ϕ-DPO: Fairness Direct Preference Optimization Approach to Continual Learning in Large Multimodal Models

Supplementary Material

A. Proof of Lemmas

A.1. Proof of Lemma 1

The DPO loss in Eqn. (7) can be rewrite as follows:

$$\begin{aligned}\mathcal{L}_{\text{DPO}}(\pi_t, \pi_{t-1}) &= \mathbb{E}_{x,y^+,y}\left[\ell_\beta\left(\Delta_t(y^+, y^-)\right)\right] \\ \ell_\beta(u) &= \log\Big(1+\exp(\beta u)\Big) \\ \Delta_t(y^+, y^-) &= \Big(\log \pi_t(y^+|x) - \log \pi_t(y^-|x)\Big) \\ &\quad - \Big(\log \pi_{t-1}(y^+|x) - \log \pi_{t-1}(y^-|x)\Big)\end{aligned} \tag{18}$$

Lemma 4 ***Pairwise Logistic Lower Bound by Margin***. *For any $u \in \mathbb{R}$ and $\beta > 0$,*

$$\ell_\beta(u) = \log(1 + e^{-\beta u}) \;\; \geq \;\; \log 2 - \tfrac{\beta}{2} u.$$

Proof. Since $\log(1+e^{-v})$ is convex and symmetric around $v = 0$ in the sense that its tangent at 0 is $\log 2 - \frac{1}{2}v$, the global underestimator follows from convexity $\log(1+e^{-v}) \;\geq\; \log 2 - \frac{1}{2}v$. Then, we substitute $v = \beta u$ follow by taking expectation over pairs:

$$\mathbb{E}[\ell_\beta(\Delta_t(y^+, y^-)) \geq \log 2 - \frac{\beta}{2}\,\mathbb{E}[\Delta_t(y^+, y^-)]. \tag{19}$$

As a result, the small value of the DPO loss forces large average margin $\mathbb{E}[\Delta_t(y^+, y^-)]$. In other words, the smaller value of DPO loss enforces the model's preference for well-retained y^+ responses stronger than for forgotten ones y^-.

Lemma 5 ***Average Margin Controls an Integral Probability Metrics (IPM)***. *Let $\mathcal{F}_1$ be the set of all 1-Lipschitz functions, If the reward function r is L-Lipschitz, then $r/L \in \mathcal{F}_1$. Then, for any pair of marginals P^+, P^- where $y^+ \sim P^+(y^+)$ and $y^- \sim P^-(y^-)$, we have*

$$\begin{aligned}\mathbb{E}[\Delta_t(y^+, y^-)] &= \mathbb{E}_{P^+}[r(x, y^+)] - \mathbb{E}_{P^-}[r(x, y^-)] \\ &\leq L\,\text{IPM}_{\mathcal{F}_1}(P^+, P^-) \leq L W_1(P^+, P^-) \\ \text{IPM}_{\mathcal{F}_1}(P^+, P^-) &= \sup_{f \in \mathcal{F}_1}\Big(\mathbb{E}_{y^+\sim P^+}[f(y^+)] - \mathbb{E}_{y^-\sim P^-}[f(y^-)]\Big)\end{aligned} \tag{20}$$

where W_1 is the 1-Wasserstein distance.

Proof. By definition of IPM over 1-Lipschitz functions and r/L is admissible, the above inequality is the Kantorovich-Rubinstein duality IPM over 1-Lipschitz functions equals W_1 provided in [23]. In addition, although Lemma 5 requires r to be an L-Lipschitz function, we have observed that a local L-Lipschitz reward function, which is satisfied in our setup, is also sufficient. Indeed, prior studies [34] rigorously derives bounds on the local Lipschitz constants of deep neural networks and shows they can be meaningfully controlled despite huge global constants. This result indicate that the LLMs behave smoothly around their high-probability outputs. In our context, we only need the log-ratio to be Lipschitz on the region visited by preference pairs, not globally over all possible outputs. Transformer-based LLMs incorporate norm control, weight decay, and normalization layers, which implicitly bound gradient magnitudes and curtail abrupt jumps in logits. Empirically, small semantic perturbations rarely cause extreme changes in logits, suggesting local smoothness holds on the data manifold [32, 111]. Thus, a locally valid Lipschitz constant suffices the requirement of Lemma 5. Therefore, while LLMs may not be globally Lipschitz, they plausibly satisfy the needed local Lipschitz continuity in the regions relevant to DPO, making Lemma 5 still valid in practice.

In addition, it can be shown that $W_1(P^+, P^-) \leq 3W_1(\pi_t, \pi_{t-1})$ follows naturally from the triangle inequality of the Wasserstein distance. In particular, if the preference distributions P^+ and P^- remain close to the current and previous policies, respectively, such that $W_1(P^+, \pi_t) \leq W_1(\pi_t, \pi_{t-1})$ and $W_1(P^-, \pi_{t-1}) \leq W_1(\pi_t, \pi_{t-1})$, then we obtain

$$\begin{aligned}W_1(P^+, P^-) &\leq W_1(P^+, \pi_t) + W_1(\pi_t, \pi_{t-1}) + W_1(\pi_{t-1}, P^-) \\ &\leq 3W_1(\pi_t, \pi_{t-1})\end{aligned} \tag{21}$$

This conditions are typically satisfied in the DPO training, where preference sampling is a monotone and non-expansive process, e.g., sampling candidates from a mixture (please refer to Remark 1 in Section A.2). In the context of continual learning of LMMs, the inequality $W_1(P^+, P^-) \leq 3W_1(\pi_t, \pi_{t-1})$ implies that the discrepancy between well-retained and forgotten knowledge is bounded by the overall policy shift between two learning steps. Intuitively, both P^+ and P^- remain anchored around their respective policies, so the overall variation between them is bounded by a constant multiple of the inter-policy shift $W_1(\pi_t, \pi_{t-1})$. Concurrently, both P^+ and P^- remain anchored around their respective policies: P^+ near the current policy π_t and P^- near the previous policy π_{t-1}. Thus, if the model update between tasks is smooth, the semantic drift between memory retention and forgetting remains limited. This highlights that our continual DPO training enforces a stable adaptation process, where catastrophic forgetting is controlled by bounding the inter-policy Wasserstein distance.

