering, Chung-Ang University
[2]Department of Artificial Intelligence, Chung-Ang University
[3]German Research Center for Artificial Intelligence (DFKI)
{hayun0406, ghdrltjd, hwanheelee, eunwoo}@cau.ac.kr Sungho.Suh@dfki.de

Abstract

Prompt-based continual learning methods have emerged to address catastrophic forgetting by leveraging large-scale foundation models. These methods keep pre-trained models frozen and tune only small sets of parameters called prompts to learn tasks sequentially. However, when a new task comes in, the key-query matching mechanism in prompt-based methods selects the most relevant prompt without adequately considering whether it is actually suitable for learning the task. To address this, we propose the CoEn (Continual Enhanced prompt pool), which dynamically manages the prompt pool each time a new task is introduced. Our goal is to transform the static management of the prompt pool into a dynamic approach, enabling greater flexibility in adapting to new tasks and reducing the risk of catastrophic forgetting. Specifically, CoEn includes a new self-enhancement mechanism that assesses whether the prompts in the prompt pool can positively transfer knowledge to a new task and selectively strengthens the prompts. We demonstrate the proposed method under image classification benchmarks for class-incremental learning. Experimental results show that the proposed method outperforms existing prompt-based methods with an average margin of 3.8% across all scenarios.

1 Introduction

Humans can naturally incorporate new information into their existing learned knowledge. Continual learning (CL) aims to emulate this cognitive function within machine learning. Although deep learning algorithms perform well on individual tasks, learning multiple tasks in sequence is challenging due to catastrophic forgetting [1, 2], where prior knowledge is lost when learning new tasks.

To address this issue, various CL approaches have been proposed [3, 4, 5], and recent prompt-based learning methods utilizing large-scale foundation models have shown promise in reducing catastrophic forgetting [6, 7, 8]. They are adopting a transfer learning approach in natural language processing called prompt-tuning [9]. This approach freezes a pre-trained model, such as a vision transformer, and trains small sets of parameters called prompts. These prompts serve as task-specific instructions for the model, avoiding direct weight adjustments [10]. L2P [6] formalized a prompt-based continual learning framework. It defines a prompt pool as a collection of task-specific prompts and introduces a key-query matching mechanism to select the most relevant prompts for learning new tasks. DualPrompt [7] additionally introduces a set of task-invariant prompts shared across all tasks,

[*]Indicates equal contribution.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

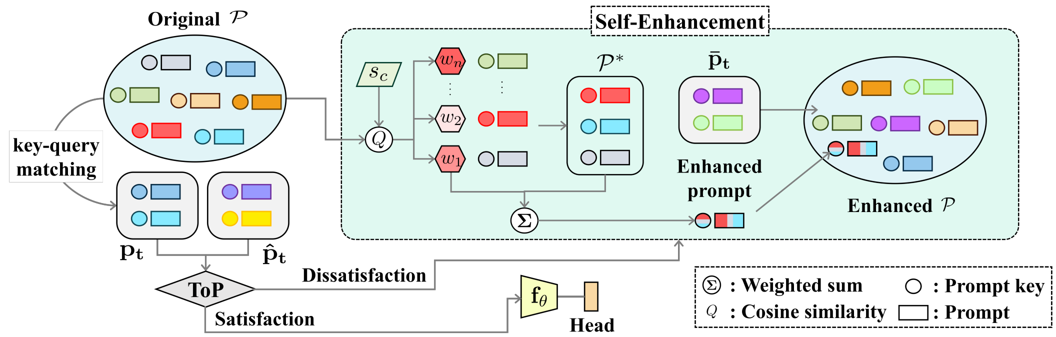

Figure 1: An overview of the proposed CoEn method for dynamic prompt pool management. CoEn evaluates the positive knowledge transferability of the current prompt pool $\mathcal{P}$ by comparing the key-query matched prompt p_t^* with a reference prompt $\hat{p}_t^*$. It computes the cosine similarity w_i between each prompt key and previously learned classes s_c and selects h prompts with the lowest values. Then, these prompts are aggregated using a weighted sum. The result is an enhanced prompt pool that better accommodates prior knowledge while absorbing new knowledge.

complementing L2P. Prompt gradient projection (PGP) [8] introduces an orthogonal constraint to the gradients of prompts, effectively reducing task interference.

