Table 1. Performance of hypernetwork-based methods KE and MEND measured by metrics in Sec. 5.1 in MAC (3898 edits), SMAC (3898 edits) on PEAK-CF and MQC (992 edits), SMQC (992 edits) on PEAK-T.

Method	PEAK-CF							PEAK-T						
	MAC			SMAC				MQC			SMQC			
	SR-S↑	SR-1↑	GSR↑	SR-S↑	SR-1↑	GSR↑	LSR↑	SR-S↑	SR-1↑	GSR↑	SR-S↑	SR-1↑	GSR↑	LSR↑
FT	0.31	12.01	0.36	0.00	0.00	0.00	5.25	0.00	3.23	0.40	0.00	0.00	0.00	2.12
FT-C	3.64	15.89	3.34	-	-	-	-	0.20	6.57	0.20	-	-	-	-
ROME	0.92	16.30	0.82	0.00	0.62	0.00	12.25	19.15	21.92	14.52	0.00	0.00	0.00	3.63
ROME-C	2.72	26.46	2.36	-	-	-	-	19.56	22.27	14.52	-	-	-	-
KE	0.92	10.31	0.72	0.00	0.00	0.00	10.51	4.23	8.06	3.13	0.00	0.00	0.00	2.98
MEND	2.36	23.09	1.90	0.00	0.56	0.00	12.35	14.11	17.54	8.77	0.00	0.00	0.00	4.02
AlphaEdit	5.44	31.01	4.57	6.98	33.24	6.16	15.17	16.33	18.44	11.09	5.65	25.11	5.04	4.11
AlphaEdit-C	2.72	34.37	2.77	-	-	-	-	15.73	17.84	9.48	-	-	-	-
WISE	39.99	65.20	38.60	0.10	5.65	0.10	15.21	27.62	32.77	24.60	0.20	3.33	0.20	4.40
T-patcher	26.64	52.54	12.83	0.00	0.59	0.00	1.01	26.81	31.68	25.20	0.00	0.00	0.00	0.32
GRACE	15.20	57.44	14.53	12.99	50.16	9.75	**27.19**	6.05	32.06	5.54	4.84	30.55	3.63	38.59
MELO	19.46	44.33	18.47	18.22	38.51	16.68	24.13	6.25	23.49	5.85	2.82	19.86	2.02	36.25
ARK	**48.97**	**75.49**	**45.69**	**43.38**	**70.72**	**39.48**	25.76	**46.17**	**68.85**	**38.31**	**28.63**	**53.73**	**21.77**	**41.37**

Table 2. Performance of hypernetwork-based methods KE and MEND measured by metrics in Sec. 5.1 in MQC (1948 edits), SMQC (1948 edits) on PEAK-CF and MAC (1984 edits), SMAC (1984 edits) on PEAK-T.

Method	PEAK-CF							PEAK-T						
	MQC			SMQC				MAC			SMAC			
	SR-S↑	SR-1↑	GSR↑	SR-S↑	SR-1↑	GSR↑	LSR↑	SR-S↑	SR-1↑	GSR↑	SR-S↑	SR-1↑	GSR↑	LSR↑
FT	5.64	21.65	4.82	0.00	0.00	0.00	2.09	0.10	1.92	0.10	0.00	0.00	0.00	3.79
FT-C	4.67	21.71	4.00	-	-	-	-	0.20	2.02	0.20	-	-	-	-
ROME	23.49	47.08	19.38	0.00	0.41	0.00	4.52	0.30	18.24	0.60	0.00	0.00	0.00	6.22
ROME-C	22.15	46.41	18.26	-	-	-	-	2.12	22.48	2.12	-	-	-	-
KE	1.03	15.40	0.77	0.00	0.00	0.00	3.92	0.10	5.04	0.10	0.00	0.00	0.00	4.56
MEND	5.23	22.38	5.05	0.00	0.00	0.00	3.77	1.01	12.40	0.81	0.00	0.00	0.00	5.41
AlphaEdit	7.69	28.57	5.64	5.33	24.92	4.21	6.92	2.02	21.78	2.12	0.50	16.28	0.41	8.45
AlphaEdit-C	7.69	28.72	5.85	-	-	-	-	1.71	27.62	1.01	-	-	-	-
WISE	46.56	70.11	44.51	0.0	3.23	0.0	7.28	23.08	**54.64**	21.37	0.20	2.92	0.20	8.40
T-patcher	33.23	60.41	31.30	0.00	0.00	0.00	0.45	20.77	43.6	18.15	0.00	0.00	0.00	0.67
GRACE	6.56	39.02	5.23	3.59	34.61	2.87	38.71	6.55	41.93	6.05	5.65	33.07	5.04	37.54
MELO	20.31	44.16	18.47	9.13	36.16	8.21	43.22	9.48	28.94	9.07	6.96	22.94	6.56	35.98
ARK	**52.51**	**73.49**	**52.41**	**26.46**	**53.44**	**20.31**	**45.73**	**41.63**	51.62	**36.90**	**38.00**	**49.60**	**35.89**	**40.97**