Now, the inequality in Eqn. (20) can be further rewritten as follows:

$$\begin{aligned}\mathbb{E}[\Delta_t(y^+, y^-)] &= \mathbb{E}_{P^+}[r(x, y^+)] - \mathbb{E}_{P^-}[r(x, y^-)] \\ &\leq 3LW_1(\pi_t, \pi_{t-1})\end{aligned} \tag{22}$$

Lemma 6 ***A Transport–Entropy Inequality***. *Since the output probability $p(y|x)$ produced by the LMM of previous learning step π_{t_1} is computed based on the softmax on the logit scores of token y, we can view π_{t-1} as a Boltzmann distribution over token sequences. Then, without a strict argument, we assume that π_{t-1} satisfies the Talagrand $T_2(C_0)$ inequality [49, 103]:*

$$W_2^2(\mu, \pi_{t-1}) \leq 2C_0\, D_{\text{KL}}(\mu \,\|\, \pi_{t-1}) \quad \textit{for all } \mu, \tag{23}$$

Then, since $W_1 \le W_2$, by substituting μ by π_t, we have the final inequality as follows:

$$W_1(\pi_t, \pi_{t-1}) \;\le\; \sqrt{2C_0\, D_{\mathrm{KL}}(\pi_t \,\|\, \pi_{t-1})} \tag{24}$$

Proof. The proof of Talagrand $T_2(C_0)$ inequality has been shown in prior studies [9].

Proof of Lemma 1. From Lemmas 4-6, we have

$$\begin{aligned}
\mathcal{L}_{\mathrm{DPO}}(\theta; x) &\ge \log 2 - \tfrac{\beta}{2}\,\mathbb{E}[\Delta_t] \\
&\ge \log 2 - \tfrac{3\beta}{2} L W_1(\pi_t, \pi_{t-1}) \\
&\ge \log 2 - \tfrac{3\beta}{2} L \sqrt{2C_0\, D_{\mathrm{KL}}(\pi_t \| \pi_{t-1})} \\
\Rightarrow \log 2 - \mathcal{L}_{\mathrm{DPO}}(\theta; x) &\le \tfrac{3\beta L}{2} \sqrt{2C_0 D_{\mathrm{KL}}(\pi_t \| \pi_{t-1})} \\
\Rightarrow D_{\mathrm{KL}}(\pi_t \| \pi_{t-1}) &\ge \frac{(\log 2 - \mathcal{L}_{\mathrm{DPO}}(\pi_t, \pi_{t-1}))^2}{\frac{1}{2}\beta^2 3^2 L^2 2C_0} \\
D_{\mathrm{KL}}(\pi_t \| \pi_{t-1}) &\ge \frac{(\log 2 - \mathcal{L}_{\mathrm{DPO}}(\pi_t, \pi_{t-1}))^2}{\beta^2 3^2 L^2 C_0}
\end{aligned} \tag{25}$$

Then, assume that there exists a constant $M \ge 1$ such that for all (x, y),

$$\frac{1}{M} \le \frac{\pi_{t-1}(y|x)}{\pi_t(y|x)} \le M. \tag{26}$$

This ensures that the predicted distributions of the LMM model at the previous learning step π_{t-1} and the current learning step π_t are mutually absolutely continuous and that their density ratio is uniformly bounded. In other words, it prevents the predictions of the LMM from collapsing across consecutive learning steps, ensuring a stable and smooth evolution of the output distribution during the continual learning procedure. Let $h(x, y) = \frac{\pi_{t-1}(y|x)}{\pi_t(y|x)}$ denote the likelihood ratio. The forward and reverse KL divergence can be rewritten as follows

$$D_{\mathrm{KL}}(\pi_{t-1} \| \pi_t) = \mathbb{E}_{(x,y)\sim\pi_t}\big[h(x, y) \log h(x, y)\big], \tag{27}$$

$$D_{\mathrm{KL}}(\pi_t \| \pi_{t-1}) = \mathbb{E}_{(x,y)\sim\pi_t}\big[-\log h(x, y)\big], \tag{28}$$

where $h(x, y) = \frac{\pi_{t-1}(y|x)}{\pi_t(y|x)}$.

Lower bound of $D_{\mathrm{KL}}(\pi_{t-1} \| \pi_t)$. The function $f(u) = u \log u$ is convex with $f''(u) = 1/u$. On the interval $[1/M, M]$, the smallest curvature is $1/M$. By the second-order convexity bound around $u = 1$,

$$u \log u \;\ge\; (u - 1) + \frac{1}{2M}(u - 1)^2. \tag{29}$$

Since $\mathbb{E}_{x,y\sim\pi_t(y|x)}[h(x, y) - 1] = 0$, taking the expectation under π_t will result in

$$D_{\mathrm{KL}}(\pi_{t-1} \| \pi_t) \;\ge\; \frac{1}{2M}\,\mathbb{E}_{\pi_t}\big[(h(x, y) - 1)^2\big]. \tag{1}$$

Upper bound of $D_{\mathrm{KL}}(\pi_t \| \pi_{t-1})$. Similarly, with $g(u) = -\log u$, we have $g''(u) = 1/u^2$, and on $[1/M, M]$, the largest curvature is M^2. Hence,

$$-\log u \;\le\; (1 - u) + \frac{M^2}{2}(u - 1)^2. \tag{30}$$

Then, taking expectation under π_t will result in

$$D_{\mathrm{KL}}(\pi_t \| \pi_{t-1}) \;\le\; \frac{M^2}{2}\,\mathbb{E}_{\pi_t}\big[(h(x, y) - 1)^2\big]. \tag{31}$$

From Eqn. (1) and and Eqn. (30), we can obtain

$$D_{\mathrm{KL}}(\pi_{t-1} \| \pi_t) \ge \frac{1}{M^3}\, D_{\mathrm{KL}}(\pi_t \| \pi_{t-1}). \tag{32}$$

Then, let us define $c = M^3$. Eqn. (25) can be further derived as follows:

$$\begin{aligned}
D_{\mathrm{KL}}(\pi_{t-1} \| \pi_t) &\ge \frac{(\log 2 - \mathcal{L}_{\mathrm{DPO}}(\theta; x))^2}{c\beta^2 3^2 L^2 C_0} \\
&\ge \frac{1}{C_{\mathrm{lower}}} (\log 2 - \mathcal{L}_{\mathrm{DPO}}(\theta; x))^2
\end{aligned} \tag{33}$$

where $C_{\mathrm{lower}} = c\beta^2 3^2 L^2 C_0$.

