

Adaptive Memory Replay for Continual Learning

James Seale Smith^{* 1,2} Lazar Valkov¹ Shaunak Halbe² Vyshnavi Gutta²
 Rogerio Feris¹ Zsolt Kira² Leonid Karlinsky¹
¹MIT-IBM Watson AI Lab ²Georgia Institute of Technology

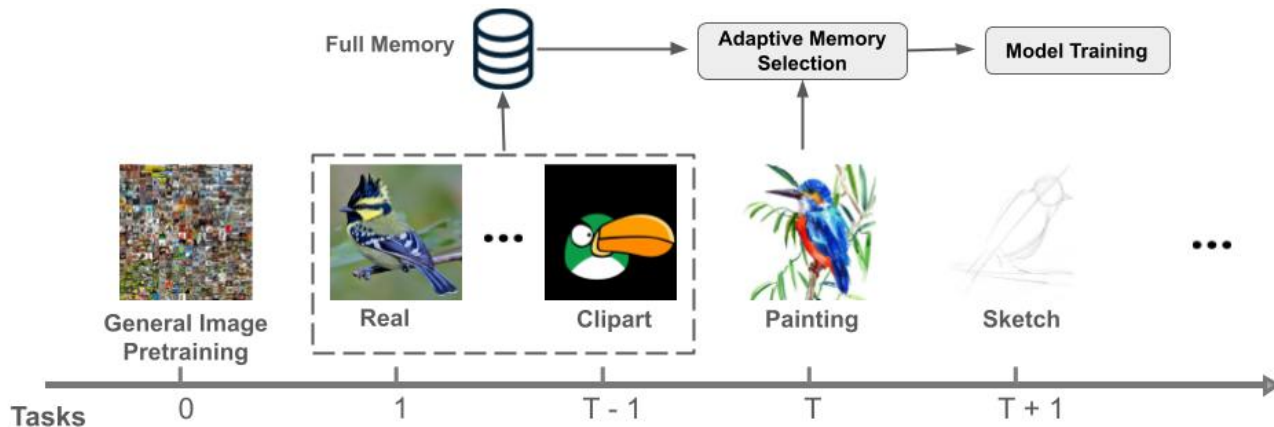


Figure 1. **Adaptive memory replay for continual pre-training.** In our setting, we begin with general image pretraining (Task 0) and transition to learn different tasks (e.g., Real, Clipart, Painting, Sketch) with full memory access to all past task data. We choose relevant samples for model training, thereby minimizing catastrophic forgetting while efficiently updating the model with new data. Further, we *replace* current task data with selected past task data, hence not adding training cost.

Abstract

Foundation Models (FMs) have become the hallmark of modern AI, however, these models are trained on massive data, leading to financially expensive training. Updating FMs as new data becomes available is important, however, can lead to ‘catastrophic forgetting’, where models underperform on tasks related to data sub-populations observed too long ago. This **continual learning** (CL) phenomenon has been extensively studied, but primarily in a setting where only a small amount of past data can be stored. We advocate for a new paradigm where memory is abundant, allowing us to keep all previous data, but computational resources are limited. In this setting, traditional replay-based CL approaches are outperformed by a simple baseline which replays past data selected uniformly at random [33], indicating that this setting necessitates a new approach. We address this by introducing a framework of **adaptive memory replay for continual learning**, where sampling of past data is phrased as a multi-armed bandit problem. We utilize Boltzmann sampling to derive a method which dynamically selects past data for train-

ing conditioned on the current task, assuming full data access and emphasizing training efficiency. Through extensive evaluations on both vision and language pre-training tasks, we demonstrate the effectiveness of our approach, which maintains high performance while reducing forgetting by up to 10% at no training efficiency cost.

1. Introduction

The concept of the Foundation Models (FMs) [1] has recently gained popularity and became ubiquitous in many downstream applications, including language [3, 30, 46], vision [34, 37, 47], and other application domains - advocating towards the ‘train-once-and-use-everywhere’ paradigm shift in AI/ML. One of the most attractive features of FMs is their ability for Zero-Shot [3] prompting, few-shot In-Context Learning (ICL) [3, 50], and great transferability to any task [30, 47]. This is due to their massive scale pre-training, often on billions [39] or trillions [30, 46] of data points. However, such power comes with high training costs. The pre-training data size is so large that normally each sample is observed only a few times or, as in language training, once (single epoch). Moreover, a common requirement for FMs is to rather frequently undergo

*Work done during internship at MIT-IBM Watson AI Lab.

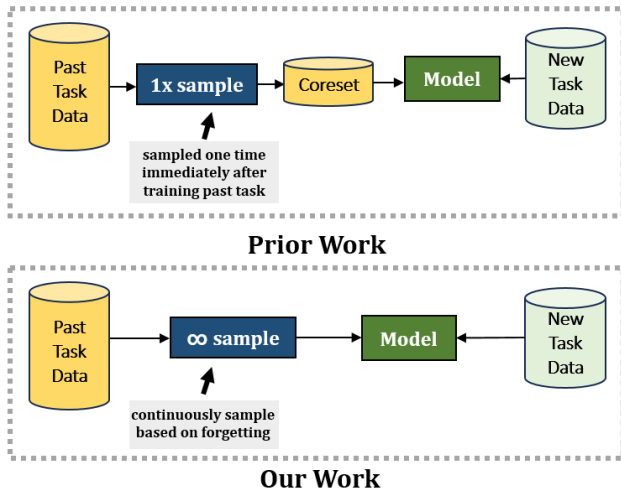


Figure 2. **Key difference of our work and prior work.** Prior work assumes that memory is expensive and constrains replay data to a fixed budget. Our work assumes memory is cheap and stores all replay data in memory, focusing on how to *dynamically select the most useful replay data* for computation-budgeted replay.

