Lifelong Knowledge Editing for Vision Language Models with Low-Rank Mixture-of-Experts

Supplementary Material

7. Details of Experimental Settings

7.1. Datasets:

E-VQA [22] is a dataset designed to facilitate the correction of error-prone samples in VQA-v2 [36] for VLLMs, containing 6,345 training samples and 2,093 testing samples. The VQA task involves providing a VLLM with an image and a related question, requiring it to analyze both the visual content and the question to produce an accurate textual response.

E-IC [22] is a dataset created for editing VLLMs to rectify errors in COCO Caption [61], consisting of 2,849 training samples and 1,000 testing samples. The IC task requires the model to generate an accurate textual description for a given image, demanding a comprehensive understanding and articulation of the visual content.

Each sample in these two datasets includes an edit sample, along with two additional samples each for modal and text generality, and two samples each for modal and text locality. Generality samples are created by rephrasing images and queries using Stable Diffusion [62] and ChatGLM [63], respectively. Meanwhile, locality samples are generated using unrelated images and queries from the OK-VQA [64] and NQ [42] datasets.

VLKEB [30] is an editor evaluation dataset constructed based on MMKG (Multi-Modal Knowledge Graph) [65]. For each sample, it also includes an editing sample, along with two samples for modal and text generality, and two samples for modal and text locality. It contains 5,000 training samples and 3,174 testing samples.

Unlike the aforementioned datasets, VLKEB uses CLIP [66] to retrieve semantically similar real-world images, which are subsequently verified manually. This process ensures higher image quality compared to datasets relying on image generation models, making VLKEB more representative of real-world scenarios.

7.2. VLLM Backbones:

Below, we introduce the VLLM backbones used in our experiments. For the specific versions utilized, please refer to the corresponding footnotes.

LLaVA[2] [28] is a Vision-Language Large Model (VLLM) that bridges the gap between visual and language understanding by leveraging GPT-4 [67] to create an instruction-tuning dataset for vision-language pre-training. It achieves the transformation of visual features into the textual representation space through a simple yet effective two-layer MLP placed between the visual encoder and the language model LLaMA [1]. This design allows LLaVA to excel at tasks such as visual question answering and instruction following with high efficiency and precision.

BLIP2 [4] introduces a visual query transformer, called Q-Former, which is trained through a two-stage pre-training process to compress visual features into fixed-length representations within the language space. This design bridges the representational gap between the frozen visual encoder [66] and the frozen language model. BLIP2 comes in several variants, and in this paper, we follow [22] to experiment with BLIP2-OPT[3]. While LLaVa utilizes the full spectrum of visual features and preserves more fine-grained details, BLIP2 significantly reduces the length of visual representations, thereby improving the inference efficiency.

MiniGPT-4[4] [29], developed based on BLIP2, keeps the Q-Former and the language model Vicuna [38] frozen, training only a linear layer positioned after the Q-Former to efficiently align visual features with Vicuna.

7.3. Baseline Editors:

FT (**F**ine-**T**uning) comprises two variants: FT-L and FT-M. According to [22], FT-L involves fine-tuning the final layer of the language transformer within the VLLM, whereas FT-M focuses on fine-tuning the visual encoder module of the VLLM for a given edit sample.

MEND (**M**odel **E**ditor **N**etworks with **D**ecomposition) [42] trains a set of small MLP hyper-networks to achieve efficient model editing. These hyper-networks generate the FFN matrix parameter offsets by inputting the decomposed back-propagation gradients of these matrices on edit samples. Through editing-specific training, these hyper-networks can produce matrix offsets that enable LLMs to satisfy editing requirements.

SERAC (**S**emi-parametric **E**diting with a **R**etrieval-**A**ugmented **C**ounterfactual) [47] is a memory-based method. It trains a scope classifier and a small counterfactual model while storing edit samples in memory. The classifier determines whether subsequent inputs are related to the edit samples. If they are related, the inputs are sent to the counterfactual model for response modification; otherwise, they are passed to the original model to gener-

[2] https://huggingface.co/liuhaotian/llava-v1.5-7b

[3] https://huggingface.co/Salesforce/blip2-opt-2.7b

[4] https://huggingface.co/Vision-CAIR/MiniGPT-4

Editors	Backbones	Edit Iterations	Optimizer	Learning Rate	Edit Layers
FT-L	LLaVA-V1.5 (7B)	25	AdamW	1e-3	The last layer of the language transformer
	BLIP2-OPT (2.7B)	25	AdamW	1e-3	The last layer of the language transformer
	MiniGPT-4 (7B)	25	AdamW	1e-3	The last layer of the language transformer
FT-M	LLaVA-V1.5 (7B)	25	AdamW	1e-3	The multi modal projector
	BLIP2-OPT (2.7B)	25	AdamW	1e-3	Qformer
	MiniGPT-4 (7B)	25	AdamW	1e-3	Qformer
TP	LLaVA-V1.5 (7B)	25	Adam	1e-2	The last layer of the language transformer
	BLIP2-OPT (2.7B)	25	Adam	1e-2	The last layer of the language transformer
	MiniGPT-4 (7B)	25	Adam	1e-2	The last layer of the language transformer
LEMoE	LLaVA-V1.5 (7B)	100	AdamW	2e-4	The 18-th layer of the language transformer
	BLIP2-OPT (2.7B)	100	AdamW	2e-4	The 18-th layer of the language transformer
	MiniGPT-4 (7B)	100	AdamW	2e-4	The 18-th layer of the language transformer

Table 3. Editing details of direct editing editors.