A.2. Proof of Lemma 2

Remark 1. Mixture Sampling and Monotone Labeling. For each prompt x, the output candidates are sampled from

$$Q_x \;=\; \alpha\, \pi_{t-1}(\cdot \mid x) \;+\; (1 - \alpha)\, \pi_t(\cdot \mid x) \quad \text{with } \alpha \in (0, 1], \tag{34}$$

and the selection kernel (human or reward model) chooses the preferred/dispreferred outputs (y^+, y^-) *monotonically* with the underlying reward, inducing pair marginals P_x^+, P_x^- that do not expand total variation beyond what is present in Q_x. Formally, we have

$$\mathrm{TV}\big(\pi_{t-1}(\cdot \mid x),\, \pi_t(\cdot \mid x)\big) \;\le\; \frac{1}{\alpha}\, \mathrm{TV}\big(P^+, P^-\big). \tag{35}$$

Remark 1 is both natural and theoretically justified in the context of continual learning via DPO. The candidate responses of DPO are typically drawn from a mixture of the previous and current policies, Q_x, to ensure balanced exposure to both past and newly adapted behaviors. The monotone labeling condition further indicates that the preference signal, whether derived from humans or a reward model, preserves the true reward ordering of outputs. Then, the total-variation inequality then follows from the data-processing principle, i.e., applying a monotone labeling kernel cannot increase statistical divergence between distributions. Intuitively, the preference selection process can only reveal discrepancies already present in the mixture Q_x, not amplify them. Consequently, Remark 1 enforces a bounded relationship between the divergence of the induced pairwise marginals (P_x^+, P_x^-) and the divergence between the underlying policies (π_{t-1}, π_t). In addition, this guarantees that updates to π_t remain geometrically close to π_{t-1}, providing a stability-adaptability balance in the continual learning setting, i.e., the model can adapt to new data or tasks while preventing catastrophic forgetting.

Remark 2. Sign Consistency. The predictor π_t is *Bayes-consistent in sign* on the support of $M_x := \frac{1}{2}(P_x^+ + P_x^-)$:

$$\mathrm{sgn}\big(q_\theta(z) - \tfrac{1}{2}\big) \;=\; \mathrm{sgn}\big(\eta(z) - \tfrac{1}{2}\big) \quad \text{for } M_x\text{-a.e. } z,$$

where $\eta(z) = \frac{dP_x^+}{d(P_x^+ + P_x^-)}(z)$ and $q_\theta(z) = \sigma(\beta s_\theta(z))$ with $\sigma(u) = \frac{1}{1+e^{-u}}$. This remark is standard in excess-risk calibration and holds whenever the logistic excess risk is sufficiently small to ensure boundary consistency.

Lemma 7 ***Logistic Calibration for Pairs***. *Given P^+ and P^-, the total variation of $TV(P^+, P^-)$ will be bounded by the DPO loss:*

$$\mathrm{TV}(P^+, P^-) \le 2\sqrt{2}\sqrt{\mathcal{L}_{\mathrm{DPO}}(\pi_t, \pi_{t-1}) \; - \; \mathcal{L}^\star_{\mathrm{DPO}}(\pi_t, \pi_{t-1})} \tag{36}$$

where $\mathcal{L}^\star_{\mathrm{DPO}}(\pi_t, \pi_{t-1})$ is the Bayes-optimal logistic pairwise loss.

Proof. Let us abbreviate $P^+ = P_x^+$, $P^- = P_x^-$, and $M = \frac{1}{2}(P^+ + P^-)$. The DPO loss and its Bayes-optimal counterpart can be written as

$$\begin{aligned} \mathcal{L}_{\mathrm{DPO}}(\pi_t, \pi_{t-1}) &= \mathbb{E}_{Z\sim M}\big[\mathrm{CE}(\eta(Z), q_\theta(Z))\big] \\ \mathcal{L}^\star_{\mathrm{DPO}}(\pi_t, \pi_{t-1}) &= \mathbb{E}_{Z\sim M}\big[\mathrm{CE}(\eta(Z), \eta(Z))\big] \end{aligned} \tag{37}$$

where $\mathrm{CE}(\cdot,\cdot)$ is the binary cross-entropy function, $\eta(\cdot)$ represents the true (Bayes-optimal) preference probability between positive and negative outcomes, $q_\theta(Z)$ is model-predicted probability obtained from the logit margin of the LMM model. Then, the excess DPO risk is formed as:

$$\begin{aligned} \mathfrak{R}_{\pi_t,\pi_{t-1}}(x) &= \mathcal{L}_{\mathrm{DPO}}(\pi_t, \pi_{t-1}) - \mathcal{L}^\star_{\mathrm{DPO}}(\pi_t, \pi_{t-1}) \\ &= \mathbb{E}_{Z\sim M}\Big[\mathrm{KL}\big(\mathrm{Bern}(\eta(Z)) \,\big\|\, \mathrm{Bern}(q_\theta(Z))\big)\Big]. \end{aligned} \tag{38}$$

where Bern is the Bernoulli distribution.

Bernoulli Pinsker Inequality. For any z, Pinsker's inequality for Bernoulli distributions gives

$$\mathrm{KL}\big(\mathrm{Bern}(\eta(z)) \,\|\, \mathrm{Bern}(q_\theta(z))\big) \;\ge\; 2\,(\eta(z) - q_\theta(z))^2. \tag{39}$$

Hence,

$$\mathbb{E}_M[(\eta - q_\theta)^2] \;\le\; \frac{1}{2}\,\mathfrak{R}_{\pi_t,\pi_{t-1}}(x). \tag{40}$$

Sign Consistency Implies Margin Control. Under Remark 2, since η and q_θ lie on the same side of $\frac{1}{2}$ for almost every z:

$$\begin{aligned} |\eta(z) - \tfrac{1}{2}| \le |\eta(z) - q_\theta(z)| + |q_\theta(z) - \tfrac{1}{2}| \le 2\,|\eta(z) - q_\theta(z)| \\ \Rightarrow |2\eta(z) - 1| \le 4\,|\eta(z) - q_\theta(z)| \end{aligned} \tag{41}$$

By definition of total variation, we have

$$\mathrm{TV}(P^+, P^-) = \mathbb{E}_{Z\sim M}\big[|2\eta(Z) - 1|\big]. \tag{42}$$

Then, applying Eqn. (41) and Cauchy–Schwarz, we will receive

$$\mathrm{TV}(P^+, P^-) \;\le\; 4\,\mathbb{E}_M|\eta - q_\theta| \;\le\; 4\,\sqrt{\mathbb{E}_M(\eta - q_\theta)^2}. \tag{43}$$

Then, substitute Eqn.(40) will result in

$$\begin{aligned} \mathrm{TV}(P^+, P^-) &\le 4\,\sqrt{\tfrac{1}{2}\,\mathfrak{R}_{\pi_t,\pi_{t-1}}(x)} \\ &= 2\sqrt{2}\,\sqrt{\mathcal{L}_{\mathrm{DPO}}(\pi_t, \pi_{t-1}) - \mathcal{L}^\star_{\mathrm{DPO}}(\pi_t, \pi_{t-1})}\,. \end{aligned} \tag{44}$$

Proof of Lemma 2. By Pinsker's inequality, we have

$$\begin{aligned} \mathrm{TV}\big(\pi_{t-1}, \pi_t\big) \;&\le\; \sqrt{\tfrac{1}{2}\,D_{\mathrm{KL}}\big(\pi_{t-1} \,\|\, \pi_t\big)}, \\ \Rightarrow D_{\mathrm{KL}}\big(\pi_{t-1} \,\|\, \pi_t\big) \;&\le\; 2\,\mathrm{TV}\big(\pi_{t-1}, \pi_t\big)^2. \end{aligned} \tag{45}$$