‘Extended Pre-Training’ (EPT) - a process of updating the model with new (massive) additional data intended to improve the model’s temporal currency. During EPT the original pre-training data cannot be naively replayed as, given the massive size of both pre-training and EPT data, it would effectively *double* the EPT cost (naturally assuming a single epoch and 50% replay mix for EPT as is customary in such cases). This would be prohibitive both in terms of the high cost (millions of dollars) as well as non-negligible negative environmental impact (extra heat emission).

However, neglecting past data during EPT is prone to the issue of *catastrophic forgetting* [19], where models updated with new data tend to underperform on previously seen data. This leads to an important question: *how can we adapt large-scale FM models to an ever-evolving world without compromising on performance or efficiency?*

The realm of *continual learning* offers some insights, but also limitations. While current benchmarks effectively highlight the challenge of catastrophic forgetting by training on non-overlapping data tasks sequentially, they are less applicable to (massive scale) EPT, as they either restrict themselves to limited memory storage (while in practical EPT all data, past and current, is usually available) and do not take into account the *training cost of replay*. For practical EPT, we argue the cost impact needs to be *minimal* in the sense that the ‘continual’ EPT needs to have similar cost as ‘naive’ EPT (disregarding old data and catastrophic forgetting issue). This is intuitive, as even the tiny overhead fraction due to replay will be applied as a factor to the training cost (measured in millions of dollars for the large-scale

models [30]).

This new setting of restricted computation and unlimited storage of past data has recently been explored for continual learning of image classification tasks [33]. The results of that work suggest that a simple baseline which randomly selects from all past data outperforms other replay-based CL methods. This creates a demand for a new approach which can better utilize all of the past data.

Taking inspiration from prior works [2, 8, 15] that have tried to select memory data intelligently by selecting the most representative samples of each memory replay dataset, we push the boundaries by considering not which past samples are the most representative (which is typically pre-decided before training future tasks), but rather *which samples most effectively prevent forgetting conditioned on the current task data* (which is decided *during the training of future tasks*). This notion is based on the intuitive concept that the model retains full access to previously seen data and that the ‘optimal’ replay data may be contingent upon the new data a model encounters during EPT.

We specifically propose an approach that dynamically adjusts the proportion of replay samples from each past task based on its propensity to be forgotten given the new task data. In such an *adaptive memory replay for continual learning* (visualized in Figure 1), our algorithm efficiently decides on the optimal allocation of memory replay samples among past tasks to minimize overall forgetting, under the vital consideration of how to do this without the requirement for drastic computation. We do this through a combination of bandit estimation and Boltzmann sampling from clusters of old datasets store in memory. We evaluate our replay strategy for both vision and language large-scale pre-training tasks. In particular, we propose and evaluate a *zero-cost protocol* that includes intelligent selection of both data to replay and reduction in the new EPT data to compensate for the (relatively small) extra cost of the selection algorithm itself.

In summary, we make the following contributions:

1. We present an *adaptive memory replay* for continual learning, a novel scheme inspired by a bandit estimation formulation that assumes full memory access and dynamically adjusts replay samples based on the new data, ensuring reduced forgetting.
2. Extensive evaluations demonstrate the efficacy of our method across both vision and language large-scale pre-training tasks.

2. Background and Related Work

Continual Learning: In the past few years, there has been significant progress in continual learning to alleviate catastrophic forgetting [29]. Regularization-based methods [22, 25, 40] modify the model parameters with additional regularization constraints to prevent catastrophic forgetting.

They store no data but explore extra regularization terms in the loss function to consolidate previous knowledge. Rehearsal approaches [5, 36, 36] memorize or generate a small fraction of data points for previous tasks and utilizes them to retain the task knowledge. Importantly, what data to retain is decided during the task itself, and subsequently used throughout future tasks. Expansion approaches expand a model’s architecture as new tasks are encountered; these are highly effective for applications where a model growing with tasks is practical [12, 23, 26, 28, 38]. Our work does not consider these methods because the model parameters grow with the number of tasks, but acknowledge that the contributions could be incorporated into these approaches.

Recently, prompt-tuning methods such as [43, 48, 49] outperformed rehearsal-based methods without using a replay buffer by learning a small number of insertable model instructions or prompts. Another line of research is the parameter isolation-based approaches [18, 27, 53] which focus on freezing the task-specific parameters and growing new branches for new tasks. [18] propose adapters which add a small number of parameters to the model for training on downstream tasks. Low-Rank Adaptation (LoRA) [20] extends on the above by using low-rank matrix counterparts of the original weights during fine-tuning, and keeps the actual weights frozen to further reduce inference costs.

Continual Learning in Transformers: The recent Vision Transformer (ViT) [10] has made a pure Transformer architecture scalable for large scale image classification and several works [11, 13, 24] have successfully applied the Transformers architecture for continual learning. In [24], for each new task, the model is copied and fixed to be used as the teacher model in the distillation phase. In [11], a unified model is learned by building upon a new architecture which dynamically expands the tokens processed by the last layer to mitigate forgetting. For each task, they learn a new task specific token per head using task-attention based decoder blocks. Recently, [13] proposes a method based on pre-trained Transformers while maintaining strict control of the memory usage and reaching state-of-the-art predictive performance. However the above methods either train a new transformer or need to fine-tune large pre-trained transformer models which requires significant compute in contrast to our objective of achieving optimal performance with limited compute.