Editors	Backbones	Optimizer	Learning Rate	Edit/Modified Layers
MEND	LLaVA-V1.5 (7B)	Adam	1e-4	layer 29,30,31 of the language transformer
	BLIP2-OPT (2.7B)	Adam	1e-4	layer 29,30,31 of the language transformer
	MiniGPT-4 (7B)	Adam	1e-4	layer 29,30,31 of the language transformer
SERAC	LLaVA-V1.5 (7B)	Adam	1e-4	-
	BLIP2-OPT (2.7B)	Adam	1e-4	-
	MiniGPT-4 (7B)	Adam	1e-4	-
LTE	LLaVA-V1.5 (7B)	Adam	5e-6	The whole language transformer
	BLIP2-OPT (2.7B)	Adam	5e-6	The whole language transformer
	MiniGPT-4 (7B)	Adam	5e-6	The whole language transformer
RECIPE	LLaVA-V1.5 (7B)	Adam	1e-5	-
	BLIP2-OPT (2.7B)	Adam	1e-5	-
	MiniGPT-4 (7B)	Adam	1e-5	-

Table 4. Editing details of training-based editing methods.

ate responses. In our experiments, we follow the setup of MMEdit [22], where BERT[5] [68] and OPT-125M[6] [69] are separately trained as the scope classifier and the counterfactual model.

TP (**T**ransformer **P**atcher) [45] assumes that each neuron in the LLM can carry a piece of knowledge. Therefore, for every new piece of edit sample, TP inserts and trains a new neuron in the final layer of the FFN within the LLM to accommodate the edit sample.

LTE (**L**earning **T**o **E**dit) [19] performs supervised fine-tuning of LLMs on edit samples, enabling the model to adapt its response when an edit sample is provided as a contextual prefix. To address sequential editing scenarios, it stores edit samples in a retrieval database and uses a pre-existing text retriever [70] to retrieve edit samples relevant to the input.

RECIPE (**RE**trieval-augmented **C**ont**I**nue **P**rompt l**E**arning) [20] aims to explore short editing prefixes to achieve lifelong editing for LLMs. By training an editing prompt generator, edit samples are converted into short continuous prompts, which serve as prefix contexts for LLMs to adapt their responses. Additionally, RECIPE trains an editing retriever to retrieve relevant edited samples based on the input.

LEMoE (**L**ifelong **E**diting with **M**ixture **o**f **E**xperts) [46] achieves lifelong editing for LLMs by inserting and maintaining a MoE in the final layer of the LLM. In lifelong editing, whenever a batch of editing requirements arises, an editing expert is initialized and trained to adapt the model's responses for that specific batch of edit samples. During the training of a new editing expert, previously trained experts also participate in the back-propagation process, but their parameters remain frozen. Additionally, a key vector is trained alongside the expert to route the input to the relevant experts.

7.4. Model Settings and Training Details

LiveEdit: LiveEdit adopts the same hyperparameters for all backbones. After hyperparameter tuning and balancing computational resources with editing performance, we set the module dimension $d_m = 1024$, the rank of the editing expert $r = 4$, and the feature extraction control parameter $k = 4$. For the index of editing layers l_e, based on the attribution analysis from [24] and our experiments in Figure 5, we set $l_e = 21$. The learning rate η is set to 1e-4, the training batch size $B = 8$, and the maximum number of training iterations is 200K. If the loss stops decreasing, we termi-

[5] https://huggingface.co/google-bert/bert-base-cased

[6] https://huggingface.co/facebook/opt-125m

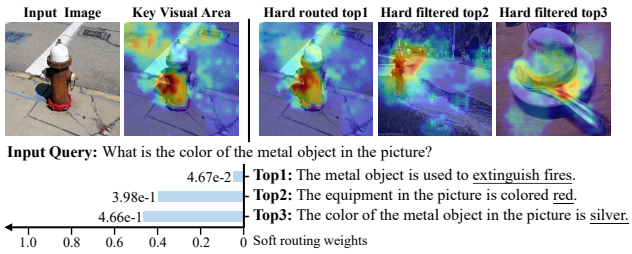

Figure 7. Instance analysis with difficult samples.

nate the training process early. A checkpoint is saved every 500 iterations, and the one with the lowest loss is selected for lifelong editing evaluation. The training process takes approximately 2 days on a single NVIDIA A800 GPU.

Other Baselines: The hyperparameters of TP [45], LTE [19], and LEMoE [46] are adopted from their respective papers. They are slightly adjusted to accommodate VLLM editing, such as mapping visual representations to the language representation space via a trainable MLP module. For the other baselines, we follow [22]. For the experimental details of direct editing methods, please refer to Table 3. For training-based editing methods, we also set the maximum number of iterations to 200K. Training details are provided in Table 4.

8. More Experiments

8.1. The Complete Lifelong Editing Results

Tables 5 and 6 present the lifelong editing performance of various editors on the E-VQA [22], VLKEB [22], and E-IC [22] datasets, with LLaVA [28], BLIP2 [4], and MiniGPT-4 [29] as backbones, respectively. The results demonstrate a similar overall trend, confirming the effectiveness of our method. It is worth noting, however, that all methods exhibit relatively lower performance on the E-IC dataset. Since the image captioning task typically requires adapting descriptions of the entire image, editing involves carrying more complex information. This leads to a decline in the editing efficacy of the editors due to the limitations in representational power.

