Foundation Models Meet Continual Learning: Recent Advances, Challenges, and Future Directions

Tarun Raheja*
Kipo AI
San Francisco, USA
tarun@kipo.ai

Nilay Pochhi*
Independent Researcher
San Francisco, USA
pochhi.nilay@gmail.com

Abstract

Foundation models (FMs) have emerged as powerful pre-trained systems capable of adapting to diverse downstream tasks, while continual learning (CL) aims to enable models to sequentially acquire new knowledge without catastrophically forgetting previous information. This paper examines the synergies between recent advances in FMs and CL techniques. We review key FM capabilities relevant to CL, analyze how FM architectures and training paradigms can enhance CL methods, and explore integrated approaches combining FM and CL principles. Our analysis suggests that FMs' robust representations, transfer abilities, and adaptable architectures offer promising avenues for advancing CL, while CL techniques can enable FMs to continually expand their capabilities in dynamic environments.

1 Introduction

The rapid progress in foundation models (FMs) Bommasani et al. [2021] and continual learning (CL) Parisi and Kanan [2019] has opened new possibilities for developing AI systems that can flexibly acquire and retain knowledge over time. FMs, exemplified by models like BERT Devlin et al. [2019], GPT-3 Brown et al. [2020], and CLIP Radford et al. [2021], demonstrate remarkable generalization and few-shot learning abilities across diverse tasks. Concurrently, CL approaches aim to overcome catastrophic forgetting McCloskey and Cohen [1989] when learning sequential tasks. This survey paper delves into how recent FM advances can be leveraged to enhance CL techniques, examining their complementary strengths and potential synergies.

2 Foundation Models for Continual Learning

2.1 Parameter-Efficient Fine-Tuning Techniques

The immense scale of FMs, while beneficial for performance, presents challenges for continual adaptation. Updating all parameters for each new task is computationally expensive and prone to overfitting. Parameter-efficient fine-tuning (PEFT) techniques have emerged to address this by selectively updating only a small subset of parameters while keeping the majority frozen, thus mitigating catastrophic forgetting by preserving pre-trained knowledge Houlsby et al. [2019]. Adapters are a prominent example of PEFT, introducing small bottleneck layers within the model to capture task-specific information. AdapterHub Pfeiffer et al. [2020] provides a framework for adapting Transformers by dynamically "stitching-in" pre-trained adapters. LoRA (Low-Rank Adaptation) Hu et al. [2021] offers another efficient approach by leveraging low-rank updates, demonstrating its effectiveness in various continual learning scenarios Wistuba et al. [2023]. Prefix Tuning modifies the

*Equal contribution

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

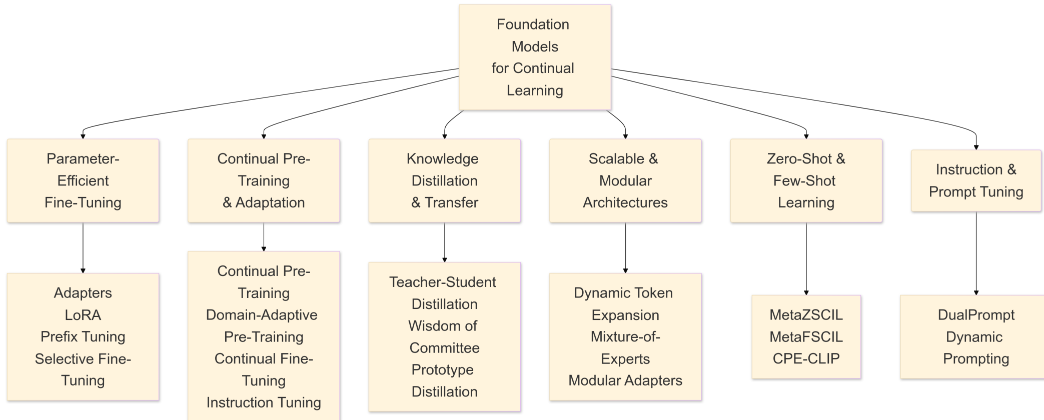

Figure 1: Overview of Foundation Models for Continual Learning

input embeddings to guide the model's behavior on different tasks. Selective parameter fine-tuning methods aim to identify crucial parameters for updating based on task relevance or gradient information, mitigating negative transfer and promoting knowledge retention Wang et al. [2024b], Zhang et al. [2023], Li et al. [2024]. These PEFT techniques offer promising avenues for adapting FMs to new tasks while striking a balance between retaining pre-trained knowledge and acquiring new information.