Then, substituting Remark 1 and Lemma 7 into Pinsker's relation will result in

$$\begin{aligned} D_{\mathrm{KL}}\big(\pi_{t-1} \,\|\, \pi_t\big) &\le 2\left(\frac{2\sqrt{2}}{\alpha}\sqrt{\mathcal{L}_{\mathrm{DPO}}(\pi_t, \pi_{t-1}) - \mathcal{L}^\star_{\mathrm{DPO}}(\pi_t, \pi_{t-1})}\right)^2 \\ &\le \frac{16}{\alpha^2}\left(\mathcal{L}_{\mathrm{DPO}}(\pi_t, \pi_{t-1}) - \mathcal{L}^\star_{\mathrm{DPO}}(\pi_t, \pi_{t-1})\right) \\ &\le \frac{16}{\alpha^2}\mathcal{L}_{\mathrm{DPO}}(\pi_t, \pi_{t-1}) \\ &= C_{\mathrm{upper}}\mathcal{L}_{\mathrm{DPO}}(\pi_t, \pi_{t-1}) \end{aligned} \tag{46}$$

where $C_{\mathrm{upper}} = \frac{16}{\alpha^2}$.

A.3. Proof of Lemma 3

Proof. Since $\log p \le 0$, one has $0 \;\le\; \alpha_\gamma(p) \;\le\; (1-p)^\gamma$, and $(1-p)^\gamma \to 0$ exponentially as $\gamma \to \infty$. Then, for any fixed $p \in (0,1)$, we have $\lim_{\gamma\to\infty} \alpha_\gamma(p) = 0$. As a result, for each group k, if $p(z) \in (0,1)$ a.s. and $\mathbb{E}[\|(p_\theta - 1)\nabla s_\theta\| \mid G_k] < \infty$, then $\lim_{\gamma\to\infty} w_k^\gamma(\theta) = 0$. By definition, we have

$$\|B_\gamma(\theta)\| \;\le\; \sum_{k=1}^{K} |q_k - q_k'|\,|w_k^\gamma(\theta)|\,\|m_k(\theta)\|. \tag{47}$$

Since $w_k^\gamma(\theta) \to 0$ for each k as $\gamma \to \infty$, the sum tends to 0.

B. DPO Data

B.1. Data Description

We release the dataset, annotations, and data preparation scripts on the project website https://uark-cviu.github.io/projects/Fai-DPO/. The repository provides the finalized annotations, data structure, and detailed instructions for reproducing the dataset used in our experiments. Due to copyright restrictions, we do not directly redistribute the raw images, as the original image copyrights remain with their respective data providers. Instead, we provide guidelines that allow users to obtain the images from the official sources and reconstruct the complete dataset in a reproducible manner.

This approach ensures that the dataset can be fully reproduced while respecting the licensing terms of the original image collections. The released resources include standardized annotations, metadata, and preprocessing protocols necessary to replicate the experimental setup reported in this paper.

B.2. Copyright and Usage Notice

All images used in this work remain the intellectual property of their original creators and dataset providers. We do not claim ownership of any raw images. The project website provides only annotations, data splits, and utilities for downloading the images from their respective official sources.

Users are responsible for ensuring that their use of the images complies with the licensing agreements and usage terms specified by the original datasets. By following the provided instructions, researchers can reconstruct the full dataset while maintaining compliance with copyright regulations and promoting transparent and reproducible research.

C. Data Quality and Noise Robustness

C.1. Data Quality

Similar to prior work that leverages LLMs for large-scale instruction or preference data generation [65, 67, 120], LLMs are only used to generate candidate negative preference pairs, which are then manually verified and filtered before training. This verification process was conducted by five specialists over approximately one month.

C.2. Noise Robustness

Large-scale preference annotations in complex settings inevitably introduce noise and potential bias. However, this noise in DPO-style training is typically considered imperfect or ambiguous negative preference pairs, especially when certain negatives are over-represented. From this perspective, our ϕ-DPO is designed to limit the influence of any single subset of preference annotations during optimization (Lemma 3), so that ambiguous or over-represented negative pairs do not negatively affect gradient updates. To further examine this issue, we conduct additional experiments by injecting controlled noise into the preference data. As shown in Table 8, ϕ-DPO maintains stable performance as noise increases, providing empirical evidence that the training process remains effective and balanced even in real-world settings with imperfect preference annotations.

Table 8. Noise Robustness Examination on the Preference Data

Noise Level	RS	Med	AD	Sci	Fin	MFT↑	MFN↑	MAA↑	BWT↑
0%	**85.68**	**69.74**	**57.73**	**61.55**	**95.28**	**74.29**	**74.00**	**75.68**	**-0.37**
5%	84.15	68.73	57.26	60.94	95.27	74.09	73.27	75.18	-1.02
15%	83.06	67.51	56.16	60.40	94.47	73.39	72.32	74.42	-1.33

D. Efficiency Analysis and Scalability

ϕ-DPO is designed to be parameter-efficient and memory-aware by training via LoRA, without updating full model parameters. As in Tab. 9, for a 7B model, ϕ-DPO requires only $\sim$0.65GB additional memory for weights of LoRA adapters and incurs a modest runtime overhead. Moreover, ϕ-DPO does not rely on replay buffers or task-specific gradient storage; preference updates are computed on-the-fly, keeping memory usage largely independent of the number of tasks. We also demonstrated scalability with different model sizes (Tab. 7), i.e., InternVL-7B and LLaVA-13B, showing consistent performance with manageable computation.

Table 9. Efficiency Analysis of Parameter-efficient Training

Method	LoRA+FT [36]	MoELoRA [13]	ϕ-DPO
Training Time	8.11s/iter	8.35s/iter	9.77s/iter
Adapter Memory	$\sim$0.65G	$\sim$0.65G	$\sim$0.65G

E. Ablation on Weighted Loss Coefficients

We adopt a fixed weight of losses in Eqn. (17) to analyze the stability–plasticity trade-off. The SFT favors adaptation to new tasks, while ϕ-DPO stabilizes updates and mitigates forgetting. As in Tab. 10, higher SFT weight improves adaptation (higher MFT) but increases forgetting (lower BWT), whereas higher ϕ-DPO weight improves retention at the cost of adaptability.