Table 3. Performance measured by SR-S and SR-AS in MAC (3898 edits), SMAC (3898 edits) on PEAK-CF and MQC (992 edits), SMQC (992 edits) on PEAK-T.

Method	PEAK-CF				PEAK-T			
	MAC		SMAC		MQC		SMQC	
	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑
FT	0.31	0.26	0.00	0.00	0.00	0.00	0.00	0.00
FT-C	3.64	3.18	-	-	0.20	0.00	-	-
ROME	0.92	0.62	0.00	0.00	19.15	15.12	0.00	0.00
ROME-C	2.72	2.62	-	-	19.56	15.83	-	-
KE	0.92	0.41	0.00	0.00	4.23	3.53	0.00	0.00
MEND	2.36	1.95	0.00	0.00	14.11	10.28	0.00	0.00
AlphaEdit	5.44	4.67	6.98	5.23	16.33	12.70	5.65	4.54
AlphaEdit-C	2.72	2.31	-	-	15.73	12.40	-	-
WISE	39.99	30.12	0.10	0.00	27.62	21.67	0.20	0.00
T-patcher	26.64	18.98	0.00	0.00	26.81	20.26	0.00	0.00
GRACE	15.20	11.90	12.99	10.31	6.05	4.54	4.84	3.93
MELO	19.46	15.03	18.22	14.73	6.25	5.04	2.82	2.28
ARK	**48.97**	**40.69**	**43.38**	**35.92**	**46.17**	**37.40**	**28.63**	**24.19**

Table 4. Performance measured by SR-S and SR-AS in MQC (1948 edits), SMQC (1948 edits) on PEAK-CF and MAC (1984 edits), SMAC (1984 edits) on PEAK-T.

METHOD	PEAK-CF				PEAK-T			
	MQC		SMQC		MAC		SMAC	
	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑
FT	5.64	3.70	0.00	0.00	0.10	0.00	0.00	0.00
FT-C	4.67	3.70	-	-	0.20	0.00	-	-
ROME	23.49	18.99	0.00	0.00	0.30	0.05	0.00	0.00
ROME-C	22.15	17.86	-	-	2.12	1.85	-	-
KE	1.03	0.81	0.00	0.00	0.10	0.00	0.00	0.00
MEND	5.23	3.59	0.00	0.00	1.01	0.05	0.00	0.00
ALPHAEDIT	7.69	6.16	5.33	4.31	2.02	1.64	0.50	0.30
ALPHAEDIT-C	7.69	5.95	-	-	1.71	1.33	-	-
WISE	46.56	37.68	0.00	0.00	23.08	17.97	0.20	0.10
T-PATCHER	33.23	26.39	0.00	0.00	20.77	16.74	0.00	0.00
GRACE	6.56	4.93	3.59	2.77	6.55	5.13	5.65	4.31
MELO	20.31	15.61	9.13	7.39	9.48	7.60	6.96	5.54
ARK	**52.51**	**43.02**	**26.46**	**22.07**	**41.63**	**33.68**	**38.00**	**30.90**

Table 5. Performance with Mistral-7B measured by metrics in Sec. 5.1 in MAC (3898 edits), SMAC (3898 edits) on PEAK-CF and MQC (992 edits), SMQC (992 edits) on PEAK-T.