2.2 Continual Pre-Training and Adaptation

Continually updating FMs with new data can further enhance their adaptability in CL. Continual pre-training involves incrementally updating FM knowledge with new data, enabling them to retain general capabilities while incorporating novel information Gupta et al. [2023], Ke et al. [2022], Li and yi Lee [2024], Çağatay Yıldız et al. [2024]. Domain-Adaptive Pre-training focuses on adapting FMs to specific domains by pre-training on domain-specific data, further enhancing their performance on downstream tasks within that domain Shi et al. [2024], Jin et al. [2022]. Continual fine-tuning leverages continual learning techniques during the fine-tuning stage, mitigating catastrophic forgetting while adapting to new tasks. COMFORT Li and Jha [2024] proposes a continual fine-tuning framework for foundation models targeted at consumer healthcare, leveraging parameter-efficient fine-tuning methods such as LoRA. Instruction tuning empowers models to follow natural language instructions, enabling them to perform new tasks by understanding instructions rather than relying solely on labeled examples Scialom et al. [2022], Wang et al. [2023c], Wu et al. [2024]. Latent replay offers a memory-efficient approach to mitigating forgetting by storing and replaying compact representations of past experiences in the latent space, proving particularly beneficial in scenarios where data privacy is a concern Ostapenko et al. [2022]. These techniques allow FMs to evolve and adapt to changing data distributions, bridging the gap between pre-training and continual learning.

2.3 Knowledge Distillation and Transfer

Knowledge distillation leverages the knowledge acquired by a teacher model to guide the training of a student model. In CL, knowledge distillation plays a crucial role in transferring knowledge from a model trained on previous tasks to a model trained on a new task, mitigating catastrophic forgetting Yu et al. [2024b], Cai et al. [2023], Chen et al. [2023]. Wisdom of Committee Liu et al. [2024b] proposes a novel distillation approach where a teaching committee comprising both foundation model teachers and complementary teachers guides the student model, achieving enhanced performance in knowledge transfer. Prototype-based distillation extends knowledge distillation by distilling knowledge from previously learned class prototypes to guide the learning of new classes, enabling efficient knowledge transfer in class-incremental learning scenarios Asadi et al. [2023], Wang et al. [2023a], Li et al. [2022]. This approach, combined with memory replay techniques like Move-to-Data Poursanidis et al. [2020], further improves knowledge retention and transfer. By transferring knowledge between models, these techniques enhance knowledge retention and facilitate efficient learning of new tasks without requiring access to past training data.

2.4 Scalable and Modular Architectures

Scalable and modular architectures are essential for enabling continual learning with FMs. Dynamic token expansion in transformer architectures allows for task-specific adaptation by dynamically expanding special tokens, enabling models to learn new tasks without significantly increasing the parameter count Douillard et al. [2021]. Mixture-of-Experts (MoE) architectures, where different experts specialize in different aspects of the input data, can be leveraged to enhance continual learning by selectively activating and updating experts for different tasks Yu et al. [2024a], Luo et al. [2024]. Modular adapters, which are small task-specific modules inserted into the pre-trained model, offer a flexible and parameter-efficient approach to continual learning by dynamically adding and composing modules for new tasks Roy et al. [2024], Wang et al. [2024c,a], Khan et al. [2022], Wang et al. [2024d]. Interference-free knowledge integration focuses on minimizing interference between tasks by ensuring that the knowledge acquired for new tasks does not disrupt the knowledge acquired for previous tasks, promoting robust knowledge retention and transfer Tang et al. [2024]. These architectures offer promising avenues for building CL systems that can efficiently adapt to new tasks while retaining previously acquired knowledge.