Table 10. Ablation on λ_{SFT} and $\lambda_{\phi\text{-DPO}}$

λ_{SFT}	$\lambda_{\phi\text{-DPO}}$	RS	Med	AD	Sci	Fin	MFT↑	MFN↑	MAA↑	BWT↑
2.0	1.0	83.62	68.56	57.18	61.32	**96.43**	**75.38**	73.42	**76.09**	-2.45
1.0	1.0	**85.68**	**69.74**	**57.73**	**61.55**	95.28	74.29	**74.00**	75.68	-0.37
1.0	2.0	83.71	67.54	55.67	59.23	93.16	72.13	71.86	73.61	**-0.33**

References

[1] Manoj Acharya, Kushal Kafle, and Christopher Kanan. Tallyqa: Answering complex counting questions. In *Proceedings of the AAAI conference on artificial intelligence*, pages 8076–8084, 2019. 6

[2] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv*, 2023. 3

[3] Gustavo Aguilar, Yuan Ling, Yu Zhang, Benjamin Yao, Xing Fan, and Chenlei Guo. Knowledge distillation from internal representations. In *Proceedings of the AAAI conference on artificial intelligence*, pages 7350–7357, 2020. 2

[4] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, et al. Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35: 23716–23736, 2022. 3

[5] Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. Qwen technical report. *arXiv*, 2023. 1, 3

[6] Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. Qwen-vl: A versatile vision-language model for understanding, localization, text reading, and beyond. *arXiv*, 2023. 3

[7] 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

[8] Meng Cao, Yuyang Liu, Yingfei Liu, Tiancai Wang, Jiahua Dong, Henghui Ding, Xiangyu Zhang, Ian Reid, and Xiaodan Liang. Continual llava: Continual instruction tuning in large vision-language models. *arXiv*, 2024. 3

[9] Patrick Cattiaux and Arnaud Guillin. A criterion for talagrand's quadratic transportation cost inequality. *arXiv*, 2003. 10

[10] Fabio Cermelli, Massimiliano Mancini, Samuel Rota Bulò, Elisa Ricci, and Barbara Caputo. Modeling the background for incremental learning in semantic segmentation. In *CVPR*, 2020. 2, 3, 4, 5

[11] Shuaichen Chang, David Palzer, Jialin Li, Eric Fosler-Lussier, and Ningchuan Xiao. Mapqa: A dataset for question answering on choropleth maps. *arXiv*, 2022. 6

[12] Yupeng Chang, Yi Chang, and Yuan Wu. Ba-lora: Bias-alleviating low-rank adaptation to mitigate catastrophic inheritance in large language models, 2025. 2

[13] Cheng Chen, Junchen Zhu, Xu Luo, Heng T Shen, Jingkuan Song, and Lianli Gao. Coin: A benchmark of continual

instruction tuning for multimodel large language models. *Advances in Neural Information Processing Systems*, 37: 57817–57840, 2024. 2, 3, 6, 7, 12

[14] Jiayi Chen and Aidong Zhang. FedMBridge: Bridgeable multimodal federated learning. In *Forty-first International Conference on Machine Learning*, 2024. 3

[15] Jinpeng Chen, Runmin Cong, Yuzhi Zhao, Hongzheng Yang, Guangneng Hu, Horace Ho Shing Ip, and Sam Kwong. Sefe: Superficial and essential forgetting eliminator for multimodal continual instruction tuning. *arXiv*, 2025. 3, 7

[16] Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, et al. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24185–24198, 2024. 8

[17] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality, 2023. 3, 6

[18] Andrea Cossu, Antonio Carta, Lucia Passaro, Vincenzo Lomonaco, Tinne Tuytelaars, and Davide Bacciu. Continual pre-training mitigates forgetting in language and vision. *Neural Networks*, 179:106492, 2024. 3

[19] Matt Deitke, Christopher Clark, Sangho Lee, Rohun Tripathi, Yue Yang, Jae Sung Park, Mohammadreza Salehi, Niklas Muennighoff, Kyle Lo, Luca Soldaini, et al. Molmo and pixmo: Open weights and open data for state-of-the-art vision-language models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 91–104, 2025. 1

[20] 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. 6

[21] Arthur Douillard, Yifu Chen, Arnaud Dapogny, and Matthieu Cord. Plop: Learning without forgetting for continual semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4040–4050, 2021. 2, 3, 4, 5

[22] 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

[23] David A Edwards. On the kantorovich–rubinstein theorem. *Expositiones Mathematicae*, 29(4):387–398, 2011. 9

[24] Deepanway Ghosal, Navonil Majumder, Ambuj Mehrish, and Soujanya Poria. Text-to-audio generation using instruction-tuned llm and latent diffusion model. *arXiv*, 2023. 3

[25] Yash Goyal, Tejas Khot, Douglas Summers-Stay, Dhruv Batra, and Devi Parikh. Making the v in vqa matter: Elevating the role of image understanding in visual question answering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6904–6913, 2017. 6

[26] Haiyang Guo, Fanhu Zeng, Ziwei Xiang, Fei Zhu, Da-Han Wang, Xu-Yao Zhang, and Cheng-Lin Liu. Hide-llava: Hierarchical decoupling for continual instruction tuning of multimodal large language model. *arXiv*, 2025. 2, 3, 6, 7

[27] Haiyang Guo, Fanhu Zeng, Fei Zhu, Wenzhuo Liu, Da-Han Wang, Jian Xu, Xu-Yao Zhang, and Cheng-Lin Liu. Federated continual instruction tuning. *arXiv*, 2025. 7

[28] Haiyang Guo, Fanhu Zeng, Fei Zhu, Jiayi Wang, Xukai Wang, Jingang Zhou, Hongbo Zhao, Wenzhuo Liu, Shijie Ma, Xu-Yao Zhang, et al. A comprehensive survey on continual learning in generative models. *arXiv*, 2025. 3

[29] Ziyu Guo, Ray Zhang, Hao Chen, Jialin Gao, Dongzhi Jiang, Jiaze Wang, and Pheng-Ann Heng. Sciverse: Unveiling the knowledge comprehension and visual reasoning of lmms on multi-modal scientific problems. *arXiv*, 2025. 6

[30] Danna Gurari, Qing Li, Abigale J Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P Bigham. Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3608–3617, 2018. 6

[31] Jiayi Han, Liang Du, Hongwei Du, Xiangguo Zhou, Yiwen Wu, Weibo Zheng, and Donghong Han. Slim: Let llm learn more and forget less with soft lora and identity mixture. *arXiv*, 2024. 2

[32] Ryo Hase, Md Rafi Ur Rashid, Ashley Lewis, Jing Liu, Toshiaki Koike-Akino, Kieran Parsons, and Ye Wang. Smoothed embeddings for robust language models. *arXiv*, 2025. 9

[33] Xuehai He, Yichen Zhang, Luntian Mou, Eric Xing, and Pengtao Xie. Pathvqa: 30000+ questions for medical visual question answering. *arXiv*, 2020. 6

[34] Calypso Herrera, Florian Krach, and Josef Teichmann. Local lipschitz bounds of deep neural networks. *arXiv*, 2020. 9