Despite the advancements in prompt-based CL approaches [6, 7, 8], most methods simply use prompts selected from key-query matching for learning tasks. While intuitive, it overlooks the potential for positive knowledge transfer from those prompts to newer tasks. For example, after learning animal-related tasks up to the current time, a new vehicle-related task may arrive the next time. Even if the most relevant prompts are selected, differences in class attributes can limit positive knowledge transfer. This issue with managing a static prompt pool can impede both the retention of previous knowledge and the acquisition of new tasks.

In this paper, we propose CoEn(**Co**ntinual **En**hanced prompt pool), which dynamically enhances the prompt pool. Inspired by task similarity detection [11], CoEn detects whether the selected prompts from the prompt pool are suitable for a new task. Specifically, we employ a statistical risk [11] to make a diagnose of whether the selected prompts can enable positive knowledge transfer to the new task. If the statistical risk is not satisfied, we determine that the current prompt pool lacks the capacity to represent the new task. To do so, we introduce a self-enhancement mechanism that incorporates prompt addition and aggregation. By adding new prompts, we expand the capacity of the pool, allowing it to accommodate and integrate knowledge from newer tasks. To optimize the growing pool size, we aggregate prompts with the least impact on previously seen classes. This dynamic management balances stability and plasticity, overcoming the limitations of existing approaches. We evaluate the proposed CoEn in a class-incremental learning scenario and achieve a 3.8% average improvement over previous methods on benchmark datasets [12, 13, 14, 15, 16].

2 Method

2.1 Framework

In CL, a sequence of tasks $D = \{D_1, D_2, \dots, D_T\}$, where T is the total number of tasks, is given. Each task D_t consists of N_t input-label pairs $\{(x_i, y_i)\}_{i=1}^{N_t}$, with (x_i, y_i) sampled from the input space X_t and the label space Y_t. The goal is to train a model $f_\theta : X \to Y$ parameterized by θ to predict the label $y = f_\theta(x) \in Y$ for an unseen test sample x from a task in the sequence.

We reformulate the existing approach with a static prompt pool into a dynamic one as illustrated in Figure 1. We define the prompt pool $\mathcal{P} = \{p_1, p_2, \dots, p_n\}$ as a set of n learnable prompts, where each prompt $p_j \in \mathbb{R}^{L_p \times d}$ has a length L_p and an embedding dimension d. Whenever a new task is introduced, we select a set of most relevant prompts from $\mathcal{P}$ using key-query matching [6]. However, they do not consider whether p_t results in positive transfer during the learning the task t.

Inspired by [11], we aim to verify whether p_t can facilitate positive knowledge transfer during the learning process. [11] employes a binary vector, named task similarity vector (TSV), to assess the presence of positive knowledge transfer between tasks, defining the statistical risk as follows:

$$\mathop{\mathbb{E}}_{(x^t,y^t)}\left[\mathcal{L}(f_{s\to t}(x^t),y^t)\right] > \mathop{\mathbb{E}}_{(x^t,y^t)}\left[\mathcal{L}(f_r(x^t),y^t)\right]. \tag{1}$$

$f_{s\to t}$ denotes the transfer model used to transfer knowledge from task s to t, while f_r denotes the reference model used to learn task t independently. If the statistical risk holds, yielding $\text{TSV}^t(s) = 1$, task s is regarded as similar to task t; otherwise, $\text{TSV}^t(s) = 0$.

Since our approach diagnoses whether the prompt pool facilitates positive knowledge transfer to the current task, we reformulate TSV as the transferability of the prompt pool (ToP) as follows:

$$\text{ToP}^t(p_t) = \begin{cases} 1, & \text{if } \mathop{\mathbb{E}}_{(x^t,y^t)}\left[\mathcal{L}(f_\theta(x^t,p_t),y^t)\right] > \mathop{\mathbb{E}}_{(x^t,y^t)}\left[\mathcal{L}(f_\theta(x^t,\hat{p}_t),y^t)\right], \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where f_θ is a pre-trained ViT, and $\hat{p}_t$ denotes a set of reference prompts independent of $\mathcal{P}$. If $\text{ToP}^t(p_t) = 1$ indicates that p_t positively aids in learning the current task. Conversely, $\text{ToP}^t(p_t) = 0$ suggests $\mathcal{P}$ lacks sufficient representational capacity to accommodate the new knowledge.