METHOD	PEAK-CF							PEAK-T						
	MAC			SMAC				MQC			SMQC			
	SR-S↑	SR-1↑	GSR↑	SR-S↑	SR-1↑	GSR↑	LSR↑	SR-S↑	SR-1↑	GSR↑	SR-S↑	SR-1↑	GSR↑	LSR↑
WISE	33.91	58.70	32.68	0.10	4.57	0.10	14.98	23.39	27.42	20.56	0.00	2.82	0.00	4.20
MELO	15.24	35.25	14.52	13.80	32.99	13.23	23.11	5.44	19.76	4.84	2.42	16.73	1.81	36.20
ARK	**40.84**	**64.03**	**39.86**	**35.73**	**59.55**	**33.30**	**24.21**	**38.31**	**57.66**	**32.06**	**23.99**	**44.96**	**19.15**	**40.27**

Table 6. Performance with Mistral-7B measured by metrics in Sec. 5.1 in MQC (1948 edits), SMQC (1948 edits) on PEAK-CF and MAC (1984 edits), SMAC (1984 edits) on PEAK-T.

METHOD	PEAK-CF							PEAK-T						
	MQC			SMQC				MAC			SMAC			
	SR-S↑	SR-1↑	GSR↑	SR-S↑	SR-1↑	GSR↑	LSR↑	SR-S↑	SR-1↑	GSR↑	SR-S↑	SR-1↑	GSR↑	LSR↑
WISE	38.96	58.73	37.27	0.00	2.71	0.00	7.32	19.25	40.93	17.84	0.00	1.21	0.00	8.02
MELO	16.94	36.96	15.40	7.60	30.29	6.88	42.10	7.86	23.99	74.60	5.75	18.95	5.24	34.76
ARK	**43.94**	**61.50**	**43.74**	**22.33**	**44.76**	**18.06**	**45.33**	**34.80**	**43.43**	**31.21**	**31.75**	**41.53**	**30.04**	**40.88**

Table 7. Performance with Mistral-7B measured by SR-S and SR-AS in MAC (3898 edits), SMAC (3898 edits) on PEAK-CF and MQC (992 edits), SMQC (992 edits) on PEAK-T.

METHOD	PEAK-CF				PEAK-T			
	MAC		SMAC		MQC		SMQC	
	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑
WISE	33.91	28.22	0.10	0.00	23.39	19.15	0.00	0.00
MELO	15.24	12.67	13.80	11.49	5.44	4.44	2.42	2.02
ARK	**40.84**	**34.02**	**35.73**	**29.86**	**38.31**	**32.06**	**23.99**	**21.98**

Table 8. Performance with Mistral-7B measured by SR-S and with SR-AS in MQC (1948 edits), SMQC (1948 edits) on PEAK-CF and MAC (1984 edits), SMAC (1984 edits) on PEAK-T.

METHOD	PEAK-CF				PEAK-T			
	MQC		SMQC		MAC		SMAC	
	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑	SR-S↑	SR-AS↑
WISE	38.96	32.44	0.00	0.00	19.25	16.43	0.00	0.00
MELO	16.94	14.12	7.60	6.30	7.86	6.65	5.75	4.84
ARK	**43.94**	**36.76**	**22.33**	**18.58**	**34.80**	**30.04**	**31.75**	**27.42**

Table 9. The comparison of GSR, LSR between ARK without l_o and ARK in non-compositional ME with 100 edits of PEAK-CF, and the comparison of PSR between ARK without l_o and ARK in non-compositional ME with 100 edits of MQuAKE.