8.2. Instance Analysis with Difficult Samples

Figure 7 demonstrates the necessity and complementarity of hard and soft routing. Hard routing filters out irrelevant options based on key visuals, preventing the third-ranked expert from receiving excessive weight in soft routing. Then, soft routing adjusts by reducing the weight of the top-ranked expert that was overlooked by hard routing. Therefore, relying solely on the top expert selected by hard routing can lead to errors.

9. Other Discussions

Balance of Multi-Objective Losses: The challenge in multi-goal optimization lies in balancing conflicting objectives. As shown in Fig. 2, the Hard Routing Loss (HRL) doesn't share parameters with Soft Routing/Editing Loss (SRL/EL), allowing it to be optimized separately. For EL and SRL, since EL can only be minimized by selecting the correct expert in a batch, and SRL helps EL choose the right expert, they align without conflict. In EL, we align with previous work [24]. In SRL, both losses boost semantic similarity. Consequently, we assign all loss weights as 1, and LiveEdit's performance confirms the effectiveness of this choice.

Time and Space Overhead of LiveEdit: The edit time and additional memory required for each new expert remain constant, regardless of the sample size. For example, with LLaVA, it takes approximately 0.32 seconds and 160KB of memory on an A800 GPU.

Discussion with MoVA: The key differences between MoVA [71] and LiveEdit are as follows: 1) Task Focus: MoVA emphasizes macro-level task performance, whereas LiveEdit focuses on the micro-level control of specific responses. 2) Intervention: MoVA routes vision encoders externally to the LLM, while LiveEdit operates internally. 3) Routing: In the coarse stage, MoVA uses In-Context Learning (ICL) to filter out task-irrelevant vision experts, whereas LiveEdit removes vision-irrelevant edit experts. In the fine stage, MoVA reduces single-expert bias, while LiveEdit minimizes the influence of prompt-irrelevant edit experts, complementing the previous stage.