[35] Yu-Chung Hsiao, Fedir Zubach, Gilles Baechler, Victor Carbune, Jason Lin, Maria Wang, Srinivas Sunkara, Yun Zhu, and Jindong Chen. Screenqa: Large-scale question-answer pairs over mobile app screenshots. *arXiv*, 2022. 6

[36] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022. 2, 7, 12

[37] Tianyu Huai, Jie Zhou, Xingjiao Wu, Qin Chen, Qingchun Bai, Ze Zhou, and Liang He. Cl-moe: Enhancing multimodal large language model with dual momentum mixture-of-experts for continual visual question answering. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 19608–19617, 2025. 7

[38] Drew A Hudson and Christopher D Manning. Gqa: A new dataset for real-world visual reasoning and compositional question answering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6700–6709, 2019. 6

[39] Atin Sakkeer Hussain, Shansong Liu, Chenshuo Sun, and Ying Shan. M2ugen: Multi-modal music understanding and generation with the power of large language models. *arXiv*, 2023. 3

[40] Goeric Huybrechts, Srikanth Ronanki, Sai Muralidhar Jayanthi, Jack Fitzgerald, and Srinivasan Veeravanallur. Document haystack: A long context multimodal image/document understanding vision llm benchmark. *arXiv*, 2025. 1

[41] Amit Kumar Jaiswal, Haiming Liu, and Ingo Frommholz. Multimodal rag enhanced visual description. *arXiv*, 2025. 1

[42] Joel Jang, Seonghyeon Ye, Changho Lee, Sohee Yang, Joongbo Shin, Janghoon Han, Gyeonghun Kim, and Minjoon Seo. Temporalwiki: A lifelong benchmark for training and evaluating ever-evolving language models. *arXiv*, 2022. 3

[43] Justin Johnson, Bharath Hariharan, Laurens Van Der Maaten, Li Fei-Fei, C Lawrence Zitnick, and Ross Girshick. Clevr: A diagnostic dataset for compositional language and elementary visual reasoning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2901–2910, 2017. 6

[44] Sahar Kazemzadeh, Vicente Ordonez, Mark Matten, and Tamara Berg. Referitgame: Referring to objects in photographs of natural scenes. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 787–798, 2014. 6

[45] Aniruddha Kembhavi, Mike Salvato, Eric Kolve, Minjoon Seo, Hannaneh Hajishirzi, and Ali Farhadi. A diagram is worth a dozen images. *arXiv*, 2016. 6

[46] Aniruddha Kembhavi, Minjoon Seo, Dustin Schwenk, Jonghyun Choi, Ali Farhadi, and Hannaneh Hajishirzi. Are you smarter than a sixth grader? textbook question answering for multimodal machine comprehension. In *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition*, pages 4999–5007, 2017. 6

[47] 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. 3, 7

[48] Frantzeska Lavda, Jason Ramapuram, Magda Gregorova, and Alexandros Kalousis. Continual classification learning using generative models. *arXiv*, 2018. 3

[49] Michel Ledoux, Ivan Nourdin, and Giovanni Peccati. Stein's method, logarithmic sobolev and transport inequalities. *Geometric and Functional Analysis*, 25(1):256–306, 2015. 9

[50] Bo Li, Hao Zhang, Kaichen Zhang, Dong Guo, Yuanhan Zhang, Renrui Zhang, Feng Li, Ziwei Liu, and Chunyuan Li. Llava-next: What else influences visual instruction tuning beyond data?, 2024. 3

[51] Bo Li, Kaichen Zhang, Hao Zhang, Dong Guo, Renrui Zhang, Feng Li, Yuanhan Zhang, Ziwei Liu, and Chunyuan Li. Llava-next: Stronger llms supercharge multimodal capabilities in the wild, 2024. 3

[52] Bo Li, Yuanhan Zhang, Dong Guo, Renrui Zhang, Feng Li, Hao Zhang, Kaichen Zhang, Yanwei Li, Ziwei Liu, and Chunyuan Li. Llava-onevision: Easy visual task transfer. *arXiv*, 2024. 3

[53] Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, and Jianfeng Gao. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *Advances in Neural Information Processing Systems*, 36, 2024. 3

[54] Feng Li, Renrui Zhang, Hao Zhang, Yuanhan Zhang, Bo Li, Wei Li, Zejun Ma, and Chunyuan Li. Llava-next: Tackling multi-image, video, and 3d in large multimodal models, 2024. 3

[55] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International Conference on Machine Learning*, pages 12888–12900. PMLR, 2022. 1

[56] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pages 19730–19742. PMLR, 2023. 1, 3

[57] KunChang Li, Yinan He, Yi Wang, Yizhuo Li, Wenhai Wang, Ping Luo, Yali Wang, Limin Wang, and Yu Qiao. Videochat: Chat-centric video understanding. *arXiv*, 2023. 3

[58] Yanwei Li, Chengyao Wang, and Jiaya Jia. Llama-vid: An image is worth 2 tokens in large language models. In *European Conference on Computer Vision*, pages 323–340. Springer, 2025. 3

[59] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 3, 7

[60] Zhang Li, Biao Yang, Qiang Liu, Zhiyin Ma, Shuo Zhang, Jingxu Yang, Yabo Sun, Yuliang Liu, and Xiang Bai. Monkey: Image resolution and text label are important things for large multi-modal models. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 26763–26773, 2024. 6

[61] Bin Lin, Yang Ye, Bin Zhu, Jiaxi Cui, Munan Ning, Peng Jin, and Li Yuan. Video-llava: Learning united visual representation by alignment before projection. *arXiv*, 2023. 3

[62] Jessy Lin, Luke Zettlemoyer, Gargi Ghosh, Wen-Tau Yih, Aram Markosyan, Vincent-Pierre Berges, and Barlas Oğuz. Continual learning via sparse memory finetuning. *arXiv*, 2025. 3

[63] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 5

[64] Dingning Liu, Xiaoshui Huang, Yuenan Hou, Zhihui Wang, Zhen fei Yin, Yongshun Gong, Peng Gao, and Wanli Ouyang. Uni3d-llm: Unifying point cloud perception, generation and editing with large language models. *arXiv*, 2024. 3

[65] Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26296–26306, 2024. 1, 3, 6, 12

[66] Haotian Liu, Chunyuan Li, Yuheng Li, Bo Li, Yuanhan Zhang, Sheng Shen, and Yong Jae Lee. Llava-next: Improved reasoning, ocr, and world knowledge, 2024. 3

[67] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36, 2024. 1, 3, 12

[68] Yuliang Liu, Zhang Li, Mingxin Huang, Biao Yang, Wenwen Yu, Chunyuan Li, Xu-Cheng Yin, Cheng-Lin Liu, Lianwen Jin, and Xiang Bai. Ocrbench: on the hidden mystery of ocr in large multimodal models. *Science China Information Sciences*, 67(12):220102, 2024. 6

