Table 1. Comparison of performance improvement of HFAT applied to five different baselines on the CIFAR-10 dataset and implemented them on the PreAct ResNet-18 and WideResNet34-10 architectures. For testing the attack methods, we selected FGSM, PGD_{20}, PGD_{100}, CW, MIM, AA_{rand}, and AA. Here, AA refers to the standard version of AutoAttack.

	PreAct ResNet-18								WideResNet34-10							
	Natural	FGSM	PGD_{20}	PGD_{100}	CW	MIM	AA_{rand}	AA	Natural	FGSM	PGD_{20}	PGD_{100}	CW	MIM	AA_{rand}	AA
AT_{PGD}	81.66	57.51	52.64	52.51	50.29	52.83	49.04	48.25	86.40	61.83	55.60	55.19	54.59	55.70	52.95	52.06
AT_{HF}	**81.88**	**60.04**	**56.39**	**56.23**	**52.60**	**56.46**	**51.64**	**50.61**	**87.53**	**65.76**	**60.06**	**59.87**	**57.64**	**60.18**	**56.55**	**55.58**
TRADES	**81.24**	59.24	55.71	55.37	50.45	55.63	49.95	49.20	83.68	61.78	59.31	59.25	54.29	59.31	54.02	53.46
$TRADES_{HF}$	80.39	**59.61**	**57.41**	**57.12**	**51.20**	**56.95**	**50.96**	**50.35**	**85.38**	**63.80**	**61.12**	**61.03**	**55.88**	**61.16**	**55.62**	**55.05**
MART	80.63	59.54	56.16	55.86	50.31	55.41	50.47	49.62	83.98	61.32	58.43	58.12	54.74	58.07	53.64	52.36
$MART_{HF}$	**81.14**	**59.73**	**57.24**	**56.97**	**51.11**	**56.36**	**51.61**	**50.74**	**84.76**	**64.03**	**61.63**	**61.47**	**56.18**	**60.73**	**55.23**	**54.77**
AWP	80.81	59.38	55.59	55.47	51.89	55.70	51.04	50.06	85.65	62.75	58.82	58.69	55.56	59.24	55.39	53.61
AWP_{HF}	**81.17**	**59.83**	**55.95**	**55.87**	**52.31**	**56.27**	**51.82**	**50.28**	**86.41**	**64.18**	**62.23**	**62.06**	**57.42**	**60.94**	**56.22**	**54.95**
HELP	80.75	59.57	56.41	56.13	52.34	56.18	50.63	49.76	83.69	62.63	59.48	59.11	55.82	60.02	55.40	53.98
$HELP_{HF}$	**81.27**	**60.04**	**57.82**	**57.50**	**52.91**	**57.05**	**51.24**	**50.31**	**85.21**	**64.29**	**62.54**	**62.21**	**57.73**	**61.25**	**56.71**	**55.21**

Tab 1. Results of SOTA in CVPR2024 [1] on CIFAR10.

Table 1. Test Accuracy and Robustness of the CIFAR-10 datasets on Wide-ResNet-3410. Both the best and last model are evaluated.

	Clean	PGD-10	PGD-20	PGD-50	C&W	AA	Ave.
PGD-AT	85.17	56.07	55.08	54.88	53.91	51.67	59.46
TRADES	85.72	56.75	56.10	55.90	53.87	53.40	60.28
MART	84.17	58.98	58.56	58.06	54.58	51.10	60.90
FAT	87.97	50.31	49.86	48.79	48.65	47.48	55.51
GAIRAT	86.30	60.64	59.54	58.74	45.57	40.30	58.51
AWP	85.57	58.92	58.13	57.92	56.03	53.90	61.74
LBGAT	88.22	56.25	54.66	54.30	54.29	52.23	59.99
LAS-AWP	57.74	61.39	60.16	59.79	58.22	**55.52**	58.80
Nasty-AT (best)	**89.15**	**63.69**	**62.34**	**62.05**	**65.10**	52.95	**65.88**
Nasty-AT (last)	87.33	65.01	63.66	63.03	63.40	50.23	65.44

Table 2. Test Accuracy and Robustness of the CIFAR-10 datasets on ResNet-18. Both the best and last model are evaluated.

	Clean	PGD-10	PGD-20	PGD-50	PGD-100	C&W	AA	Ave.
PGD-AT	84.25	46.88	46.56	44.85	44.76	45.75	41.69	50.67
MART	81.61	52.38	51.28	50.93	50.80	47.77	46.09	54.40
TRADES	83.64	52.05	50.67	50.38	50.20	49.68	48.41	55.00
FAT	87.32	45.80	43.53	43.11	42.98	43.50	40.76	49.57
LBGAT	85.73	53.12	52.05	51.78	51.68	50.63	49.04	56.29
CAS	86.24	51.38	51.49	51.77	51.04	53.66	46.69	56.03
AWP	79.45	55.04	54.47	54.36	54.30	51.17	49.40	56.88
LAS-AT	82.39	54.74	53.70	53.70	53.72	51.96	49.94	57.16
AGAIN-AWP	86.52	59.99	59.35	59.11	58.85	61.19	**51.89**	62.41
Nasty-AT(best)	**90.86**	**62.37**	**60.94**	**60.19**	**59.91**	**62.54**	50.18	**63.85**
Nasty-AT(last)	90.28	61.86	60.00	59.44	58.89	61.67	48.96	63.01

Tab 2. Results of NAT (ours) on CIFAR10.

[1] Li Q, Hu Y, Dong Y, et al. Focus on hiders: Exploring hidden threats for enhancing adversarial training[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 24442-24451.