3 Continual Learning Paradigms with Foundation Models

3.1 Zero-Shot and Few-Shot Learning in Continual Contexts

Zero-shot and few-shot learning play crucial roles in enabling continual learning in scenarios where labeled data is limited. Zero-shot learning leverages semantic information about classes, such as class attributes or textual descriptions, to recognize unseen classes without any labeled examples Yu et al. [2024b], Zheng et al. [2023b], Wei et al. [2020, 2021], Yi and Elhoseiny [2021]. MetaZSCIL Wu et al. [2023] introduces a meta-learning approach for generalized zero-shot class incremental learning, enabling models to incrementally learn unseen classes without training samples. Few-shot learning aims to learn new classes from a very limited number of labeled examples, often by leveraging meta-learning techniques or by adapting pre-trained models using parameter-efficient fine-tuning methods Roy et al. [2024], Hu et al. [2023], Hung et al. [2019], Zhou et al. [2022], Ke et al. [2022]. Multimodal Parameter-Efficient Few-Shot Class Incremental Learning D'Alessandro et al. [2023] proposes CPE-CLIP, a method that leverages the knowledge acquired by CLIP to improve performance and prevent forgetting in FSCIL. MetaFSCIL Chi et al. [2022] employs a meta-learning approach for few-shot class incremental learning, demonstrating significant performance improvements on benchmark datasets. By effectively leveraging limited labeled data, these techniques facilitate continual learning in realistic scenarios where new classes emerge over time and labeled data may be scarce.

3.2 Instruction and Prompt Tuning

Instruction and prompt tuning provide powerful mechanisms for adapting FMs to new tasks in continual learning, enabling models to learn new skills without extensive parameter updates. Prompt-based continual learning employs prompts, which are small learnable parameters, to guide the model's behavior on different tasks Ahrens et al. [2023], Hu et al. [2023], Gao et al. [2024], Khan et al. [2023], Wang et al. [2021]. These prompts can be dynamically selected or updated based on the task, allowing for flexible adaptation while retaining pre-trained knowledge. DualPrompt Wang et al. [2022] presents a novel approach to attach complementary prompts to a pre-trained model, achieving state-of-the-art performance in continual learning without rehearsal. Dynamic prompting extends this approach by allowing prompts to evolve over time, further enhancing the model's adaptability to new tasks Gao et al. [2024]. By effectively leveraging the pre-trained knowledge embedded in FMs and by introducing a succinct memory system through prompts, these techniques enable efficient continual adaptation while mitigating catastrophic forgetting.

4 Theoretical and Empirical Foundations

Understanding the empirical and theoretical foundations of continual learning with FMs is vital for developing robust and effective CL systems. Empirical evaluations on various benchmarks provide valuable insights into the effectiveness of different CL methods, aiding in identifying the strengths and weaknesses of various approaches Zheng et al. [2023a], Li and Jha [2024], Ermiş et al.

[2022], Chitale et al. [2023], Wu et al. [2022]. Latent replay has demonstrated its effectiveness in mitigating forgetting by preserving a compact representation of past experiences Ostapenko et al. [2022]. Hierarchical task decomposition, where complex tasks are broken down into simpler subtasks, is a promising direction for enhancing continual learning by promoting knowledge transfer and minimizing interference between tasks Wang et al. [2023b]. Mode connectivity, referring to the existence of low-loss valleys connecting different local minima in the loss landscape, provides insights into the relationship between different learned tasks and can be leveraged to improve knowledge transfer and reduce forgetting Ren et al. [2024]. The Neural Tangent Kernel (NTK) offers a promising theoretical tool for analyzing the training dynamics of neural networks, providing valuable insights into the interplay between feature representation, task orthogonality, and generalization in continual learning scenarios Liu et al. [2024a]. These theoretical and empirical foundations pave the way for designing more effective and robust CL systems.

5 Open Challenges and Future Directions

Despite the promising synergies between FMs and CL, several challenges remain and demand further investigation. Scalability and efficiency are crucial for handling a vast number of tasks and large datasets, requiring innovative solutions to manage computational resources and training time Mehta et al. [2021], Çağatay Yıldız et al. [2024], Ermiş et al. [2022]. Mitigating negative transfer, where learning a new task hinders performance on previous ones, necessitates careful consideration of task relationships and the development of strategies to minimize interference Ke et al. [2022], Adel [2024]. Comprehensive evaluation protocols are essential for rigorously comparing methods, taking into account factors like forgetting, transfer learning ability, sample efficiency, and computational cost Li and Jha [2024], Zheng et al. [2023a]. Ethical considerations, including data bias, fairness, and responsible use, are increasingly important as FMs and CL technologies continue to advance Shi et al. [2024]. Addressing these challenges will pave the way for the development and deployment of more robust, efficient, and ethical continual learning systems powered by foundation models.