METHOD	GSR↑	LSR↑	PSR↑
ARK-l_o	82.00	54.21	53.23
ARK	**82.00**	**54.53**	**54.52**

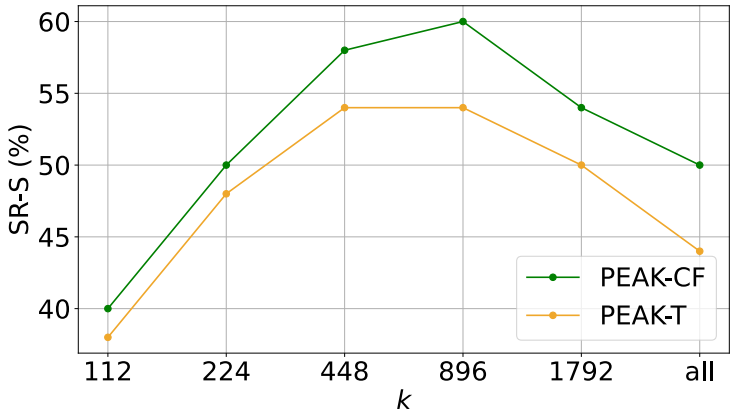
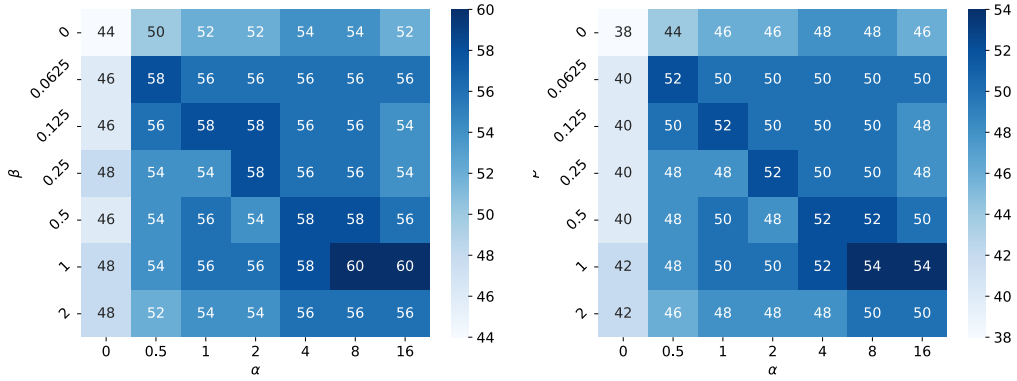


Figure 1. The sensitivity analysis for the hyperparameter k on PEAK-CF and PEAK-T. The variation trend of SR-S with respect to k is similar in different datasets. k is not sensitive across different datasets.



(a) PEAK-CF (b) PEAK-T

Figure 2. The sensitivity analysis for the hyperparameters α and β on PEAK-CF and PEAK-T. ARK is not sensitive to the selection of α and β. The variation trends of SR-S with respect to α and β are similar in different datasets. α and β are not sensitive across different datasets. Setting α to be 8 times β often leads to the best performance.

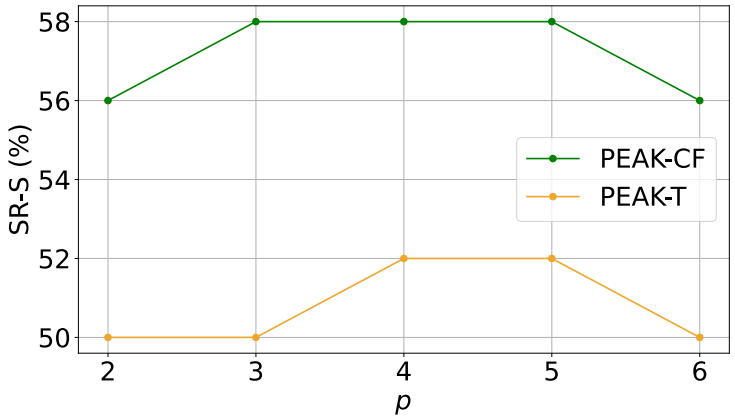


Figure 3. The sensitivity analysis for the hyperparameter p on PEAK-CF and PEAK-T. ARK is not sensitive to the selection of p. The variation trend of SR-S with respect to p is similar in different datasets. p is not sensitive across different datasets.

