

Table 1: **(To Reviewer1-Q1)** Performance of one-step editing on LLaVA1.5 (We average the results on E-IC and E-VQA).

Method	Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Avg
FT	67.4	57.9	54.8	67.1	63.2	62.1
KE	72.7	65.4	55.3	82.2	67.3	68.4
T-Patcher	89.0	76.5	69.0	81.2	81.3	79.6
MEND	95.4	92.6	78.3	83.5	80.3	86.0
IKE	93.4	85.1	77.9	27.7	3.2	57.5
SERAC	<b>96.3</b>	92.4	85.5	83.3	7.7	73.0
<b>UniKE</b>	95.7	<b>92.8</b>	<b>88.4</b>	<b>86.0</b>	<b>86.4</b>	<b>89.9</b>

Table 2: **(To Reviewer1-Q1)** Performance of cross-task editing on LLaVA1.5.

Method	Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Avg
FT	66.9	57.2	51.0	62.3	54.4	56.4
KE	69.8	60.3	52.5	79.3	62.1	64.8
T-Patcher	81.2	60.0	57.4	77.4	76.5	70.5
MEND	90.4	84.3	73.8	78.6	76.0	80.6
SERAC	92.1	88.3	82.5	82.2	1.2	69.3
<b>UniKE</b>	<b>92.2</b>	<b>89.2</b>	<b>83.8</b>	<b>82.7</b>	<b>84.7</b>	<b>86.5</b>

Table 3: **(To Reviewer1-Q1)** Performance of cross-task editing on BLIP-2 OPT.

Method	Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Avg
FT	57.2	49.9	43.2	52.2	49.7	50.4
KE	64.2	60.1	57.2	83.5	59.2	64.8
T-Patcher	83.1	69.7	65.9	84.5	77.9	76.2
MEND	84.2	82.4	74.9	91.4	80.2	82.6
SERAC	90.8	89.2	84.1	90.0	1.7	71.2
<b>UniKE</b>	<b>91.1</b>	<b>90.6</b>	<b>88.2</b>	<b>91.7</b>	<b>85.6</b>	<b>89.4</b>

Table 4: **(To Reviewer1-Q2)** Comparison with recent baselines for one-step editing on MiniGPT-4 (We average the results on E-IC and E-VQA).

Method	Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Avg
MENMET	97.0	96.2	82.4	98.0	85.2	91.8
WISE	97.2	92.2	88.7	98.4	<b>88.2</b>	93.0
<b>UniKE</b>	<b>97.4</b>	<b>96.6</b>	<b>92.6</b>	<b>98.8</b>	88.1	<b>94.7</b>

Table 5: **(To Reviewer1-Q2)** Comparison with recent baselines for cross-task editing on MiniGPT-4.

Method	Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Avg
MENMET	88.4	87.2	78.0	86.1	82.5	84.4
WISE	89.2	85.4	83.4	87.8	83.6	85.9
<b>UniKE</b>	<b>90.7</b>	<b>88.2</b>	<b>86.8</b>	<b>90.4</b>	<b>83.8</b>	<b>88.0</b>

Table 6: **(To Reviewer1-Q3)** Performance of each method on LLM editing task (ZsRE) for one-step editing and 200-step editing.

Method	ONE-STEP EDITING				200-STEP EDITING			
	Rel.	Gen.	Loc.	Avg	Rel.	Gen.	Loc.	Avg
FT	77.4	76.7	35.5	63.2	19.5	17.2	5.4	14.0
KE	20.6	20.1	81.3	40.7	7.6	6.8	65.8	26.7
T-Patcher	97.1	95.0	96.2	96.1	81.4	70.6	91.3	81.1
MEND	98.2	97.7	97.4	97.8	0.0	0.0	0.0	0.0
IKE	99.4	97.2	59.2	85.3	-	-	-	-
SERAC	88.6	87.9	99.9	92.1	24.0	23.2	<b>96.4</b>	47.9
MENMET	99.1	86.8	97.4	94.4	82.9	73.6	90.2	82.2
WISE	98.8	96.3	<b>99.9</b>	98.3	82.8	74.7	95.5	84.3
<b>UniKE</b>	<b>99.5</b>	<b>97.9</b>	99.6	<b>99.0</b>	<b>85.1</b>	<b>76.7</b>	95.6	<b>85.8</b>

Table 7: **(To Reviewer2-Q1)** The computational speed, resource utilization and performance of each method. We use the average results of five metrics (Reliability, T-Generality, M-Generality, T-Locality, and M-Locality) as the performance measure.