Coreset Replay for Continual Learning: Rehearsal-based methods use a memory buffer to store selective samples of the previous tasks. These samples are then replayed with new task data to prevent catastrophic forgetting. A notable rehearsal-based method, Experience Replay (ER) from memory [36] interleaves the previous task samples with the current task data for optimizing the network parameters. E2E [8] deploys a herding algorithm to bolster coreset representativeness of the past task training data.

ERT [6] further extends ER by a balanced sampling strategy and bias control. Selective replay [15] proposes task-based rehearsal strategies for sample selection based on class-margin boundary, minimum confidence etc. DER++ [5] mixes rehearsal with a distillation loss for preventing catastrophic forgetting. HAL [9] integrates ER with an additional objective of keeping the predictions on anchor points of past tasks intact. MIR [35], GDumb [32] and ASER [41] store samples based on parameter updates, order of sample arrival and memory-based class boundaries respectively. Finally, ACE [2] explores various alternative population strategies to select coreset replay data. While the above methods consider the *representativeness of data stored in the memory*, they fail to take into account the *relationship between memory and the current task at hand*, as is explored in our work.

3. Preliminaries

Memory Replay In continual learning (CL)¹, the objective during task T , is to find parameters θ which minimize the loss L over the current dataset X_T and all previously seen datasets:

$$O := \min_{\theta} \left[\sum_{x \in X_T} \frac{L(x; \theta)}{|X_T|} + \sum_{t=1}^{T-1} \sum_{x \in X_t} \frac{L(x; \theta)}{|X_t|} \right] \quad (1)$$

Typically, CL approaches assume that past data cannot all be stored. Instead, experience replay approaches store a subset of past data from all previous tasks in a memory buffer $\mathcal{M} \subset \cup_{t=1}^{T-1} X_t$, where the size of \mathcal{M} is much smaller compared to the combined number of data points from all past tasks. These approaches [36] use \mathcal{M} to approximate the true objective (Eq. 1) and minimize:

$$\min_{\theta} \left[\sum_{x \in X_T} \frac{L(x; \theta)}{|X_T|} + \alpha \sum_{x \in \mathcal{M}} \frac{L(x; \theta)}{|\mathcal{M}|} \right] \quad (2)$$

where α is a hyper-parameter. The memory buffer is updated after each task but the total number of stored items is constant. The resulting method’s computational requirements scale well with the number of learned tasks, but the limited size of \mathcal{M} means that it becomes less effective at representing past data, as the number of learned tasks increases.

K-armed Bandits The (stochastic) K-armed bandit problem [4] considers a setting in which there are K available actions, referred to as *arms*. Performing one of the actions returns a stochastic reward drawn from an unknown distribution. The problem is selecting a number of actions in a

¹In our setting, the model does *not* have access to the task id during inference.

way which minimizes the expected regret, defined as the expected difference between the rewards obtained by always choosing the optimal action and the rewards obtained by following our strategy. At each step, a bandit strategy approximates the parameters of the reward distribution of each of the K actions. Thereafter, the strategy needs to select an action to perform. One such strategy is *Boltzmann Exploration* [21] which computes the mean of the observed rewards for each action, and then uses all means to define a categorical distribution, from which the choice of action is drawn. Finally, if the action’s distributions change between steps, the bandit problem is referred to as *non-stationary* [54].

4. Adaptive Memory Replay

In this section, we first modify the typical CL setting by challenging the restrictive assumption that past data cannot be accessed. Second, we modify the objective which we minimize, in order to better reflect the new CL setting. Third, we link the resulting problem to that of multi-armed bandit allocation [4] and detail our approach.

4.1. Replay: A New Perspective

For FM extended pre-training, we challenge the common CL assumption that past data is unavailable and alter the common CL setting after making two observations. First, data storage is cheap, therefore it is possible to store and sample from any dataset we have seen before. Second, computation is expensive, meaning that we cannot re-train on all of the past data in memory. Following this insight, we modify the CL setting so that all of the past data can be stored, but a CL algorithm’s computational demands to access and use the past data need to be constant and compensated during training to lead to zero extra cost (compared to naive training only on new data).

Being able to access all of the past data naturally leads to a rehearsal-based approach where one can only replay a limited subset \mathcal{M} , due to computational constraints. In contrast to the typical memory replay solutions, \mathcal{M} is not fixed while training on a new task, but is instead allowed to adapt to the new data and the changing model while maintaining training efficiency. Therefore, we refer to it as an *adaptive memory buffer*.

We begin developing our approach by modifying the objective function to better reflect our goal. First, we express the main objective (Eq. 1) in terms of its forgetting \mathcal{F} on past data, compared to the performance of the optimal parameters for the previous task:

$$O = \min_{\theta} \left[\sum_{x \in X_T} \frac{L(x; \theta)}{|X_T|} + \sum_{t=1}^{T-1} \sum_{x \in X_t} \frac{\mathcal{F}(x; \theta)}{|X_t|} + C \right]$$

where C is a constant. This change reflects the fact that we fine-tune the previously optimal model on the new task and that our focus is on minimizing forgetting, rather than improving our performance on past data. Next, we define the optimal memory buffer $\mathcal{M}^*(\theta)$ as the subset of past data for which our model has the maximum forgetting. We then following replay-based methods (Eq. 2) and similarly optimize:

$$\min_{\theta} \left[\sum_{x \in X_T} \frac{L(x; \theta)}{|X_T|} + \alpha \sum_{x \in \mathcal{M}^*(\theta)} \frac{\mathcal{F}(x; \theta)}{|\mathcal{M}^*(\theta)|} + C \right] \quad (3)$$

Minimizing this objective leads to our rehearsal-based CL algorithm with adaptive memory $\mathcal{M}^*(\theta)$ which replays the past data points $x \in \mathcal{M}^*(\theta)$ that have currently been forgotten the most. Importantly, to minimize the overhead which replaying imposes on the training process, we keep the number of computed gradients constant by discarding a portion of the data on the new task. To do this, we remove $|\mathcal{M}^*(\theta)|$ randomly selected data points for each batch of inputs from the new task, thus keeping the total number of processed inputs the same.