[69] Sylvain Lobry, Diego Marcos, Jesse Murray, and Devis Tuia. Rsvqa: Visual question answering for remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12):8555–8566, 2020. 6

[70] Pan Lu, Swaroop Mishra, Tanglin Xia, Liang Qiu, Kai-Wei Chang, Song-Chun Zhu, Oyvind Tafjord, Peter Clark, and Ashwin Kalyan. Learn to explain: Multimodal reasoning via thought chains for science question answering. *Advances in Neural Information Processing Systems*, 35: 2507–2521, 2022. 6

[71] 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

[72] Junhua Mao, Jonathan Huang, Alexander Toshev, Oana Camburu, Alan L Yuille, and Kevin Murphy. Generation and comprehension of unambiguous object descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 11–20, 2016. 6

[73] Anand Mishra, Shashank Shekhar, Ajeet Kumar Singh, and Anirban Chakraborty. Ocr-vqa: Visual question answering by reading text in images. In *2019 international conference on document analysis and recognition (ICDAR)*, pages 947–952. IEEE, 2019. 6

[74] Hoang-Quan Nguyen, Thanh-Dat Truong, Xuan Bac Nguyen, Ashley Dowling, Xin Li, and Khoa Luu. Insect-foundation: A foundation model and large-scale 1m dataset for visual insect understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21945–21955, 2024. 1

[75] Trong-Thuan Nguyen, Pha Nguyen, Jackson Cothren, Alper Yilmaz, and Khoa Luu. Hyperglm: Hypergraph for video scene graph generation and anticipation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 29150–29160, 2025. 3

[76] Xuan-Bac Nguyen, Manuel Serna-Aguilera, Arabinda Kumar Choudhary, Pawan Sinha, Xin Li, and Khoa Luu. Cobra: A continual learning approach to vision-brain understanding: X. nguyen et al. *International Journal of Computer Vision*, 134(1):30, 2026. 3

[77] Yiqiao Qiu, Yixing Shen, Zhuohao Sun, Yanchong Zheng, Xiaobin Chang, Weishi Zheng, and Ruixuan Wang. Sats: Self-attention transfer for continual semantic segmentation. *Pattern Recognition*, 138:109383, 2023. 2

[78] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023. 4

[79] Anastasia Razdaibiedina, Yuning Mao, Rui Hou, Madian Khabsa, Mike Lewis, and Amjad Almahairi. Progressive prompts: Continual learning for language models. *arXiv*, 2023. 3

[80] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv*, 2017. 4

[81] Manuel Serna-Aguilera, Raegan Anderes, Page Dobbs, and Khoa Luu. Nico-rag: Multimodal hypergraph retrieval-augmented generation for understanding the nicotine public health crisis. *arXiv*, 2026. 1

[82] Wei Shen, Jiangbo Pei, Yi Peng, Xuchen Song, Yang Liu, Jian Peng, Haofeng Sun, Yunzhuo Hao, Peiyu Wang, Jianhao Zhang, and Yahui Zhou. Skywork-r1v3 technical report, 2025. 3

[83] Haizhou Shi, Zihao Xu, Hengyi Wang, Weiyi Qin, Wenyuan Wang, Yibin Wang, Zifeng Wang, Sayna Ebrahimi, and Hao Wang. Continual learning of large language models: A comprehensive survey. *ACM Computing Surveys*, 2024. 3

[84] Wenhao Shi, Zhiqiang Hu, Yi Bin, Junhua Liu, Yang Yang, See-Kiong Ng, Lidong Bing, and Roy Ka-Wei Lee. Math-llava: Bootstrapping mathematical reasoning for multimodal large language models. *arXiv*, 2024. 6

[85] Zenglin Shi, Pei Liu, Tong Su, Yunpeng Wu, Kuien Liu, Yu Song, and Meng Wang. Densely distilling cumulative knowledge for continual learning. *arXiv*, 2024. 2

[86] Marco Siino. Mcrock at semeval-2024 task 4: Mistral 7b for multilingual detection of persuasion techniques in memes. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 53–59, 2024. 3

[87] Chonghao Sima, Katrin Renz, Kashyap Chitta, Li Chen, Hanxue Zhang, Chengen Xie, Jens Beißwenger, Ping Luo, Andreas Geiger, and Hongyang Li. Drivelm: Driving with graph visual question answering. In *European conference on computer vision*, pages 256–274. Springer, 2024. 6

[88] Amanpreet Singh, Vivek Natarajan, Meet Shah, Yu Jiang, Xinlei Chen, Dhruv Batra, Devi Parikh, and Marcus Rohrbach. Towards vqa models that can read. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8317–8326, 2019. 6

[89] 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

[90] Alane Suhr and Yoav Artzi. Continual learning for instruction following from realtime feedback. *Advances in Neural Information Processing Systems*, 36:32340–32359, 2023. 3

[91] Ryota Tanaka, Taichi Iki, Taku Hasegawa, Kyosuke Nishida, Kuniko Saito, and Jun Suzuki. Vdocrag: Retrieval-augmented generation over visually-rich documents. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 24827–24837, 2025. 1

[92] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive representation distillation. *arXiv*, 2019. 2

[93] Huu-Thien Tran, Thanh-Dat Truong, and Khoa Luu. Bima: Bijective maximum likelihood learning approach to hallucination prediction and mitigation in large vision-language models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5302–5311, 2025. 3

[94] Thanh-Dat Truong, Chi Nhan Duong, Ngan Le, Son Lam Phung, Chase Rainwater, and Khoa Luu. Bimal: Bijective maximum likelihood approach to domain adaptation in semantic scene segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8548–8557, 2021. 3

[95] Thanh-Dat Truong, Ngan Le, Bhiksha Raj, Jackson Cothren, and Khoa Luu. Fredom: Fairness domain adaptation approach to semantic scene understanding. In *IEEE/CVF Computer Vision and Pattern Recognition (CVPR)*, 2023. 3

[96] Thanh-Dat Truong, Hoang-Quan Nguyen, Bhiksha Raj, and Khoa Luu. Fairness continual learning approach to semantic scene understanding in open-world environments. *Advances in Neural Information Processing Systems*, 36: 65456–65467, 2023. 2, 3

[97] Thanh-Dat Truong, Utsav Prabhu, Dongyi Wang, Bhiksha Raj, Susan Gauch, Jeyamkondan Subbiah, and Khoa Luu. Eagle: Efficient adaptive geometry-based learning in cross-view understanding. *Advances in Neural Information Processing Systems*, 37:137309–137333, 2024. 3

[98] Thanh-Dat Truong, Christophe Bobda, Nitin Agarwal, and Khoa Luu. Mango: Multimodal attention-based normalizing flow approach to fusion learning. *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 3