6 Conclusion

This paper explored how recent advances in foundation models can enhance continual learning techniques. We highlighted the potential of FMs' robust representations, transfer capabilities, and adaptable architectures to address key CL challenges. Integrated approaches combining FM and CL paradigms offer promising directions for developing more flexible and capable AI systems. As research in both fields progresses, the convergence of FM and CL principles may lead to significant breakthroughs in building AI systems that can continually learn and adapt in dynamic environments.

References

Tameem Adel. Similarity-based adaptation for task-aware and task-free continual learning, 2024.

Kyra Ahrens, Hans Hergen Lehmann, Jae Hee Lee, and Stefan Wermter. Read between the layers: Leveraging intra-layer representations for rehearsal-free continual learning with pre-trained models, 2023.

Nader Asadi, Mohammad Davar, S. Mudur, Rahaf Aljundi, and Eugene Belilovsky. Prototype-sample relation distillation: Towards replay-free continual learning, 2023.

Rishi Bommasani, Drew A. Hudson, E. Adeli, R. Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, J. Bohg, Antoine Bosselut, E. Brunskill, Erik Brynjolfsson, S. Buch, Dallas Card, Rodrigo Castellon, Niladri S. Chatterji, Annie S. Chen, Kathleen A. Creel, Jared Davis, Dora Demszky, Chris Donahue, M. Doumbouya, Esin Durmus, Stefano Ermon, J. Etchemendy, Kawin Ethayarajh, L. Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah D. Goodman, S. Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas F. Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, G. Keeling, Fereshte Khani, O. Khattab, Pang Wei Koh, M. Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, J. Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir

Mirchandani, E. Mitchell, Zanele Munyikwa, Suraj Nair, A. Narayan, D. Narayanan, Benjamin Newman, Allen Nie, Juan Carlos Niebles, H. Nilforoshan, J. Nyarko, Giray Ogut, Laurel J. Orr, Isabel Papadimitriou, J. Park, C. Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Robert Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher R'e, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, K. Srinivasan, Alex Tamkin, Rohan Taori, A. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, M. Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models, 2021.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, J. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, T. Henighan, R. Child, A. Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Ma teusz Litwin, Scott Gray, B. Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, I. Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.

Yuliang Cai, Jesse Thomason, and Mohammad Rostami. Task-attentive transformer architecture for continual learning of vision-and-language tasks using knowledge distillation, 2023.

Kanghao Chen, Sijia Liu, Ruixuan Wang, and Weishi Zheng. Adaptively integrated knowledge distillation and prediction uncertainty for continual learning, 2023.

Zhixiang Chi, Li Gu, Huan Liu, Yang Wang, Yuanhao Yu, and Jingshan Tang. Metafscil: A meta-learning approach for few-shot class incremental learning, 2022.

Rajas Chitale, Ankit Vaidya, Aditya Kane, and Archana Ghotkar. Task arithmetic with lora for continual learning, 2023.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

Arthur Douillard, Alexandre Ram'e, Guillaume Couairon, and M. Cord. Dytox: Transformers for continual learning with dynamic token expansion, 2021.

Marco D'Alessandro, Alberto Alonso, Enrique Calabrés, and M. Galar. Multimodal parameter-efficient few-shot class incremental learning, 2023.

B. Ermiş, Giovanni Zappella, Martin Wistuba, Aditya Rawal, and C. Archambeau. Memory efficient continual learning with transformers, 2022.

Xinyuan Gao, Songlin Dong, Yuhang He, Qiang Wang, and Yihong Gong. Beyond prompt learning: Continual adapter for efficient rehearsal-free continual learning, 2024.

Kshitij Gupta, Benjamin Th'erien, Adam Ibrahim, Mats L. Richter, Quentin G. Anthony, Eugene Belilovsky, I. Rish, and Timothée Lesort. Continual pre-training of large language models: How to (re)warm your model?, 2023.

N. Houlsby, A. Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and S. Gelly. Parameter-efficient transfer learning for nlp, 2019.

J. E. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021.

Zhiyuan Hu, J. Lyu, Dashan Gao, and N. Vasconcelos. Pop: Prompt of prompts for continual learning, 2023.