Baseline	# Edit	Editors	E-VQA						VLKEB						E-IC					
			Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Average	Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Average	Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Average
LLaVA (7B)	1	FT-L	93.88	87.98	80.25	99.61	94.78	91.30 (±0.42)	94.29	87.00	92.22	91.16	91.37	91.21 (±1.09)	73.48	72.98	65.79	99.28	99.06	82.12 (±0.82)
		FT-M	87.29	76.11	53.23	100.00	96.95	82.72 (±1.05)	76.31	65.57	59.43	100.00	92.35	78.73 (±0.76)	56.19	56.55	49.94	100.00	100.00	72.54 (±0.85)
		MEND	91.23	90.05	91.29	91.02	90.22	90.76 (±0.64)	92.13	91.28	90.22	89.19	90.13	90.59 (±1.24)	92.82	**91.81**	90.59	96.38	93.69	**93.06** (±1.40)
		SERAC	89.33	83.72	84.97	82.05	23.78	72.77 (±0.36)	89.77	89.11	87.92	66.68	14.20	69.54 (±0.83)	88.18	81.03	85.61	84.01	28.58	73.48 (±1.19)
		TP	35.95	36.12	28.65	93.87	97.61	58.44 (±0.33)	50.77	55.70	51.65	87.93	90.43	67.30 (±0.29)	57.63	59.23	55.34	60.90	88.00	64.22 (±0.49)
		LTE	94.16	93.57	**93.59**	94.08	86.26	92.33 (±1.56)	94.42	93.57	93.22	86.84	79.69	89.55 (±1.41)	93.35	91.30	**92.77**	95.77	91.98	93.03 (±0.55)
		RECIPE	91.37	86.51	87.73	94.27	88.88	89.75 (±1.13)	92.67	92.35	91.01	89.67	82.85	89.71 (±0.57)	84.45	76.97	81.57	96.53	96.37	87.18 (±1.08)
		LEMoE	93.60	92.77	89.99	99.28	96.98	94.52 (±1.09)	94.85	93.09	91.67	87.03	87.88	90.90 (±0.29)	**93.80**	91.42	90.61	95.14	93.00	92.79 (±0.41)
		LiveEdit	**94.28**	**94.51**	88.01	**100.00**	**100.00**	**95.36** (±0.57)	**96.43**	**95.22**	**93.72**	**100.00**	**100.00**	**97.08** (±0.62)	82.16	81.01	78.27	**100.00**	**100.00**	88.29 (±1.42)
	10	FT-L	90.57	84.14	73.21	95.56	81.50	85.00 (±1.07)	88.05	85.32	85.23	74.53	85.74	83.77 (±1.22)	68.74	67.05	60.95	97.05	91.20	77.00 (±0.80)
		FT-M	84.90	73.53	49.99	100.00	55.98	72.88 (±0.63)	68.63	57.57	56.56	100.00	82.99	73.15 (±0.23)	57.23	55.92	50.09	100.00	90.31	70.71 (±1.18)
		MEND	3.58	3.55	3.53	2.10	1.26	2.80 (±0.02)	0.18	0.24	0.05	0.03	0.19	0.14 (±0.00)	66.49	64.77	58.42	87.25	86.15	72.62 (±1.13)
		SERAC	88.09	83.40	83.57	64.91	15.50	67.10 (±0.92)	81.55	74.49	80.24	54.71	13.15	60.83 (±0.98)	56.57	56.88	52.62	59.96	14.66	48.14 (±0.40)
		TP	32.71	31.23	28.58	75.10	91.17	51.76 (±0.60)	44.56	47.52	45.36	52.21	66.61	51.25 (±0.69)	45.28	47.25	42.78	19.74	59.59	42.93 (±0.38)
		LTE	92.83	91.41	**90.82**	86.38	85.52	89.39 (±0.34)	90.06	81.52	88.11	83.40	81.48	84.91 (±0.78)	52.16	55.70	48.39	90.74	89.09	67.21 (±0.89)
		RECIPE	90.22	85.92	86.24	90.34	88.11	88.17 (±1.48)	83.92	76.23	82.84	86.33	83.69	82.60 (±0.72)	56.00	56.19	52.14	91.80	95.31	70.29 (±0.98)
		LEMoE	91.95	86.54	79.82	85.19	49.81	78.66 (±1.03)	91.55	84.58	81.03	67.19	72.81	79.43 (±0.52)	**89.00**	**85.23**	**83.24**	86.39	82.86	85.34 (±0.91)
		LiveEdit	**93.79**	**93.21**	86.42	**100.00**	**100.00**	**94.68** (±1.03)	**95.54**	**94.52**	**91.25**	**100.00**	**100.00**	**96.26** (±0.33)	81.93	80.80	75.55	**100.00**	**100.00**	**87.66** (±0.49)
	100	FT-L	79.67	70.05	64.07	83.47	54.44	70.34 (±0.95)	75.41	73.67	74.16	70.01	82.05	75.06 (±1.17)	65.08	60.90	58.46	86.82	89.19	72.09 (±0.65)
		FT-M	82.90	72.83	47.26	100.00	43.39	69.28 (±0.43)	60.60	59.79	56.29	100.00	68.07	68.95 (±0.78)	59.41	55.44	52.16	100.00	72.93	67.99 (±0.43)
		MEND	2.22	2.20	2.21	0.21	0.62	1.49 (±0.02)	0.56	0.58	0.66	0.18	0.07	0.41 (±0.00)	56.83	56.89	53.07	87.63	84.64	67.81 (±0.93)
		SERAC	88.08	81.53	82.48	62.13	12.90	65.42 (±0.32)	72.25	62.43	70.68	53.73	13.69	54.56 (±0.49)	53.35	53.70	49.41	48.04	17.25	44.35 (±0.50)
		TP	29.37	28.72	24.66	14.64	45.01	28.48 (±0.33)	19.71	20.07	19.36	11.40	24.05	18.92 (±0.08)	22.88	25.81	20.90	3.59	14.87	17.61 (±0.06)