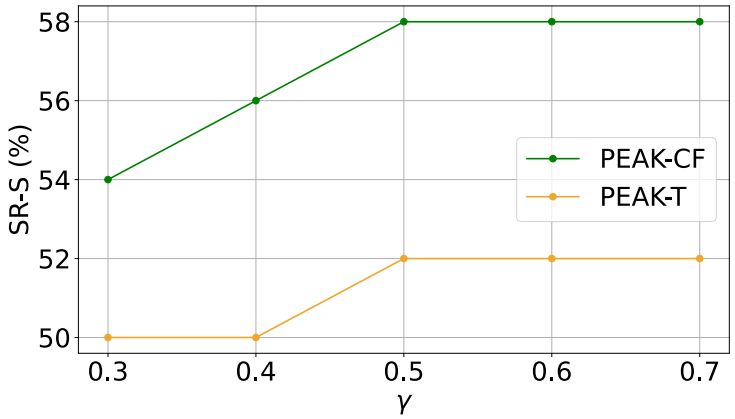


Figure 4. The sensitivity analysis for the hyperparameter γ on PEAK-CF and PEAK-T. ARK is not sensitive to the selection of γ. The variation trend of SR-S with respect to γ is similar in different datasets. γ is not sensitive across different datasets.

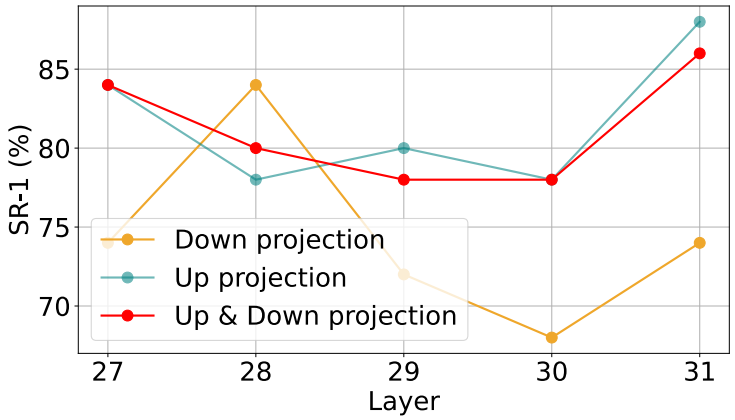


Figure 5. Performance of non-compositional ME measured by *SR-1* when editing only $\mathbf{W}_{\text{down}}$, only $\mathbf{W}_{\text{up}}$, both $\mathbf{W}_{\text{up}}$ and $\mathbf{W}_{\text{down}}$ in different layers.

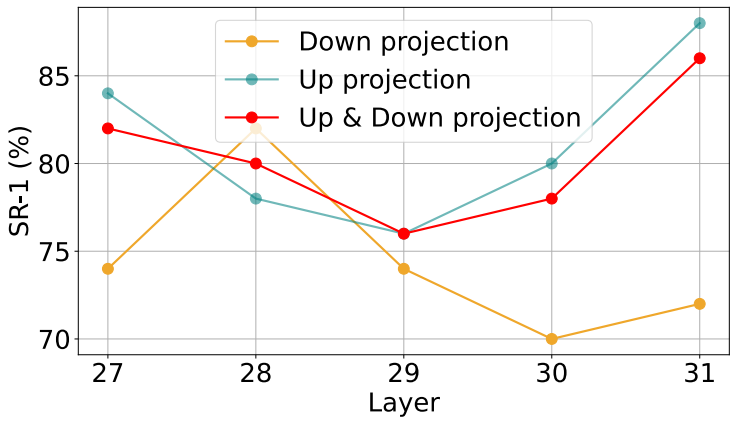


Figure 6. Performance of non-compositional ME measured by *SR-1* when editing only $\mathbf{W}_{\text{down}}$, only $\mathbf{W}_{\text{up}}$, both $\mathbf{W}_{\text{up}}$ and $\mathbf{W}_{\text{down}}$ in different layers with fake relations and answers.

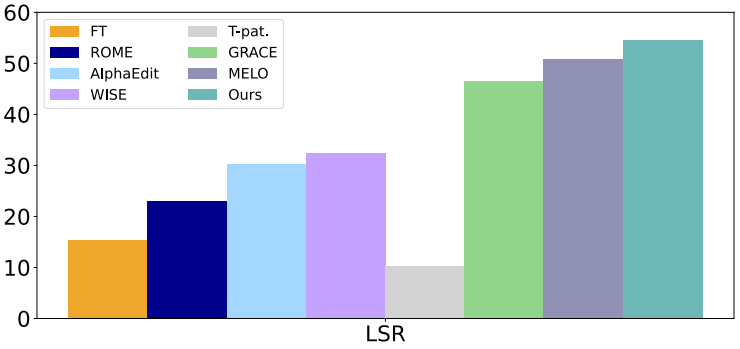


Figure 7. The comparison of *LSR* between ARK and baselines in non-compositional ME with 100 edits of PEAK-CF.