[99] Thanh-Dat Truong, Hoang-Quan Nguyen, Xuan-Bac Nguyen, Ashley Dowling, Xin Li, and Khoa Luu. Insect-foundation: A foundation model and large multimodal dataset for vision-language insect understanding. *International Journal of Computer Vision*, pages 1–26, 2025. 3

[100] Thanh-Dat Truong, Utsav Prabhu, Bhiksha Raj, Jackson Cothren, and Khoa Luu. Falcon: Fairness learning via contrastive attention approach to continual semantic scene understanding. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 15065–15075, 2025. 2, 3

[101] Thanh-Dat Truong, Huu-Thien Tran, Tran Thai Son, Bhiksha Raj, and Khoa Luu. Directed-tokens: A robust multi-modality alignment approach to large language-vision models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 3

[102] Bryan Wang, Gang Li, Xin Zhou, Zhourong Chen, Tovi Grossman, and Yang Li. Screen2words: Automatic mobile ui summarization with multimodal learning. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pages 498–510, 2021. 6

[103] Shuchan Wang, Photios A Stavrou, and Mikael Skoglund. Generalizations of talagrand inequality for sinkhorn distance using entropy power inequality. *Entropy*, 24(2):306, 2022. 9

[104] Weiyun Wang, Zhangwei Gao, Lixin Gu, Hengjun Pu, Long Cui, Xingguang Wei, Zhaoyang Liu, Linglin Jing, Shenglong Ye, Jie Shao, et al. Internvl3.5: Advancing open-source multimodal models in versatility, reasoning, and efficiency, 2025. 3

[105] Xiao Wang, Tianze Chen, Qiming Ge, Han Xia, Rong Bao, Rui Zheng, Qi Zhang, Tao Gui, and Xuanjing Huang. Orthogonal subspace learning for language model continual learning. *arXiv*, 2023. 3, 7

[106] Yidong Wang, Zhuohao Yu, Jindong Wang, Qiang Heng, Hao Chen, Wei Ye, Rui Xie, Xing Xie, and Shikun Zhang. Exploring vision-language models for imbalanced learning. *International Journal of Computer Vision*, 132(1):224–237, 2024. 2

[107] 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, 7

[108] Yuetian Weng, Mingfei Han, Haoyu He, Xiaojun Chang, and Bohan Zhuang. Longvlm: Efficient long video understanding via large language models. *arXiv*, 2024. 3

[109] Runsen Xu, Xiaolong Wang, Tai Wang, Yilun Chen, Jiangmiao Pang, and Dahua Lin. Pointllm: Empowering large language models to understand point clouds. In *ECCV*, 2024. 3

[110] Senqiao Yang, Jiaming Liu, Ray Zhang, Mingjie Pan, Zoey Guo, Xiaoqi Li, Zehui Chen, Peng Gao, Yandong Guo, and Shanghang Zhang. Lidar-llm: Exploring the potential of large language models for 3d lidar understanding. *arXiv*, 2023. 3

[111] Muchao Ye, Ziyi Yin, Tianrong Zhang, Tianyu Du, Jinghui Chen, Ting Wang, and Fenglong Ma. Unit: a unified look at certified robust training against text adversarial perturbation. *Advances in Neural Information Processing Systems*, 36:22351–22368, 2023. 9

[112] Wenpeng Yin, Jia Li, and Caiming Xiong. Contintin: Continual learning from task instructions. *arXiv*, 2022. 3

[113] Kaining Ying, Fanqing Meng, Jin Wang, Zhiqian Li, Han Lin, Yue Yang, Hao Zhang, Wenbo Zhang, Yuqi Lin, Shuo Liu, et al. Mmt-bench: A comprehensive multimodal benchmark for evaluating large vision-language models towards multitask agi. *arXiv*, 2024. 6

[114] Shi Yu, Chaoyue Tang, Bokai Xu, Junbo Cui, Junhao Ran, Yukun Yan, Zhenghao Liu, Shuo Wang, Xu Han, Zhiyuan Liu, et al. Visrag: Vision-based retrieval-augmented generation on multi-modality documents. *arXiv*, 2024. 1

[115] Daoguang Zan, Bei Chen, Dejian Yang, Zeqi Lin, Minsu Kim, Bei Guan, Yongji Wang, Weizhu Chen, and Jian-Guang Lou. Cert: continual pre-training on sketches for library-oriented code generation. *arXiv*, 2022. 3

[116] Fanhu Zeng, Fei Zhu, Haiyang Guo, Xu-Yao Zhang, and Cheng-Lin Liu. Modalprompt: Towards efficient multimodal continual instruction tuning with dual-modality guided prompt. *arXiv*, 2024. 3

[117] Han Zhang, Yu Lei, Lin Gui, Min Yang, Yulan He, Hui Wang, and Ruifeng Xu. Cppo: Continual learning for reinforcement learning with human feedback. In *The Twelfth International Conference on Learning Representations*, 2024. 3

[118] Rongzhi Zhang, Jiaming Shen, Tianqi Liu, Jialu Liu, Michael Bendersky, Marc Najork, and Chao Zhang. Do not blindly imitate the teacher: Using perturbed loss for knowledge distillation. *arXiv*, 2023. 2

[119] Renrui Zhang, Xinyu Wei, Dongzhi Jiang, Ziyu Guo, Shicheng Li, Yichi Zhang, Chengzhuo Tong, Jiaming Liu, Aojun Zhou, Bin Wei, et al. Mavis: Mathematical visual instruction tuning with an automatic data engine. *arXiv*, 2024. 6

[120] Ruohong Zhang, Liangke Gui, Zhiqing Sun, Yihao Feng, Keyang Xu, Yuanhan Zhang, Di Fu, Chunyuan Li, Alexander G Hauptmann, Yonatan Bisk, and Yiming Yang. Direct preference optimization of video large multimodal models from language model reward. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 694–717. Association for Computational Linguistics, 2025. 12

[121] Xin Zhang, Liang Bai, Xian Yang, and Jiye Liang. C-lora: Continual low-rank adaptation for pre-trained models. *arXiv*, 2025. 2

[122] Yuanhan Zhang, Bo Li, haotian Liu, Yong jae Lee, Liangke Gui, Di Fu, Jiashi Feng, Ziwei Liu, and Chunyuan Li. Llava-next: A strong zero-shot video understanding model, 2024. 3

[123] Hongbo Zhao, Fei Zhu, Haiyang Guo, Meng Wang, Rundong Wang, Gaofeng Meng, and Zhaoxiang Zhang. Mllm-cl: Continual learning for multimodal large language models. *arXiv*, 2025. 1, 2, 3, 6, 7

[124] Yue Zhao, Ishan Misra, Philipp Krähenbühl, and Rohit Girdhar. Learning video representations from large language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6586–6597, 2023. 3