Method	GPU memory	editing time for each sample	Avg performance
FT	22G	6.1s	60.6
KE	24G	5.8s	74.7
T-Patcher	18G	4.7s	80.4
MEND	36G	5.2s	90.3
IKE	20G	1.6s	65.5
SERAC	49G	3.6s	76.4
<b>UniKE</b>	18G	5.0s	<b>95.2</b>

Table 8: **(To Reviewer3-Q1)** The results of CKA evaluation on MiniGPT-4 with the setup of one-step editing.

Method	T-Patcher	MEND	IKE	<b>UniKE</b>
CKA↑	1.38	1.33	1.47	<b>1.58</b>

Table 9: **(To Reviewer3-Q2)** Performance of counterfactual editing on MiniGPT-4.

Method	Rel.	T-Gen.	M-Gen.	T-Loc.	M-Loc.	Avg
T-Patcher	80.0	65.9	57.7	84.2	88.3	75.2
MEND	90.6	83.2	74.1	93.5	82.1	84.7
IKE	90.3	83.7	<b>81.5</b>	44.1	5.0	60.9
<b>UniKE</b>	<b>90.8</b>	<b>84.7</b>	80.7	<b>94.9</b>	<b>94.5</b>	<b>89.1</b>

Table 10: **(To Reviewer4-Q2)** Editing time cost and performance with/without encoders for UniKE. The time refers to the average editing or inference time for one sample. Gen is the average result of T-Generality and M-Generality; while Loc is the average result of T-Locality and M-Locality.

Method	GPU Memory	Editing time	Inference time	Rel.	Gen.	Loc.
w/o encoders	17.7GB	4.92s	0.212s	96.2	91.2	90.3
<b>UniKE</b>	17.8GB	5.04s	0.217s	<b>97.4</b>	<b>94.6</b>	<b>93.5</b>

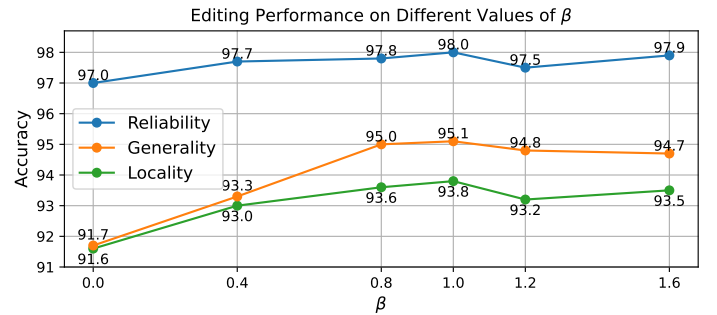


Figure 1: **(To Reviewer4-Q3)** Editing performance on different value of  $\beta$ . Generality is the average result of T-Generality and M-Generality; while Locality is the average result of T-Locality and M-Locality.

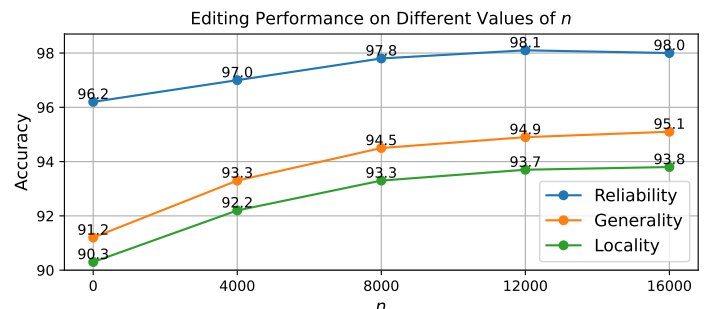


Figure 2: **(To Reviewer2-Q4)** Editing performance on different value of  $n$ .