2.2 Self-Enhancement Mechanism

A prompt pool with limited representation capacity causes forgetting of previous tasks and impedes acquiring new knowledge. If the statistical risk is not satisfied, it suggests that the current capacity is insufficient to encode any additional knowledge into the pool. Therefore, we expand the capacity of the pool by adding a newly initialized prompt $\bar{p}_t$ to the pool. To prevent the prompt pool from expanding excessively, we introduce prompt aggregation, which combines prompts with low similarity to previously learned classes into a single enhanced prompt. After each incremental learning step, we extract the class-wise feature vector $v_c = f_\theta(x_c)$ for each seen class $c \in C$, where C is the total number of classes, to assess the similarity between prompt keys and previously learned classes. Instead of storing v_c, we store the mean $\mu_c \in \mathbb{R}^d$ and covariance $\Sigma_c \in \mathbb{R}^{d\times d}$ for each class c [17]. From the distributions with μ and Σ, we sample the feature s_c for each seen class. The similarity w_i between s_c and the i-th prompt keys k_i is calculated using the cosine similarity, $sim(\cdot)$ as

$$w_i = \frac{1}{C}\sum_{c=1}^{C} sim(k_i, s_c), i \in \{0,\ldots,N\}, \tag{3}$$

where N is the total number of prompts in the pool. We select the set of h prompts $\mathcal{P}^* = \{(k_{\rho(1)}, p_{\rho(1)}), (k_{\rho(2)}, p_{\rho(2)}), \ldots, (k_{\rho(h)}, p_{\rho(h)})\} \subseteq \mathcal{P}$, where $\rho(i)$ gives the index corresponding to the i-th lowest value in w. The prompt keys and prompts in $\mathcal{P}^*$ are aggregated into a set of enhanced key and prompt through a weighted sum based on the w_i as follows:

$$k_{enh} = \sum_{i=1}^{h} \frac{w_{\rho(i)} k_{\rho(i)}}{\sum_{i=1}^{h} w_{\rho(i)}}, \quad p_{enh} = \sum_{i=1}^{h} \frac{w_{\rho(i)} p_{\rho(i)}}{\sum_{i=1}^{h} w_{\rho(i)}}. \tag{4}$$

The enhanced key k_{enh} and corresponding prompt p_{enh} are integrated into $\mathcal{P}$. Note that CoEn applies the matching loss and the cross-entropy loss for classification [7].

3 Experiments

3.1 Setup

We conducted on class-incremental learning using various benchmark datasets: CIFAR-100 [12], ImageNet-R [13], EuroSAT [14], RESISC45 [15], and CUB-200 [16]. We define tasks by partitioning the entire set of classes into disjoint subsets in a dataset. Note that "x-Split" refers to dividing a dataset into x tasks, each comprising a subset of the classes of the dataset. We used a ViT backbone [18] pre-trained on ImageNet and compared with L2P [6], DualPrompt [7], and PGP [8].

3.2 Results

3.2.1 General-domain tasks

We used CIFAR-100 [12] and ImageNet-R [13], which consist of general object images, divided into 10 and 20 tasks, respectively, as shown in Table 1. For CIFAR-100, CoEn outperformed L2P,

Table 1: Comparison results on 10-Split-CIFAR100, 20-Split-CIFAR100, 10-Split-ImageNet-R, and 20-Split-ImageNet-R. Accuracy denotes the average accuracy for each task after incremental learning. Forgetting refers to the average decrease in performance on previous tasks after learning a new task.

Method	10-Split-CIFAR100		20-Split-CIFAR100		10-Split-ImageNet-R		20-Split-ImageNet-R	
	Accuracy	Forgetting	Accuracy	Forgetting	Accuracy	Forgetting	Accuracy	Forgetting
L2P	83.5	6.9	81.6	9.4	65.1	5.1	57.0	9.5
DualPrompt	86.1	5.8	83.5	7.8	69.2	**4.7**	65.7	**7.1**
PGP	86.7	5.5	83.5	8.1	69.1	5.8	**65.9**	**7.1**
CoEn (ours)	**86.8**	**4.9**	**84.3**	**6.4**	**69.6**	5.6	64.9	8.0

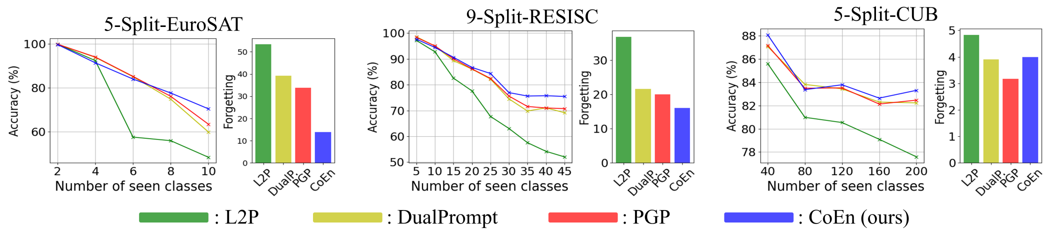

Figure 2: Comparison results on 5-Split-EuroSAT, 9-Split-RESISC45, and 5-Split-CUB. The left plot depicts accuracy, while the right bar graph indicates the degree of forgetting.