Steven C. Y. Hung, Cheng-Hao Tu, Cheng-En Wu, Chien-Hung Chen, Yi-Ming Chan, and Chu-Song Chen. Compacting, picking and growing for unforgetting continual learning, 2019.

Xisen Jin, Dejiao Zhang, Henghui Zhu, Wei Xiao, Shang-Wen Li, Xiaokai Wei, Andrew O. Arnold, and Xiang Ren. Lifelong pretraining: Continually adapting language models to emerging corpora, 2022.

Zixuan Ke, Haowei Lin, Yijia Shao, Hu Xu, Lei Shu, and Bin Liu. Continual training of language models for few-shot learning, 2022.

Muhammad Gul Zain Ali Khan, Muhammad Ferjad Naeem, L. Gool, D. Stricker, F. Tombari, and Muhammad Zeshan Afzal. Introducing language guidance in prompt-based continual learning, 2023.

Shadab Khan, Surbhi Agarwal, and Khan Agarwal. Lifelong language learning with adapter based transformers, 2022.

Chen-An Li and Hyung yi Lee. Examining forgetting in continual pre-training of aligned large language models, 2024.

Chia-Hao Li and N. Jha. Comfort: A continual fine-tuning framework for foundation models targeted at consumer healthcare, 2024.

Haoling Li, Xin Zhang, Xiao Liu, Yeyun Gong, Yifan Wang, Yujiu Yang, Qi Chen, and Peng Cheng. Gradient-mask tuning elevates the upper limits of llm performance, 2024.

Xiaodi Li, Zhuoyi Wang, Dingcheng Li, L. Khan, and B. Thuraisingham. Lpc: A logits and parameter calibration framework for continual learning, 2022.

Jingren Liu, Zhong Ji, Yunlong Yu, Jiale Cao, Yanwei Pang, Jungong Han, and Xuelong Li. Parameter-efficient fine-tuning for continual learning: A neural tangent kernel perspective, 2024a.

Zichang Liu, Qingyun Liu, Yuening Li, Liang Liu, Anshumali Shrivastava, Shuchao Bi, Lichan Hong, Ed H. Chi, and Zhe Zhao. Wisdom of committee: Distilling from foundation model to specialized application model, 2024b.

Tongxu Luo, Jiahe Lei, Fangyu Lei, Weihao Liu, Shizhu He, Jun Zhao, and Kang Liu. Moelora: Contrastive learning guided mixture of experts on parameter-efficient fine-tuning for large language models, 2024.

M. McCloskey and N. J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem, 1989.

Sanket Vaibhav Mehta, Darshan Patil, Sarath Chandar, and Emma Strubell. An empirical investigation of the role of pre-training in lifelong learning, 2021.

O. Ostapenko, Timothée Lesort, P. Rodriguez, Md Rifat Arefin, Arthur Douillard, I. Rish, and Laurent Charlin. Continual learning with foundation models: An empirical study of latent replay, 2022.

G. I. Parisi and Christopher Kanan. Rethinking continual learning for autonomous agents and robots, 2019.

Jonas Pfeiffer, Andreas Rücklé, Clifton A. Poth, Aishwarya Kamath, Ivan Vulic, Sebastian Ruder, Kyunghyun Cho, and Iryna Gurevych. Adapterhub: A framework for adapting transformers, 2020.

Miltiadis Poursanidis, J. Benois-Pineau, A. Zemmari, Boris Mansencal, and A. Rugy. Move-to-data: A new continual learning approach with deep cnns, application for image-class recognition, 2020.

Alec Radford, Jong Wook Kim, Chris Hallacy, A. Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision, 2021.

Weijieying Ren, Xinlong Li, Lei Wang, Tianxiang Zhao, and Wei Qin. Analyzing and reducing catastrophic forgetting in parameter efficient tuning, 2024.

Shuvendu Roy, Elham Dolatabadi, Arash Afkanpour, and A. Etemad. Few-shot tuning of foundation models for class-incremental learning, 2024.

Thomas Scialom, Tuhin Chakrabarty, and S. Muresan. Continual-t0: Progressively instructing 50+ tasks to language models without forgetting, 2022.

Haizhou Shi, Zihao Xu, Hengyi Wang, Weiyi Qin, Wenyuan Wang, Yibin Wang, and Hao Wang. Continual learning of large language models: A comprehensive survey, 2024.