In order to compute Eq. 3, we need to be able to select $\mathcal{M}^*(\theta)$ — the subset of all past data with the highest forgetting. This subset changes as we update the parameters θ , and it is computationally infeasible to evaluate the forgetting of each of the past data points. Instead, we seek to divide all past data into clusters A_i of items expected to have similar forgetting values.

This allows us to infer the forgetting values of the data points in a cluster, based on a small number of evaluations, and use this to select data points for our adaptive memory buffer which exhibits a high amount of forgetting. Currently, we place all of the data from the same previous task into the same cluster, i.e. $A_i = X_i$, and assume that it would exhibit similar forgetting values. We leave more elaborate clustering techniques for future work.

4.2. Adaptive Memory as a Bandit Optimization

Formally, we divide all past data into K disjoint subsets A_i , s.t. $\cup_{i=1}^K A_i = \cup_{i=1}^{T-1} X_i$, where the forgetting of an input is distributed according to a subset-specific distribution: $\mathcal{F}(x) \sim \mathcal{D}_i$, for $x \in A_i$. For the parameters θ_j at training iteration j , we would like to select a subset $\mathcal{M}(\theta_j)$ which exhibits close to the worst forgetting, minimizing the following quantity:

$$r(\mathcal{M}(\theta_j)) := \sum_{x \in \mathcal{M}^*(\theta_j)} \mathcal{F}(x; \theta_j) - \sum_{x \in \mathcal{M}(\theta_j)} \mathcal{F}(x; \theta_j).$$

We frame this as a non-stationary K -armed bandit (KAB) problem [42], where pulling an arm and receiving a reward corresponds to sampling a data point from a cluster and evaluating its forgetting. Then, at each training iteration,

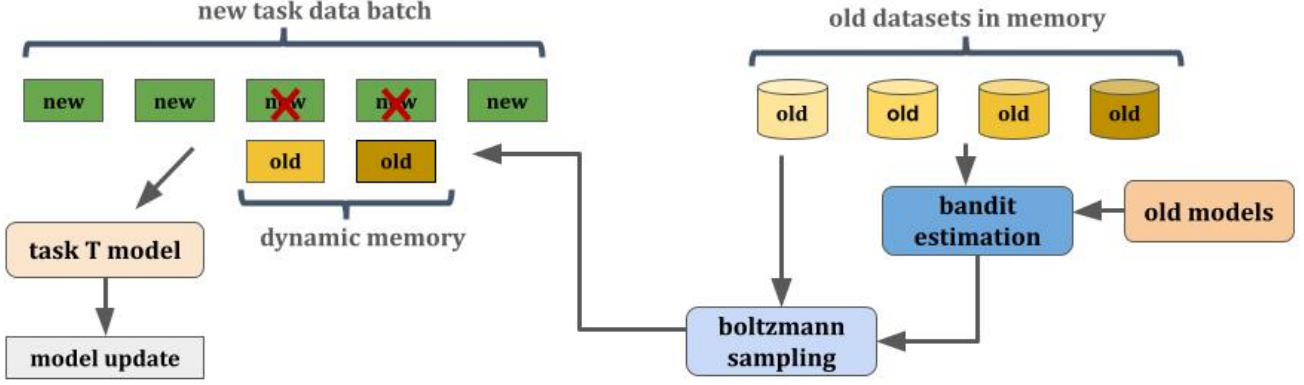


Figure 3. **Overview of our adaptive memory replay approach.** New task data is integrated with selectively rehearsed old data from full replay memory to update the task model. Unlike simple iid replay, our rehearsal data is chosen through a combination of bandit estimation and Boltzmann sampling from clusters of old datasets stored in memory. To reduce computation costs associated with data replay, we randomly discard samples from the training data to be replaced with the selected replay data. This ensures a cost-effective balance between incorporating new information and retaining knowledge of previous tasks, thus mitigating catastrophic forgetting with minimal computational overhead.

we have to choose which of the K arms to pull in order to select $|\mathcal{M}|$ data points with maximum forgetting. As we select different M over all training steps, we would like to reduce the *expected regret*: $\mathbb{E}R := \mathbb{E}[\sum_j r(\mathcal{M}(\theta_j))]$, which is the expected difference in forgetting values between \mathcal{M} and \mathcal{M}^* over all training steps. In this work, we implement the Boltzmann Exploration [21] approach which, at training step j , approximates the mean reward of each arm, denoted by $\mu_i^{(j)}$, and then uses all arms’ means as parameters for a categorical distribution over the choice of arms to pull. This distribution is then used to sample $|\mathcal{M}|$ arms and in turn sample the adaptive memory buffer $\mathcal{M}(\theta_j)$.