DualPrompt, and PGP in the 10-split scenario with gains of 3.3%, 0.7%, and 0.1%, respectively, and in the 20-split scenario, it achieved gains of 2.7%, 0.8%, and 0.8%. For forgetting, we reduced it by an average of 1.2% in the 10-split task and 2.0% in the 20-split task compared to other methods. These improvements highlight the proposed method, which balances stability and plasticity in CL. For ImageNet-R, CoEn performed comparably to L2P, DualPrompt, and PGP in the 10-split task, outperforming them by 4.5%, 0.4%, and 0.5%, respectively. In the 20-split, it remained close to PGP but outperformed L2P by 7.9%. CoEn maintained the forgetting rate of 5.6%, similar to DualPrompt, demonstrating its effectiveness in retaining knowledge while learning new tasks.

3.2.2 Specific-domain tasks

In the specific-domain setup, we used EuroSAT [14], RESISC45 [15], and CUB-200 [16], which contain specific categories, such as satellite imagery and fine-grained bird species, divided into 5, 9, and 5 tasks, respectively, as shown in Figure 2. For EuroSAT, CoEn achieved the highest accuracy of 70.5%, with gains of 22.0%, 8.9%, and 3.4% over L2P, DualPrompt, and PGP, respectively. Moreover, it significantly reduced forgetting to 13.8%, compared to higher rates of 29.5%. For RESISC45, CoEn reached an accuracy of 75.5%, outperforming the others by margins of 6.1%, 6.3%, and 4.7%, respectively. Additionally, it exhibited a forgetting rate of 16.0%, which is, on average, 5.3% lower than that of the other methods. For CUB-200, CoEn achieved 83.3% accuracy with a low 4.0% forgetting rate, highlighting its strength in handling fine-grained features.

4 Conclusion

In this work, we have proposed CoEn that dynamically manages the prompt pool to address the limitations of static management of existing works. The dynamic pool management strategy continuously expands and refines the prompt pool to enable positive knowledge transfer to all tasks, facilitating new task learning and mitigating catastrophic forgetting. Experimental results show that CoEn outperforms existing methods, achieving an average accuracy gain of 3.8% across all scenarios.

Acknowledgements

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (2021-0-01341, Artificial Intelligence Graduate School Program(Chung-Ang University) and No.2022-0-00124 (RS-2022-II220124), Development of Artificial Intelligence Technology for Self-Improving Competency-Aware Learning Capabilities).

References

[1] Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, pages 109–165. Elsevier, 1989.

[2] Robert M French. Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4):128–135, 1999.

[3] 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.

[4] Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. *arXiv preprint arXiv:1606.04671*, 2016.

[5] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017.

[6] Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 139–149, 2022.

[7] Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun, Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, et al. Dualprompt: Complementary prompting for rehearsal-free continual learning. In *European Conference on Computer Vision*, pages 631–648. Springer, 2022.

[8] Jingyang Qiao, Xin Tan, Chengwei Chen, Yanyun Qu, Yong Peng, Yuan Xie, et al. Prompt gradient projection for continual learning. In *The Twelfth International Conference on Learning Representations*, 2023.

[9] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35, 2023.

[10] Renze Lou, Kai Zhang, and Wenpeng Yin. A comprehensive survey on instruction following. *arXiv preprint arXiv:2303.10475*, 2023.

[11] Zixuan Ke, Bing Liu, and Xingchang Huang. Continual learning of a mixed sequence of similar and dissimilar tasks. *Advances in neural information processing systems*, 33:18493–18504, 2020.

[12] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[13] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[14] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019.

[15] Gong Cheng, Junwei Han, and Xiaoqiang Lu. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 105(10):1865–1883, 2017.

[16] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

[17] Gengwei Zhang, Liyuan Wang, Guoliang Kang, Ling Chen, and Yunchao Wei. Slca: Slow learner with classifier alignment for continual learning on a pre-trained model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 19148–19158, October 2023.

[18] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.