Longxiang Tang, Zhuotao Tian, Kai Li, Chunming He, Hantao Zhou, Hengshuang Zhao, Xiu Li, and Jiaya Jia. Mind the interference: Retaining pre-trained knowledge in parameter efficient continual learning of vision-language models, 2024.

Huiyi Wang, Haodong Lu, Lina Yao, and Dong Gong. Self-expansion of pre-trained models with mixture of adapters for continual learning, 2024a.

Jue Wang, Dajie Dong, Lidan Shou, Ke Chen, and Gang Chen. Effective continual learning for text classification with lightweight snapshots, 2023a.

Liyuan Wang, Jingyi Xie, Xingxing Zhang, Mingyi Huang, Hang Su, and Jun Zhu. Hierarchical decomposition of prompt-based continual learning: Rethinking obscured sub-optimality, 2023b.

Liyuan Wang, Jingyi Xie, Xingxing Zhang, Hang Su, and Jun Zhu. Hide-pet: Continual learning via hierarchical decomposition of parameter-efficient tuning, 2024b.

Mingyang Wang, Heike Adel, Lukas Lange, Jannik Strotgen, and Hinrich Schutze. Learn it or leave it: Module composition and pruning for continual learning, 2024c.

Mingyang Wang, Heike Adel, Lukas Lange, Jannik Strotgen, and Hinrich Schütze. Rehearsal-free modular and compositional continual learning for language models, 2024d.

Zhen Wang, Rameswar Panda, Leonid Karlinsky, R. Feris, Huan Sun, and Yoon Kim. Multitask prompt tuning enables parameter-efficient transfer learning, 2023c.

Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer G. Dy, and Tomas Pfister. Learning to prompt for continual learning, 2021.

Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun, Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer G. Dy, and Tomas Pfister. Dualprompt: Complementary prompting for rehearsal-free continual learning, 2022.

Kun-Juan Wei, Cheng Deng, and Xu Yang. Lifelong zero-shot learning, 2020.

Kun-Juan Wei, Cheng Deng, Xu Yang, and D. Tao. Incremental zero-shot learning, 2021.

Martin Wistuba, Prabhu Teja Sivaprasad, Lukas Balles, and Giovanni Zappella. Continual learning with low rank adaptation, 2023.

Tongtong Wu, Massimo Caccia, Zhuang Li, Yuan-Fang Li, G. Qi, and Gholamreza Haffari. Pretrained language model in continual learning: A comparative study, 2022.

Xinbo Wu, Max Hartman, Vidhata Arjun Jayaraman, and L. Varshney. Switchcit: Switching for continual instruction tuning of large language models, 2024.

Yanan Wu, Tengfei Liang, Songhe Feng, Yi Jin, Gengyu Lyu, Haojun Fei, and Yang Wang. Metazscil: A meta-learning approach for generalized zero-shot class incremental learning, 2023.

Kai Yi and Mohamed Elhoseiny. Domain-aware continual zero-shot learning, 2021.

Jiazuo Yu, Yunzhi Zhuge, Lu Zhang, Ping Hu, Dong Wang, Huchuan Lu, and You He. Boosting continual learning of vision-language models via mixture-of-experts adapters, 2024a.

Yu-Chu Yu, Chi-Pin Huang, Jr-Jen Chen, Kai-Po Chang, Yung-Hsuan Lai, Fu-En Yang, and Yu-Chiang Frank Wang. Select and distill: Selective dual-teacher knowledge transfer for continual learning on vision-language models, 2024b.

Wenxuan Zhang, P. Janson, Rahaf Aljundi, and Mohamed Elhoseiny. Overcoming general knowledge loss with selective parameter finetuning, 2023.

Junhao Zheng, Shengjie Qiu, and Qianli Ma. Learn or recall? revisiting incremental learning with pre-trained language models, 2023a.

Zangwei Zheng, Mingyu Ma, Kai Wang, Ziheng Qin, Xiangyu Yue, and Yang You. Preventing zero-shot transfer degradation in continual learning of vision-language models, 2023b.

Da-Wei Zhou, Han-Jia Ye, and De chuan Zhan. Few-shot class-incremental learning by sampling multi-phase tasks, 2022.

Çağatay Yıldız, Nishaanth Kanna Ravichandran, Prishruit Punia, Matthias Bethge, and B. Ermiş. Investigating continual pretraining in large language models: Insights and implications, 2024.