2nd INTERNATIONAL CONFERENCE ON
Artificial Intelligence
31 May - 1 June 2026

IICAI
Second International Interdisciplinary Conference on Artificial Intelligence

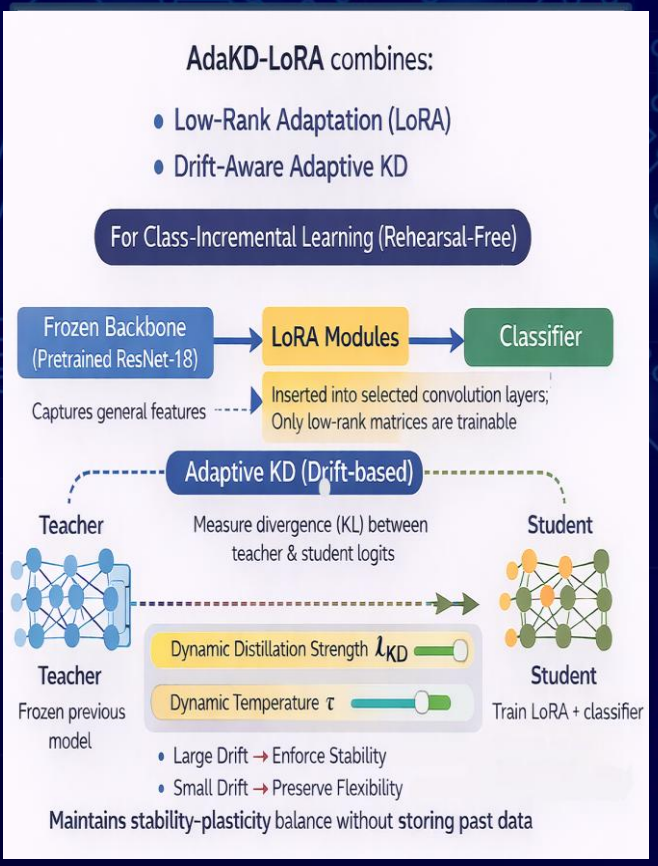
AdaKD-LoRA: Adaptive Knowledge Distillation with Low-Rank Adaptation for Rehearsal Free Continual Learning

Mohammad Ghabel Rahmat, Samira Vejdanparast, Azam Bastanfard, Abbas Jalilvand*
Corresponding Author: PhD Candidate, Computer Engineering, Email: M.ghabelrahmat@iau.ac.ir
PhD Candidate Computer Engineering m.ghabelrahmat@iau.ac.ir, PhD Candidate, Computer Engineering s.vejdanparast@iau.ac.ir,
Assistant Professor, Department of Computer Engineering Bastanfard@kiau.ac.ir, Assistant Professor, Department of Computer Engineering Jalilvand@iau.ac.ir

Abstract

Rehearsal free continual learning requires models to acquire new task-specific knowledge without erasing previously learned information, despite the absence of stored exemplars. However, many distillation-based strategies employ a static knowledge-distillation weight, applying a fixed regularization strength across all tasks. This static formulation is unable to reflect the varying degrees of model drift that occur during incremental learning, often resulting in either excessive rigidity or insufficient retention. To address this limitation, we introduce AdaKD-LoRA, a parameter efficient framework that combines low-rank adaptation with dynamically regulated distillation. In this approach, a pretrained backbone remains frozen to preserve general representations, while lightweight LoRA modules encode task-specific updates. The strength of the distillation term is adaptively adjusted based on the divergence between consecutive task models, allowing the method to increase emphasis on stability when drift is high and to maintain flexibility when new classes require stronger plasticity. Experiments on a ten-task Class-Incremental Learning benchmark show that this adaptive formulation substantially reduces forgetting and achieves more stable knowledge preservation compared with Fine-Tuning, LwF, and LoRA-only baselines. AdaKD-LoRA delivers higher accuracy on earlier tasks, more favorable backward transfer, and competitive overall performance, demonstrating that replacing static distillation with an adaptive mechanism provides a more effective and scalable solution for rehearsal-free continual learning.

Research method



conclusion

we proposed **AdaKD-LoRA**, a parameter-efficient and rehearsal-free framework for Class-Incremental Learning that combines low-rank adaptation with adaptive knowledge distillation. By freezing the pretrained backbone and updating only LoRA modules, the method reduces structural interference across tasks.
Unlike static KD approaches, the distillation strength and temperature are dynamically adjusted based on model divergence, enabling a better stability–plasticity balance.
Results on Split CIFAR-100 show higher average accuracy, lower forgetting, and improved backward transfer compared to strong rehearsal-free baselines, confirming the effectiveness of the drift-aware adaptive distillation strategy.