To approximate the mean forgetting values $\mu_i^{(j)}$ of cluster i at training step j , we first sample a small number of data points from the cluster A_i and evaluate the average of their forgetting values — $\bar{f}_i^{(j)}$. We would like to compute $\mu_i^{(j)}$ based on the previously computed mean value $\mu_i^{(j-1)}$ and the currently computed forgetting average $\bar{f}_i^{(j)}$. However, we note that the forgetting values depend on our model’s parameters θ_j , thus change between training iterations. We account for this by using a moving average, which is used for KAB when the underlying distributions are non-stationary [44]: $\mu_i^{(j)} = \beta \bar{f}_i^{(j)} + (1 - \beta) \mu_i^{(j-1)}$.

Once we have approximated the mean forgetting values for all clusters, we use them to create a categorical distribution over the choice of clusters, with the help of the tempered softmax function [17]. We compute: $p(A_i)^{(j)} = \exp\{\mu_i^{(j)}/t\}/Z$, where t is the temperature hyperparameter and Z is the normalization constant. Afterwards, we use this distribution to sample $|\mathcal{M}|$ cluster indices. Finally, we sample one input from each selected cluster, uniformly

Algorithm 1 Adaptive Memory Replay

Require: Pre-trained model with parameters θ , datasets $\{X_i\}_{i=1}^T$, constant C , learning rate η , replay buffer size $|\mathcal{M}|$, and temperature t

Ensure: Updated model θ that has minimal forgetting

- 1: Initialize replay buffer $\mathcal{M} \leftarrow \emptyset$
 - 2: Initialize clusters $A_i \leftarrow X_i$ for each past dataset
 - 3: Initialize mean forgetting $\mu_i^{(0)} \leftarrow 0$ for each cluster A_i
 - 4: **for** each training iteration j **do**
 - 5: Sample a batch B from new data X_T
 - 6: **for** each cluster A_i **do**
 - 7: Sample points from A_i to est. forgetting $\bar{f}_i^{(j)}$
 - 8: Update means $\mu_i^{(j)} \leftarrow \beta \bar{f}_i^{(j)} + (1 - \beta) \mu_i^{(j-1)}$
 - 9: **end for**
 - 10: Compute distribution $p(A_i)^{(j)}$ using softmax
 - 11: Sample based on $p(A_i)^{(j)}$ to fill buffer $\mathcal{M}(\theta_j)$
 - 12: Remove $|\mathcal{M}(\theta_j)|$ random data points from B
 - 13: Update model to minimize objective fn (Eq. 3)
 - 14: **end for**
-

at random, and combine the samples to create the adaptive memory buffer for the current training step. **Our full method is summarized in Algorithm 1.**

5. Experiments

We evaluate the efficacy of our proposed adaptive memory replay for continual learning of FM pre-training (i.e., extended pre-training) in both the vision and language domains. We utilize two distinct pre-trained models as our backbones for these experiments: a Vision Masked Autoencoder (MAE) [16] pre-trained on ImageNet-1K for vision-

Table 1. **Results on DomainNet [31]**. Results are included for 6 tasks. Final Loss and Forgetting are reported using test data loss across all tasks after training on the entire sequence, normalized by the Oracle (full re-training on all datasets at each task). Training time for each approach is normalized using the Oracle training time. The ‘0-cost’ result indicates the number of training steps of our approach is reduced to align with the training time of naive fine-tuning.

Approach	Normalized MAE Loss		Normalized Training Time		
	Final Loss (↓)	Forgetting (↓)	Total	Selecting Replay Data	Training Model
Oracle	0.0%	0.0%	100.0%	0.0%	100.0%
Base	100%	0.0%	0.0%	0.0%	0.0%
Naive	54.73%	70.95%	34.19%	0.0%	34.19%
Standard Rehearsal	30.41%	22.97%	34.61%	0.0%	34.61%
Our Rehearsal	23.65%	4.39%	36.07%	2.45%	33.61%
Our Rehearsal (0 Cost)	26.69%	12.84%	34.26%	2.31%	31.93%

Table 2. **Results on Medical MNIST [52]**. Results are included for 5 tasks. Final Loss and Forgetting are reported using test data loss across all tasks after training on the entire sequence, normalized by the Oracle (full re-training on all datasets at each task). Training time for each approach is normalized using the Oracle training time. The ‘0-cost’ result indicates the number of training steps of our approach is reduced to align with the training time of naive fine-tuning.

Approach	Normalized MAE Loss		Normalized Training Time		
	Final Loss (↓)	Forgetting (↓)	Total	Selecting Replay Data	Training Model
Oracle	0.0%	0.0%	100.0%	0.0%	100.0%
Base	100%	0.0%	0.0%	0.0%	0.0%
Naive	82.36%	98.80%	35.22	0.0%	35.22
Standard Rehearsal	12.57%	5.19%	36.05%	0.0%	36.05%
Our Rehearsal	9.73%	1.61%	37.39%	1.69%	35.71%
Our Rehearsal (0 Cost)	11.44%	1.98%	35.52%	1.60%	33.92%

related tasks, and LLaMA [45] with 7 billion parameters for language experiments. The evaluation metrics for our experiments are twofold: test data **Final Loss** and test data loss **Forgetting**. The Loss metrics are standard for FM training (especially in Language Modeling) as they are known to correlate with downstream use due to the massive pre-training volumes and de-facto seeing most of the samples only once. These metrics are normalized between 0% and 100%, where 0% represents an offline upper bound with all data trained independently and identically distributed (iid), and 100% corresponds to the performance of the pre-trained model without any fine-tuning.