		LTE	88.92	87.89	85.72	84.34	81.60	85.69 (±0.37)	80.27	64.25	80.13	81.62	79.11	77.08 (±0.85)	48.37	49.63	45.08	87.66	87.62	63.67 (±0.60)
		RECIPE	89.86	83.32	84.82	87.37	85.08	86.09 (±0.55)	73.97	63.73	72.69	86.20	82.59	75.84 (±0.58)	53.23	53.75	49.36	87.64	95.52	67.90 (±0.36)
		LEMoE	42.41	36.60	34.33	78.57	53.28	49.04 (±0.45)	83.07	75.36	71.72	54.09	49.68	66.78 (±0.60)	55.36	53.14	51.12	87.77	81.35	65.75 (±1.02)
		LiveEdit	**93.54**	**92.34**	**85.89**	**100.00**	**99.31**	**94.21** (±0.34)	**94.56**	**90.65**	**89.56**	**100.00**	**100.00**	**94.95** (±1.34)	**80.81**	**78.77**	**63.52**	**100.00**	**100.00**	**84.62** (±0.62)
	1000	FT-L	71.39	59.83	57.41	55.55	48.99	58.63 (±0.17)	68.14	66.38	66.98	65.61	75.35	68.49 (±0.32)	59.78	54.99	54.17	65.37	78.96	62.65 (±0.43)
		FT-M	69.57	56.34	44.07	100.00	41.47	62.29 (±0.40)	53.41	48.80	43.16	100.00	57.03	60.48 (±0.50)	49.21	47.75	43.81	100.00	35.14	55.18 (±0.88)
		MEND	0.04	0.05	0.05	0.08	0.09	0.06 (±0.00)	0.03	0.05	0.07	0.06	0.08	0.06 (±0.00)	54.39	54.14	50.99	83.87	80.60	64.80 (±0.35)
		SERAC	85.57	75.58	82.01	62.46	15.69	64.26 (±0.37)	60.93	56.49	60.06	52.94	15.04	49.09 (±0.36)	52.93	53.44	49.01	49.91	16.65	44.39 (±0.73)
		TP	16.56	16.80	15.65	7.28	15.60	14.38 (±0.14)	5.46	4.81	5.51	2.77	7.19	5.15 (±0.07)	10.28	13.14	9.75	1.71	4.45	7.87 (±0.13)
		LTE	83.93	82.55	81.34	83.97	73.09	80.98 (±1.36)	64.51	56.26	64.80	80.85	76.52	68.59 (±0.60)	48.83	49.96	45.68	85.17	86.41	63.21 (±0.67)
		RECIPE	87.00	76.81	83.09	86.95	87.03	84.18 (±0.80)	62.00	56.84	61.50	85.37	82.07	69.56 (±0.31)	53.11	53.48	48.99	87.93	94.84	67.67 (±1.11)
		LEMoE	30.80	25.75	24.32	71.45	46.23	39.71 (±0.23)	67.97	61.07	58.16	48.48	44.06	55.95 (±0.36)	34.50	31.38	28.14	82.09	75.88	50.40 (±0.26)
		LiveEdit	**92.93**	**90.16**	**84.30**	**100.00**	**96.43**	**92.76** (±0.20)	**92.22**	**83.97**	**82.75**	**100.00**	**100.00**	**91.79** (±0.55)	**72.80**	**69.95**	**57.05**	**100.00**	**99.79**	**79.92** (±0.72)
BLIP2 (2.7B)	1	FT-L	52.86	48.80	32.94	98.24	94.27	65.42 (±0.69)	54.31	54.27	54.08	98.40	94.37	71.09 (±1.05)	45.02	44.47	40.72	99.02	98.27	65.50 (±0.49)
		FT-M	91.70	87.24	33.30	100.00	85.22	79.49 (±0.72)	92.64	80.97	63.62	100.00	83.02	84.05 (±0.70)	67.14	61.76	43.34	100.00	96.76	73.80 (±0.16)
		MEND	93.13	92.76	93.07	92.00	75.81	89.35 (±0.93)	94.91	93.81	93.84	94.98	86.54	92.82 (±0.82)	**94.96**	**92.45**	**92.33**	94.95	88.86	92.71 (±1.45)
		SERAC	88.39	84.50	84.25	85.82	26.00	73.79 (±1.01)	87.95	84.67	85.20	68.10	17.75	68.73 (±0.97)	88.71	83.81	84.38	84.28	24.70	73.18 (±0.25)
		TP	70.14	65.80	53.05	98.11	85.33	74.49 (±0.38)	50.98	49.47	50.88	94.76	78.57	64.93 (±1.02)	49.65	48.58	46.02	93.69	78.95	63.38 (±1.00)
		LTE	95.74	93.86	86.90	97.93	87.97	92.48 (±0.70)	94.13	91.93	92.23	93.89	92.27	92.89 (±1.01)	92.58	91.94	90.90	97.80	91.42	**92.93** (±1.37)
		RECIPE	89.42	86.24	87.53	99.87	89.16	90.45 (±1.46)	92.38	89.74	89.17	97.13	94.46	92.58 (±1.16)	85.20	81.44	82.71	100.00	94.59	88.79 (±1.23)
		LEMoE	93.56	92.23	91.40	98.50	85.21	92.18 (±0.73)	94.59	93.14	92.37	94.53	61.53	87.23 (±0.34)	93.07	91.37	83.28	94.45	60.44	84.52 (±0.38)
		LiveEdit	**96.67**	**94.20**	**93.82**	**100.00**	**100.00**	**96.94** (±1.32)	**98.77**	**98.08**	**94.89**	**100.00**	**100.00**	**98.35** (±1.58)	80.60	80.12	76.88	**100.00**	**100.00**	87.52 (±0.31)
	10	FT-L	53.77	48.91	37.66	97.68	80.40	63.68 (±0.88)	55.81	55.55	55.29	93.37	75.45	67.09 (±0.71)	46.20	44.07	40.54	98.37	92.59	64.35 (±0.75)
		FT-M	85.27	81.09	34.60	100.00	49.00	69.99 (±0.15)	92.48	86.02	71.48	100.00	60.34	82.06 (±0.47)	63.79	56.88	47.75	100.00	52.41	64.17 (±0.42)
		MEND	40.40	32.21	27.73	87.92	67.74	51.20 (±0.39)	54.47	52.59	53.61	91.96	84.58	67.44 (±0.16)	7.15	7.48	7.01	19.74	21.70	12.62 (±0.05)
