

Table 3: **Comparison with invariance regularization-based defense methods on CIFAR datasets.** We report clean and robust accuracy (AutoAttack; AA). 12

Defense		CIFAR10			CIFAR100		
		Clean	AA	Sum.	Clean	AA	Sum.
ResNet-18	AT	83.77	42.42	126.19	55.82	19.50	75.32
	TRADES	81.25	48.54	129.79	54.74	23.60	78.34
	MART	82.15	47.83	129.98	54.54	26.04	80.58
	LBGAT	85.00 \pm 0.47	48.85 \pm 0.46	133.86 \pm 0.65	65.87 \pm 0.74	23.19 \pm 0.74	89.07 \pm 0.73
	ARREST*	<u>86.63</u>	46.14	<u>132.77</u>	-	-	-
	AR-AT (ours)	87.82 \pm 0.19	<u>49.02 \pm 0.47</u>	136.84 \pm 0.33	67.51 \pm 0.13	23.38 \pm 0.19	<u>90.89 \pm 0.29</u>
	AR-AT+SWA (ours)	86.44 \pm 0.05	50.28 \pm 0.14	<u>136.72 \pm 0.19</u>	<u>67.17 \pm 0.18</u>	<u>24.36 \pm 0.20</u>	91.53 \pm 0.25
WRN-34-10	AT	86.06	46.26	132.32	59.83	23.94	83.77
	TRADES	84.33	<u>51.75</u>	136.08	57.61	26.88	84.49
	MART	86.10	49.11	135.21	57.75	24.89	82.64
	LBGAT	88.19 \pm 0.11	<u>52.56 \pm 0.34</u>	140.75 \pm 0.34	68.17 \pm 0.56	<u>26.92 \pm 0.32</u>	95.09 \pm 0.64
	ARREST*	<u>90.24</u>	50.20	140.44	73.05	24.32	<u>97.37</u>
	AR-AT (ours)	90.89 \pm 0.22	50.77 \pm 0.50	<u>141.66 \pm 0.51</u>	<u>72.51 \pm 0.51</u>	24.18 \pm 0.46	96.70 \pm 0.56
	AR-AT+SWA (ours)	<u>90.06 \pm 0.08</u>	54.03 \pm 0.31	144.09 \pm 0.39	72.37 \pm 0.13	27.31 \pm 0.13	99.69 \pm 0.26