Because our primary contribution lies in our novel perspective of full-memory replay, we compare our approach with full memory-access iid data replay as opposed to typical continual learning methods. We hypothesize that coresets selection replay methods are effectively upper-bounded by full iid replay, given their goal of identifying the most representative data for replay (rather than our perspective of identifying the most forgotten data for replay as a function of the current data). Consequently, we do not compare our

method against sampled data replay in our main results tables as we store all data in memory.

Furthermore, our experiments are designed to demonstrate the advantages of our adaptive memory replay approach over traditional iid replay, especially in terms of computational efficiency and reduced forgetting. We consider gains of our approach to be orthogonal to the realms of non-replay regularization-based continual learning methods, and thus these comparisons are not the main focus of our results. Besides, from the perspective of computational efficiency, recent work has found such approaches to be impractical for computationally bounded continual learning [14].

The hyperparameters for our experiments were meticulously chosen based on a series of small task experiments. We update our model on 10,000 new data examples per task. In the interest of computational resources for the larger Llama model, we approximate the training of all the model parameters with LoRA finetuning [20] in the language modeling experiments. In our experience, conclusions attained for LoRA finetuning reflect the same in full model train-

Table 3. **Results on Synthetic Visual Concepts [7]**. Results are included for 4 tasks. Final Loss and Forgetting are reported using test data loss across all tasks after training on the entire sequence, normalized by the Oracle (full re-training on all datasets at each task). Negative forgetting indicates forward transfer (which is only present in this unique dataset). Training time for each approach is normalized using the Oracle training time. The ‘0-cost’ result indicates the number of training steps of our approach is reduced to align with the training time of naive fine-tuning.

Approach	Normalized MAE Loss		Normalized Training Time		
	Final Loss (↓)	Forgetting (↓)	Total	Selecting Replay Data	Training Model
Oracle	0.0%	0.0%	100.0%	0.0%	100.0%
Base	100%	0.0%	0.0%	0.0%	0.0%
Naive	52.11%	27.93%	38.87%	0.0%	38.87%
Standard Rehearsal	35.86%	-6.61%	41.96%	0.0%	41.96%
Our Rehearsal	34.82%	-7.78%	46.40%	44.94%	1.46%
Our Rehearsal (0 Cost)	35.36%	-7.43%	37.31%	1.36%	35.94%

Table 4. **Results on 5-task Causal Language Modeling benchmark**. Final Loss and Forgetting are reported using test data loss across all tasks after training on the entire sequence, normalized by the Oracle (full re-training on all datasets at each task). Training time for each approach is normalized using the Oracle training time. The ‘0-cost’ result indicates the number of training steps of our approach is reduced to align with the training time of naive fine-tuning.

Approach	Normalized MAE Loss		Normalized Training Time		
	Final Loss (↓)	Forgetting (↓)	Total	Selecting Replay Data	Training Model
Oracle	0.0%	0.0%	100.0%	0.0%	100.0%
Base	100%	0.0%	0.0%	0.0%	0.0%
Naive	137.54%	180.36%	36.05%	0.0%	36.05%
Standard Rehearsal	44.11%	76.97%	40.32%	0.0%	40.32%
Our Rehearsal	12.88%	38.06%	45.76%	40.33%	5.42%
Our Rehearsal (0 Cost)	14.60%	39.75%	36.80%	4.04%	32.76%

ing. We use a learning rate of $2e - 5$ for full model fine-tuning and $2e - 4$ for LoRA-based fine-tuning. For our proposed adaptive memory replay bandit scheme, we found that a temperature of $t = 0.1$ and forgetting mean update ratio of $\beta = 0.01$ performed best. We compose our replay batches for both iid replay and our adaptive memory replay with a 1:1 ratio of replay data to new task training data. We conducted evaluations on a hold-out test dataset comprising 500 samples per dataset.

5.1. Results for Vision SSL

In Tables 1,2,3, we benchmark our proposed approach on 3 different continual pre-training sequences composed of vision datasets. Our goal was to demonstrate the robustness of our findings with a variety of unique and practical dataset sequences. The first dataset is the DomainNet [31] dataset (Table 1), containing 6 different domains of common objects. The next is the Medical MNIST dataset [52] (Table 2), from which we sampled 5 standardized biomedical image datasets containing the highest number of samples. Finally,

we use 4 attribute splits from the Synthetic Visual Concepts (SyViC) dataset [7] (3).

Our results demonstrate the advantages of our adaptive memory replay method in the vision domain. Our approach consistently outperforms full memory iid replay (which serves as an upper bound for other replay-based continual learning methods that sample from a limited coreset), achieving lower final loss and forgetting rates. The slight increase in normalized training time is negligible compared to the performance gains, and furthermore, we show a 0-cost result where we reduce the number of training steps of our approach to align with the training time of naive fine-tuning, and show that even this result outperforms iid replay in all three benchmarks.

We note that our strongest performance gains come from the DomainNet results. The gains for the medical data sequence and synthetic data pre-training are much more modest, yet remain pronounced. The synthetic data sequence is interesting in that forward transfer (i.e., negative forgetting) appears in all results - however, our method still has more

forward transfer compared to iid replay. In practical terms, our results imply that vision systems equipped with our continual learning strategy would exhibit improved robustness over time, adapting to new data without significant loss of prior knowledge or computational costs. We re-iterate that there is much more room for improvement from our full-memory continual learning perspective - advanced strategies can close the gap between our method and the upper bound with fixed time costs by exploring interesting questions such as *how to better cluster the data* and *which new data is more or less favorable to discard*.