		SERAC	86.12	82.24	82.52	70.24	16.06	67.43 (±1.10)	78.43	67.72	75.12	56.21	14.54	58.40 (±0.73)	47.69	46.94	43.46	65.29	16.65	44.01 (±0.15)
		TP	55.86	52.39	43.00	91.26	63.62	61.22 (±0.54)	50.19	51.61	50.13	89.26	68.48	61.93 (±0.16)	44.57	44.88	39.82	65.84	51.55	49.33 (±0.37)
		LTE	94.96	92.43	85.23	94.19	87.01	90.76 (±1.19)	92.02	80.25	90.30	94.66	90.39	89.52 (±1.47)	47.56	50.73	44.64	97.05	92.90	66.57 (±0.78)
		RECIPE	88.27	85.37	84.80	97.78	88.49	88.94 (±0.76)	81.39	70.31	79.19	94.40	95.59	84.18 (±0.74)	47.26	46.30	42.83	100.00	95.38	66.35 (±0.76)
		LEMoE	91.59	85.92	86.82	83.64	29.71	75.54 (±0.86)	91.40	90.71	89.12	58.20	49.49	75.79 (±0.48)	**87.45**	**84.54**	71.74	85.98	71.23	80.19 (±0.60)
		LiveEdit	**95.65**	**93.46**	**90.98**	**100.00**	**100.00**	**96.02** (±0.93)	**98.57**	**97.92**	**94.63**	**100.00**	**100.00**	**98.22** (±0.48)	80.65	79.38	**75.66**	**100.00**	**100.00**	**87.14** (±0.96)
	100	FT-L	52.47	43.68	39.43	91.13	47.72	54.88 (±0.83)	57.26	56.23	56.73	86.19	68.87	65.06 (±0.15)	50.81	48.70	42.79	90.12	52.09	56.90 (±0.89)
		FT-M	48.04	40.71	29.86	100.00	33.25	50.37 (±0.43)	56.49	58.16	53.13	100.00	43.52	62.26 (±0.91)	56.23	53.18	43.37	100.00	31.59	56.87 (±0.44)
		MEND	17.69	16.39	18.31	91.52	67.94	42.37 (±0.45)	39.91	40.96	40.37	92.32	84.32	59.58 (±0.19)	7.97	8.19	8.04	20.15	23.23	13.52 (±0.13)
		SERAC	86.98	81.27	80.97	71.04	15.55	67.17 (±0.60)	66.73	53.69	64.37	59.18	17.86	52.37 (±0.70)	43.55	42.16	39.09	57.77	15.84	39.68 (±0.50)
		TP	44.26	38.24	33.28	43.81	38.54	39.63 (±0.10)	46.61	48.12	47.22	64.25	43.21	49.88 (±0.49)	37.51	37.62	33.80	10.26	20.84	28.00 (±0.47)
		LTE	92.78	90.36	83.88	94.33	81.79	88.63 (±0.48)	78.41	60.70	78.55	94.34	90.13	80.42 (±0.72)	43.03	42.52	39.63	96.20	91.95	62.67 (±0.81)
		RECIPE	88.12	82.49	83.10	98.97	86.53	87.84 (±1.06)	68.79	54.96	67.20	94.61	97.12	76.54 (±0.77)	43.40	42.19	39.02	98.40	95.55	63.71 (±0.86)
		LEMoE	29.61	21.70	27.05	79.48	32.61	38.09 (±0.11)	44.13	44.78	42.55	58.13	51.76	48.27 (±0.71)	57.13	51.73	46.73	91.72	58.06	61.08 (±0.34)
		LiveEdit	**95.25**	**92.94**	**85.48**	**100.00**	**99.76**	**94.69** (±0.50)	**98.20**	**97.67**	**93.96**	**100.00**	**100.00**	**97.97** (±0.34)	**79.18**	**77.44**	**74.12**	**100.00**	**100.00**	**86.15** (±0.33)
	1000	FT-L	45.10	34.62	35.42	48.42	41.24	40.96 (±0.29)	55.39	54.34	53.87	50.80	54.00	53.68 (±0.80)	53.61	48.81	45.92	52.45	59.09	51.98 (±0.70)
		FT-M	40.40	31.46	27.85	100.00	27.44	45.43 (±0.68)	47.03	49.68	46.99	100.00	41.41	57.02 (±0.13)	48.24	45.55	43.03	100.00	23.76	52.12 (±0.16)
		MEND	15.84	14.35	17.73	91.74	70.17	41.97 (±0.12)	37.22	38.03	37.19	91.49	84.10	57.61 (±0.58)	6.54	6.51	6.50	13.52	20.38	10.69 (±0.14)
		SERAC	83.35	70.80	80.32	67.66	13.13	63.05 (±0.87)	53.58	45.78	52.42	56.81	16.90	45.10 (±0.38)	43.12	41.69	38.72	48.08	14.88	37.30 (±0.27)
		TP	20.63	15.09	18.41	8.65	8.25	14.21 (±0.18)	24.36	24.21	24.25	16.37	19.96	21.83 (±0.14)	26.03	26.31	24.85	4.11	11.77	18.62 (±0.12)
		LTE	89.32	82.82	81.51	94.86	69.83	83.67 (±1.05)	61.67	51.05	61.60	94.78	90.94	72.01 (±0.66)	44.52	43.56	41.29	96.45	90.86	63.33 (±0.34)
		RECIPE	84.99	74.20	82.04	96.82	87.73	85.16 (±1.32)	54.64	46.54	54.10	94.60	96.93	69.37 (±1.04)	43.02	41.63	38.59	99.68	92.96	63.18 (±0.16)
		LEMoE	19.73	17.34	18.22	72.01	31.06	31.67 (±0.14)	34.74	33.43	32.05	55.55	50.04	41.16 (±0.58)	43.46	43.34	37.69	93.27	67.52	57.06 (±0.38)
		LiveEdit	**94.42**	**91.98**	**84.65**	**100.00**	**97.38**	**93.69** (±0.67)	**97.00**	**91.92**	**87.53**	**100.00**	**100.00**	**95.29** (±1.48)	**72.86**	**70.34**	**67.92**	**100.00**	**100.00**	**82.22** (±0.62)