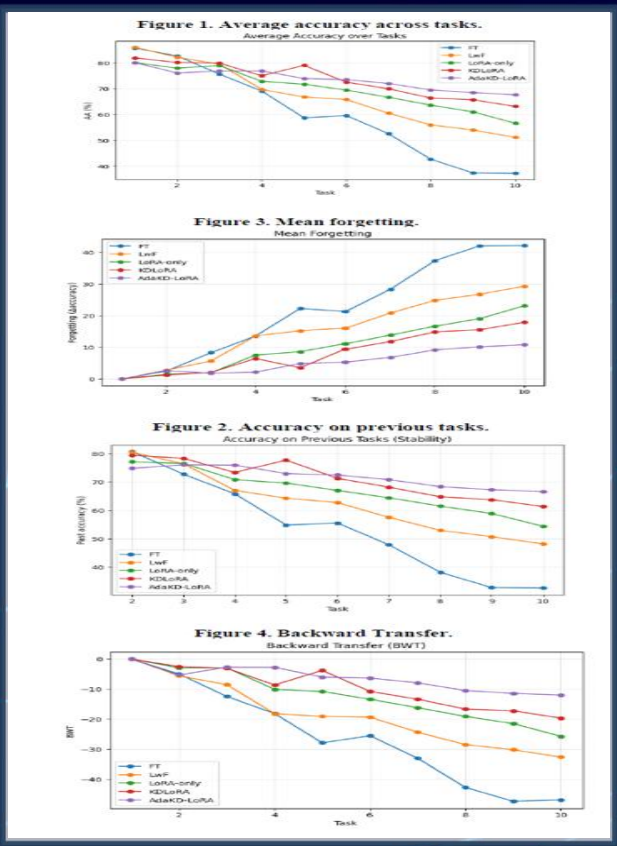
The purpose of the research

AdaKD-LoRA is a rehearsal-free Class-Incremental Learning framework that integrates Low-Rank Adaptation (LoRA) with drift-aware adaptive knowledge distillation.
The pretrained backbone is frozen to preserve previously learned representations, while lightweight LoRA modules enable parameter-efficient task adaptation. Only low-rank parameters and the classifier head are updated.
To prevent catastrophic forgetting, the divergence between teacher and student predictions is measured at each task. The distillation strength is dynamically adjusted:
High drift → stronger distillation (stability)
Low drift → weaker distillation (plasticity)
This adaptive mechanism maintains an effective stability–plasticity balance without storing past data.

Key Contributions

- First drift-aware adaptive KD in rehearsal-free Class-IL
- Dynamic task-dependent stability regulation
- LoRA-based structural constraint for reduced interference
- Significant forgetting reduction over static KD baselines

Practical or simulation results



References

[1] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," Neural Networks, vol. 113, pp. 54–71, 2019.
[2] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars,"A continual learning survey: Defying forgetting in classification tasks," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022

[3] G. M. van de Ven, T. Tuytelaars, and A. S. Tolias,"Three types of incremental learning," Nature Machine Intelligence, vol. 4, no. 12, pp. 1185–1197, 2022.
[4] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. van de Weijer,"Class-incremental learning: Survey and performance evaluation on image classification,"IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
[5] M. McCloskey and N. J. Cohen,"Catastrophic interference in connectionist networks: The sequential learning problem,"Psychology of Learning and Motivation, vol. 24, pp. 109–165, 1989.
[6] S.-A. Rebuffi, A. Kolesnikov, and C. H. Lampert,"iCaRL: Incremental classifier and representation learning," in Proc. CVPR, 2017, pp. 2001–2010.
[7] J. Kirkpatrick et al.,"Overcoming catastrophic forgetting in neural networks,"Proceedings of the National Academy of Sciences, vol. 114, no. 13, pp. 3521–3526, 2017.
[8] F. Zenke, B. Poole, and S. Ganguli,"Continual learning through synaptic intelligence," in Proc. ICML, 2017, pp. 3987–3995.
[9] E. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen,"LoRA: Low-rank adaptation of large language models," in Proc. ICLR, 2022
[10] Z. Li and D. Hoiem,"Learning without forgetting," in Proc. ECCV, 2016, pp. 614–629.G.
[11] Hinton, O. Vinyals, and J. Dean,"Distilling the knowledge in a neural network," arXiv preprint, arXiv:1503.02531, 2015.
[12] D. Lopez-Paz and M. Ranzato, "Gradient episodic memory for continual learning," Proc. NeurIPS, 2017.
[13] A. Chaudhry et al., "Efficient lifelong learning with A-GEM," Proc. ICLR, 2019.
[14] R. Aljundi et al., "Gradient-based sample selection for online continual learning," Proc. NeurIPS, 2019.
[15] H. Shin et al., "Continual learning with deep generative replay," Proc. NeurIPS, 2017.
[16] N. Houlsby et al., "Parameter-efficient transfer learning for NLP," Proc. ICML, 2019.
[17] E. Hu et al., "LoRA: Low-rank adaptation of large language models," Proc. ICLR, 2022.
[18] O. Zaken et al., "BitFit: Simple and effective fine-tuning for transformer-based masked language models," arXiv:2106.10199, 2021.
[19] B. Liu et al., "Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning," arXiv:2205.05638, 2022.
[20] M. Jia et al., "Visual prompt tuning," Proc. ECCV, 2022.
[21] G. Hinton et al., "Distilling the knowledge in a neural network," arXiv:1503.02531, 2015.
[22] A. Romero et al., "FitNets: Hints for thin deep nets," arXiv:1412.6550, 2014.
[23] T. Furlanello et al., "Born-again neural networks," Proc. ICML, 2018.
[24] Y. Li et al., "Rethinking the transferability of pretrained models," Proc. CVPR, 2022.
[25] X. Li, S. Grandvalet, and F. Davoine, "Explicit inductive bias for transfer learning," Proc. ICLR, 2018.
[26] R. Azimi, R. Rishav, M. Teichmann, and S. E. Kahou, "KD-LoRA: A hybrid approach to efficient fine-tuning with LoRA and knowledge distillation," in Proc. 4th NeurIPS Workshop Efficient Natural Language and Speech Processing (ENLSP-IV), 2024, Art. no. 2410.20777.
[27] A. Krizhevsky,"Learning multiple layers of features from tiny images," Tech. Rep., University of Toronto, 2009.
[28] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in Proc. Int. Conf. Learn. Represent. (ICLR), 2022
[29] He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–77