5.2. Results for Causal Language Modeling

Our approach is further affirmed through our language experiments using the Llama model. In Table 4, we benchmark on a 5-dataset sequence using datasets from Huggingface [51]. These datasets were chosen based on the significant variations in loss observed post fine-tuning, thus providing a rigorous test for our approach. The datasets encompass a broad range of language tasks, ensuring that our results are representative of diverse language modeling scenarios.

The performance of our adaptive memory replay in our language experiments mirrors the success observed in the vision tasks. We see a significant reduction in both forgetting and final loss compared to the iid full-memory replay. Furthermore, the ‘0-cost’ variant of our method is particularly noteworthy, as it manages to retain a high level of performance without additional computational expenditure compared to naive fine-tuning. This aspect is crucial for applications where computational resources are limited, especially fitting LLM extended pre-training where due to high data volumes and enormous model sizes even a tiny fraction of extra cost is intolerable.

5.3. Additional Analysis

In Figure 4, we present a comprehensive comparison of the final loss versus training time for our adaptive memory replay method against the Oracle, using the Synthetic Visual Concepts dataset sequence [7]. This plot demonstrates how our method converges towards the Oracle’s performance as we increase the compute budget of our method (via using more replay samples and discarding fewer new task samples). With a limited budget, there is a notable difference in the final loss between our method and the Oracle. However, as training progresses, our method steadily approaches the Oracle’s level of performance, matching and even outperforming Oracle (which as a fixed compute budget itself, pre-defined by the number of training steps we use) with a lower compute cost. The horizontal dotted line marks the point at which our approach reaches the Oracle’s normalized loss, showcasing the efficiency of our method in terms of both loss minimization and computational time. This re-

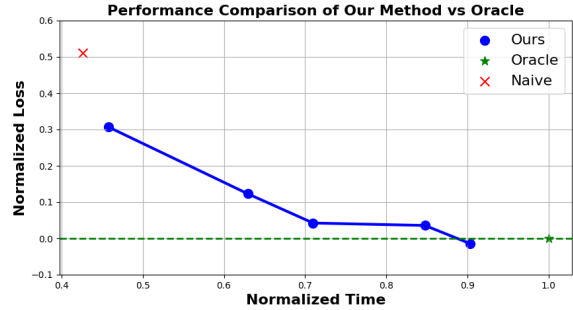


Figure 4. Final Loss vs Training Time for adaptive memory replay vs Oracle using the Synthetic Visual Concepts [7] sequence.

sult is significant as it not only validates the effectiveness of our adaptive memory replay in reducing the final loss but also highlights its capability to achieve this with a substantially lower training time.

6. Conclusion

In this paper, we underscored the importance of adapting machine learning methodologies to the ever-evolving demands of real-world large-scale continual learning. Our findings, rooted in extensive evaluations across both vision and language tasks, validate the advantages of our *adaptive memory replay for continual learning*. By dynamically selecting past training data samples, our method offers a nuanced balance, ensuring minimized forgetting without imposing costly computational requirements. Ultimately, we hope to inspire methodologies that are both computationally efficient and effective in real-world continual learning scenarios, where data access and computational resources are bound by practical constraints. This research trajectory serves as a stepping stone towards the development of continually learning systems that efficiently and intelligently utilize all available data, enhancing their learning and adaptability across a series of tasks throughout their life-cycle.

Limitations & Future work. Future work should focus on refining the adaptive memory replay mechanism, particularly exploring more sophisticated bandit-based selection strategies to further enhance the balance between retaining old knowledge and accommodating new information. The decision of which data to discard during the replay phase also warrants deeper investigation to avoid potential loss of critical information (and furthermore potential unrealized gains in discarding repetitive or similar samples). There is also a great need to develop advancing clustering techniques to capture the subtleties of data evolution which can lead to more representative memory buffers. In addition, bringing greater realism into continual learning models by incorporating real-world constraints and scenarios will be crucial, such as blurred task boundaries and online clustering.

References

- [1] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avani Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models, 2022. **1**
- [2] Daniel Brignac, Niels Lobo, and Abhijit Mahalanobis. Improving replay sample selection and storage for less forgetting in continual learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3540–3549, 2023. **2, 3**
- [3] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020. **1**
- [4] Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1): 1–122, 2012. **3, 4**
- [5] Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: a strong, simple baseline. *Advances in neural information processing systems*, 33:15920–15930, 2020. **3**
- [6] Pietro Buzzega, Matteo Boschini, Angelo Porrello, and Simone Calderara. Rethinking experience replay: a bag of tricks for continual learning. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2180–2187. IEEE, 2021. **3**
- [7] Paola Cascante-Bonilla, Khaled Shehada, James Seale Smith, Sivan Doveh, Donghyun Kim, Rameswar Panda, Gul Varol, Aude Oliva, Vicente Ordonez, Rogerio Feris, et al. Going beyond nouns with vision & language models using synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20155–20165, 2023. **7, 8**
- [8] Francisco M Castro, Manuel J Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. In *Proceedings of the European conference on computer vision (ECCV)*, pages 233–248, 2018. **2, 3**
- [9] Arslan Chaudhry, Albert Gordo, Puneet Dokania, Philip Torr, and David Lopez-Paz. Using hindsight to anchor past knowledge in continual learning. In *Proceedings of the AAAI conference on artificial intelligence*, pages 6993–7001, 2021. **3**
- [10] A Dosovitskiy, L Beyer, A Kolesnikov, D Weissenborn, X Zhai, and T Unterthiner. Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. **3**
- [11] Arthur Douillard, Alexandre Ramé, Guillaume Couairon, and Matthieu Cord. Dytox: Transformers for continual learning with dynamic token expansion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9285–9295, 2022. **3**
- [12] Sayna Ebrahimi, Franziska Meier, Roberto Calandra, Trevor Darrell, and Marcus Rohrbach. Adversarial continual learning. *arXiv preprint arXiv:2003.09553*, 2020. **3**
- [13] Beyza Ermis, Giovanni Zappella, Martin Wistuba, Aditya Rawal, and Cedric Archambeau. Memory efficient continual learning with transformers. *Advances in Neural Information Processing Systems*, 35:10629–10642, 2022. **3**
- [14] Yasir Ghunaim, Adel Bibi, Kumail Alhamoud, Motasem Alfara, Hasan Abed Al Kader Hammoud, Ameya Prabhu, Philip HS Torr, and Bernard Ghanem. Real-time evaluation in online continual learning: A new hope. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11888–11897, 2023. **6**
- [15] Tyler L Hayes and Christopher Kanan. Selective replay enhances learning in online continual analogical reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3502–3512, 2021. **2, 3**
- [16] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022. **5**
- [17] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. **5**
- [18] Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona

- Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning*, pages 2790–2799. PMLR, 2019. 3
- [19] Yen-Chang Hsu, Yen-Cheng Liu, Anita Ramasamy, and Zsolt Kira. Re-evaluating continual learning scenarios: A categorization and case for strong baselines. *arXiv preprint arXiv:1810.12488*, 2018. 2
- [20] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021. 3, 6
- [21] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996. 4, 5
- [22] 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. 2
- [23] Soochan Lee, Junsoo Ha, Dongsu Zhang, and Gunhee Kim. A neural dirichlet process mixture model for task-free continual learning. *arXiv preprint arXiv:2001.00689*, 2020. 3
- [24] Duo Li, Guimei Cao, Yunlu Xu, Zhazhan Cheng, and Yi Niu. Technical report for iccv 2021 challenge sslad-track3b: Transformers are better continual learners. *arXiv preprint arXiv:2201.04924*, 2022. 3
- [25] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 2
- [26] Vincenzo Lomonaco and Davide Maltoni. Core50: a new dataset and benchmark for continuous object recognition. *arXiv preprint arXiv:1705.03550*, 2017. 3
- [27] Arun Mallya, Dillon Davis, and Svetlana Lazebnik. Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In *Proceedings of the European conference on computer vision (ECCV)*, pages 67–82, 2018. 3
- [28] Davide Maltoni and Vincenzo Lomonaco. Continuous learning in single-incremental-task scenarios. *Neural Networks*, 116:56–73, 2019. 3
- [29] Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, pages 109–165. Elsevier, 1989. 2
- [30] OpenAI. Gpt-4 technical report, 2023. 1, 2
- [31] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1406–1415, 2019. 6, 7
- [32] Ameya Prabhu, Philip HS Torr, and Puneet K Dokania. Gdumb: A simple approach that questions our progress in continual learning. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 524–540. Springer, 2020. 3
- [33] Ameya Prabhu, Hasan Abed Al Kader Hammoud, Puneet K Dokania, Philip HS Torr, Ser-Nam Lim, Bernard Ghanem, and Adel Bibi. Computationally budgeted continual learning: What does matter? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3698–3707, 2023. 1, 2
- [34] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021. 1
- [35] Aljundi Rahaf and Caccia Lucas. Online continual learning with maximally interfered retrieval. In *NIPS*, 2019. 3
- [36] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. *Advances in Neural Information Processing Systems*, 32, 2019. 3
- [37] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2022. 1
- [38] 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. 3
- [39] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. Laion-5b: An open large-scale dataset for training next generation image-text models, 2022. 1
- [40] Joan Serra, Didac Suris, Marius Miron, and Alexandros Karatzoglou. Overcoming catastrophic forgetting with hard attention to the task. In *International conference on machine learning*, pages 4548–4557. PMLR, 2018. 2
- [41] Dongsub Shim, Zheda Mai, Jihwan Jeong, Scott Sanner, Hyunwoo Kim, and Jongseong Jang. Online class-incremental continual learning with adversarial shapley value. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9630–9638, 2021. 3
- [42] Aleksandrs Slivkins et al. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019. 4
- [43] James Seale Smith, Leonid Karlinsky, Vyshnavi Gutta, Paola Cascante-Bonilla, Donghyun Kim, Assaf Arbelle, Rameswar Panda, Rogerio Feris, and Zsolt Kira. Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11909–11919, 2023. 3
- [44] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018. 5
- [45] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023. 6
- [46] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste

- Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models, 2023. [1](#)
- [47] Junke Wang, Dongdong Chen, Zuxuan Wu, Chong Luo, Luowei Zhou, Yucheng Zhao, Yujia Xie, Ce Liu, Yu-Gang Jiang, and Lu Yuan. Omnivl:one foundation model for image-language and video-language tasks, 2022. [1](#)
- [48] 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. [3](#)
- [49] 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. [3](#)
- [50] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023. [1](#)
- [51] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Huggingface’s transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*, 2019. [8](#)
- [52] Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41, 2023. [6](#), [7](#)
- [53] Jaehong Yoon, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. Lifelong learning with dynamically expandable networks. *arXiv preprint arXiv:1708.01547*, 2017. [3](#)
- [54] Peng Zhao, Lijun Zhang, Yuan Jiang, and Zhi-Hua Zhou. A simple approach for non-stationary linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 746–755. PMLR, 2020. [4](#)