Table 5. Lifelong editing performance on BLIP2-OPT (2.7B) and LLaVA-V1.5 (7B) across the E-VQA and VLKEB datasets.

# Edit	Editors	E-VQA						VLKEB						E-IC					
		Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Average	Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Average	Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Average
1	FT-L	93.85	86.25	88.58	99.47	84.50	90.53 $_{(\pm 0.20)}$	82.17	81.64	78.45	98.45	75.06	83.16 $_{(\pm 0.99)}$	60.40	59.47	52.79	99.95	92.46	73.01 $_{(\pm 1.00)}$
	FT-M	91.42	85.82	49.09	100.00	86.72	82.61 $_{(\pm 0.51)}$	93.51	91.28	68.30	100.00	81.78	86.97 $_{(\pm 1.36)}$	71.60	67.34	49.55	100.00	91.69	76.04 $_{(\pm 0.71)}$
	MEND	93.47	91.32	91.22	81.51	74.97	86.50 $_{(\pm 1.17)}$	90.28	89.18	89.55	93.63	86.69	89.87 $_{(\pm 0.78)}$	90.28	83.91	79.05	95.92	93.97	88.63 $_{(\pm 0.89)}$
	SERAC	87.66	69.24	85.96	82.11	24.11	69.81 $_{(\pm 0.84)}$	89.28	89.06	87.35	64.88	12.72	68.66 $_{(\pm 0.76)}$	85.09	78.96	81.70	83.36	24.50	70.72 $_{(\pm 0.54)}$
	TP	52.45	48.39	51.56	93.93	83.61	65.99 $_{(\pm 0.91)}$	49.17	51.88	48.92	90.63	79.79	64.08 $_{(\pm 0.98)}$	52.19	51.49	49.81	83.33	72.04	61.77 $_{(\pm 0.78)}$
	LTE	93.54	77.59	84.07	94.03	88.14	87.48 $_{(\pm 0.45)}$	92.92	85.16	86.25	88.29	81.96	86.91 $_{(\pm 0.71)}$	**91.30**	84.62	83.55	94.96	91.88	89.26 $_{(\pm 1.01)}$
	RECIPE	89.23	70.39	87.94	95.49	90.90	86.79 $_{(\pm 0.42)}$	93.20	92.40	90.71	90.29	84.35	90.19 $_{(\pm 0.29)}$	81.25	74.44	76.96	97.18	92.17	84.40 $_{(\pm 0.91)}$
	LEMoE	92.59	90.49	89.16	97.01	84.33	90.71 $_{(\pm 0.94)}$	92.26	89.28	88.27	93.87	67.49	86.23 $_{(\pm 0.89)}$	90.32	**86.30**	**85.75**	96.66	69.17	85.64 $_{(\pm 1.42)}$
	LiveEdit	**94.90**	**92.63**	**92.06**	**100.00**	**100.00**	**95.92** $_{(\pm 0.36)}$	**96.41**	**95.56**	**92.85**	**100.00**	**100.00**	**96.96** $_{(\pm 1.04)}$	86.84	83.82	76.95	**100.00**	**100.00**	**89.52** $_{(\pm 1.20)}$
10	FT-L	84.14	76.41	71.19	97.84	53.98	76.71 $_{(\pm 0.73)}$	84.20	81.22	82.12	95.47	68.37	82.28 $_{(\pm 0.73)}$	62.28	59.67	52.88	98.07	55.84	65.75 $_{(\pm 0.83)}$
	FT-M	88.05	81.83	53.26	100.00	53.01	75.23 $_{(\pm 0.58)}$	91.31	88.82	63.64	100.00	63.40	81.43 $_{(\pm 1.15)}$	79.04	70.57	53.82	100.00	53.19	71.33 $_{(\pm 0.31)}$
	MEND	37.34	31.66	35.52	44.31	37.24	37.21 $_{(\pm 0.18)}$	57.01	55.39	55.34	72.03	67.66	61.49 $_{(\pm 0.47)}$	69.98	69.35	59.25	93.39	91.52	76.69 $_{(\pm 1.20)}$
	SERAC	86.61	67.86	83.14	67.69	14.47	63.95 $_{(\pm 0.42)}$	79.90	71.80	78.44	52.64	13.08	59.17 $_{(\pm 0.41)}$	50.72	49.60	48.36	62.87	16.01	45.51 $_{(\pm 0.47)}$
	TP	35.71	30.29	36.38	74.53	55.67	46.52 $_{(\pm 0.34)}$	44.70	47.40	44.74	69.61	70.01	55.29 $_{(\pm 0.13)}$	47.02	47.29	46.06	50.10	42.77	46.65 $_{(\pm 0.64)}$
	LTE	88.55	73.87	83.91	89.44	88.02	84.76 $_{(\pm 0.25)}$	89.10	81.76	85.96	85.53	81.06	84.68 $_{(\pm 1.26)}$	54.77	55.69	52.24	90.84	93.06	69.32 $_{(\pm 1.04)}$
	RECIPE	88.44	69.59	85.61	90.68	89.78	84.82 $_{(\pm 1.27)}$	82.58	74.01	81.18	88.33	85.84	82.39 $_{(\pm 0.20)}$	50.38	48.96	48.11	92.27	92.69	66.48 $_{(\pm 0.78)}$
	LEMoE	92.55	86.51	83.31	79.76	42.61	76.94 $_{(\pm 0.65)}$	89.02	86.94	85.18	67.60	50.75	75.90 $_{(\pm 1.13)}$	**88.13**	**84.15**	75.86	92.84	47.92	77.78 $_{(\pm 0.76)}$