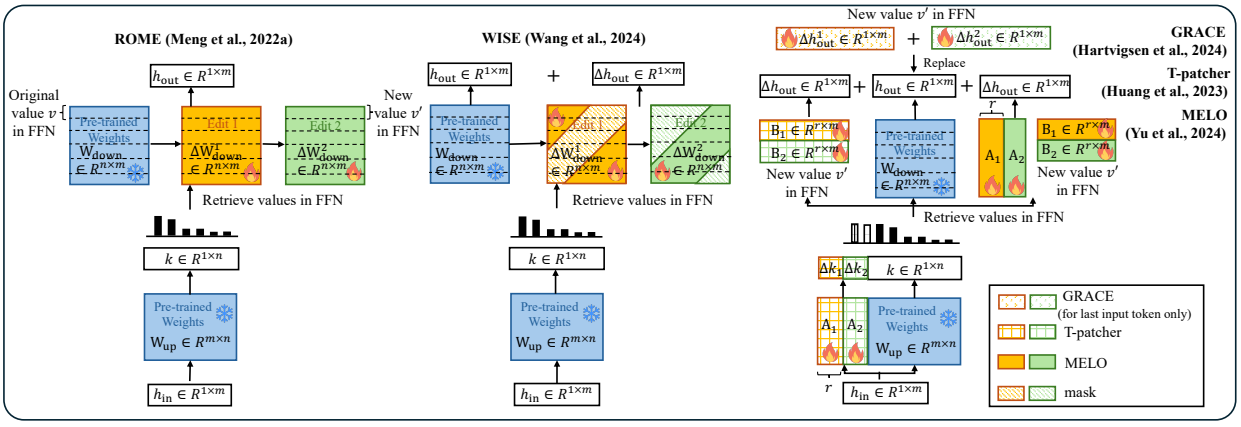


Figure 8. The visualization of different baselines. ROME and WISE overwrite the value v in $\mathbf{W}_{\text{down}}$ at the original key k while GRACE, T-patcher, and MELO add new values v for $\mathbf{W}_{\text{down}}$. ROME and WISE edit the $\mathbf{W}_{\text{down}}$ with/without gradient mask. GRACE, T-patcher, and MELO edit the $\mathbf{W}_{\text{down}}$ in different low-rank forms. Note that GRACE edits only the last input token, while others edit all tokens.

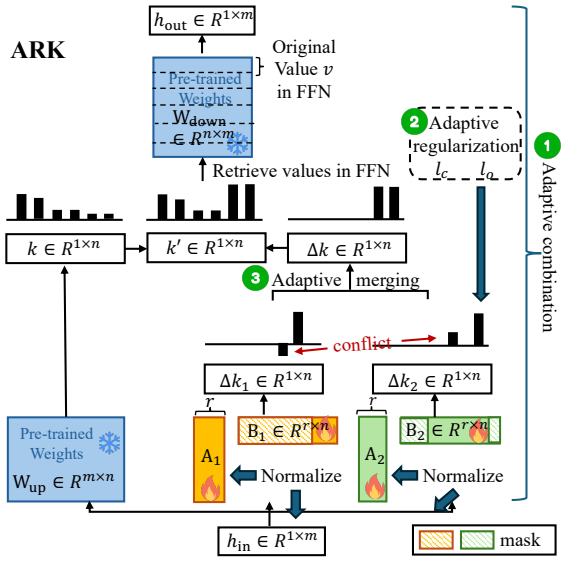


Figure 9. Our proposed ARK, including adaptive combination (Sec. 4.1), regularization (Sec. 4.2) and merging (Sec. 4.3). ARK redirects the specific input to the new key k' by adjusting $\mathbf{W}_{\text{up}}$, in order to activate its correspondind value v' retained in $\mathbf{W}_{\text{down}}$ for the edit.