	LiveEdit	**94.06**	**91.84**	**90.60**	**100.00**	**99.92**	**95.28** $_{(\pm 0.97)}$	**95.75**	**95.13**	**91.99**	**100.00**	**100.00**	**96.57** $_{(\pm 1.50)}$	85.79	83.52	**76.09**	**100.00**	**100.00**	**89.08** $_{(\pm 1.33)}$
100	FT-L	64.81	58.11	54.23	93.62	44.98	63.15 $_{(\pm 1.06)}$	69.98	69.05	68.09	88.74	63.07	71.79 $_{(\pm 0.16)}$	62.35	57.49	54.13	96.15	46.86	63.39 $_{(\pm 0.17)}$
	FT-M	66.06	51.17	47.24	100.00	36.65	60.22 $_{(\pm 0.28)}$	60.03	60.18	57.00	100.00	51.24	65.69 $_{(\pm 0.54)}$	58.93	54.35	47.84	100.00	36.80	59.59 $_{(\pm 0.14)}$
	MEND	23.32	20.75	21.42	47.27	42.60	31.07 $_{(\pm 0.18)}$	38.85	40.02	37.99	72.82	68.89	51.71 $_{(\pm 0.39)}$	51.92	53.96	48.24	86.02	83.88	64.80 $_{(\pm 0.79)}$
	SERAC	87.39	66.53	81.70	65.13	14.04	62.96 $_{(\pm 0.64)}$	69.38	58.31	67.57	51.71	13.72	52.14 $_{(\pm 0.78)}$	47.41	46.17	45.22	49.37	13.52	40.34 $_{(\pm 0.55)}$
	TP	19.46	20.32	22.03	41.85	32.91	27.32 $_{(\pm 0.39)}$	39.60	42.52	39.14	36.15	43.68	40.21 $_{(\pm 0.22)}$	39.60	39.64	38.48	17.29	26.04	32.21 $_{(\pm 0.40)}$
	LTE	85.05	71.80	79.61	87.06	86.10	81.92 $_{(\pm 0.50)}$	79.45	63.55	79.46	84.18	79.42	77.21 $_{(\pm 1.26)}$	50.47	49.57	48.17	88.12	92.87	65.84 $_{(\pm 0.70)}$
	RECIPE	88.26	67.74	81.48	91.77	88.23	83.49 $_{(\pm 1.18)}$	71.54	60.04	70.35	85.57	86.85	74.87 $_{(\pm 1.06)}$	47.27	46.06	45.12	89.52	94.25	64.44 $_{(\pm 0.68)}$
	LEMoE	27.84	26.38	26.73	68.45	33.73	36.63 $_{(\pm 0.21)}$	44.75	43.33	41.96	73.90	60.38	52.86 $_{(\pm 0.62)}$	52.85	50.20	48.11	92.08	49.63	58.57 $_{(\pm 0.53)}$
	LiveEdit	**93.23**	**91.27**	**88.34**	**100.00**	**99.61**	**94.49** $_{(\pm 0.26)}$	**94.06**	**94.62**	**89.42**	**100.00**	**100.00**	**95.62** $_{(\pm 1.26)}$	**84.38**	**83.53**	**71.85**	**100.00**	**100.00**	**87.95** $_{(\pm 0.86)}$
1000	FT-L	58.97	46.34	50.77	72.04	41.92	54.01 $_{(\pm 0.84)}$	61.85	60.82	60.41	76.17	60.34	63.92 $_{(\pm 0.45)}$	58.88	53.98	52.98	93.31	46.49	61.13 $_{(\pm 0.74)}$
	FT-M	51.18	42.79	40.08	100.00	37.28	54.27 $_{(\pm 0.76)}$	51.26	54.35	51.13	100.00	50.72	61.49 $_{(\pm 1.03)}$	52.65	47.90	47.25	100.00	33.69	56.30 $_{(\pm 0.58)}$
	MEND	31.84	26.60	33.98	43.08	44.90	36.08 $_{(\pm 0.15)}$	42.35	45.07	42.68	66.98	62.32	51.88 $_{(\pm 0.69)}$	49.30	51.46	46.70	81.03	82.53	62.20 $_{(\pm 0.30)}$
	SERAC	84.50	60.36	81.83	63.26	13.40	60.67 $_{(\pm 0.22)}$	57.02	51.66	56.80	48.48	13.24	45.44 $_{(\pm 0.39)}$	47.03	45.79	44.82	42.12	15.31	39.01 $_{(\pm 0.56)}$
	TP	9.24	8.96	10.25	20.54	17.03	13.20 $_{(\pm 0.19)}$	23.84	25.00	23.54	18.79	25.58	23.35 $_{(\pm 0.06)}$	25.07	24.50	24.09	14.77	18.05	21.30 $_{(\pm 0.30)}$
	LTE	81.93	66.81	75.96	88.24	78.95	78.38 $_{(\pm 0.56)}$	64.28	56.91	64.07	83.96	78.59	69.56 $_{(\pm 0.51)}$	51.31	49.25	49.19	88.19	90.53	65.69 $_{(\pm 0.91)}$
	RECIPE	85.33	61.79	76.35	90.33	89.08	80.58 $_{(\pm 1.12)}$	58.18	52.31	58.14	85.49	83.84	67.59 $_{(\pm 0.29)}$	47.02	45.76	44.85	89.80	92.26	63.94 $_{(\pm 0.79)}$
	LEMoE	25.07	23.87	26.68	89.11	50.50	43.04 $_{(\pm 0.29)}$	39.48	38.24	36.52	82.35	58.01	50.92 $_{(\pm 0.12)}$	50.61	47.35	47.19	95.11	47.62	57.58 $_{(\pm 0.91)}$
	LiveEdit	**92.75**	**89.81**	**85.12**	**100.00**	**96.82**	**92.90** $_{(\pm 0.21)}$	**93.41**	**89.87**	**85.21**	**100.00**	**100.00**	**93.70** $_{(\pm 0.31)}$	**81.43**	**80.34**	**62.25**	**100.00**	**99.73**	**84.75** $_{(\pm 1.42)}$

Table 6. Lifelong editing performance on MiniGPT-4 (7B) across the E-VQA and VLKEB datasets.