TABLE VI: The certified robust accuracy and certified robustness rate of different approaches on various datasets within smaller vicinity.

Approach	Certified Robust Accuracy			
	CIFAR-100	CIFAR-10	SVHN	MNIST
ERM	33.45	48.85	59.34	48.01
DA	54.43	83.50	84.79	81.23
PGDT	44.59	83.23	87.98	95.89
TRADES	58.86	80.57	82.45	95.39
MART	56.73	81.35	73.84	95.22
RS	53.93	88.98	86.03	90.48
IBP	33.45	54.41	67.34	97.74
PRL	53.99	91.74	91.97	98.99
Ours	**57.27**	**93.58**	**92.85**	**97.15**

$\kappa = 10^{-2}, 1-\alpha = 0.99$; L^∞ bound at 0.1 for MNIST, and 2/255 for CIFAR and SVHN. See Table I for L^∞ bound at 0.3 for MNIST, and 8/255 for CIFAR and SVHN.

TABLE VII: Comparison of the influence of different κ values on the certified robust accuracy of CIFAR-10.

Approach	$\kappa = 0$ (Deterministic)	$\kappa = 10^{-3}$	$\kappa = 10^{-2}$	$\kappa = 10^{-1}$
ERM	-	1.25	1.25	25.09
DA	-	73.50	76.07	86.59
PGDT	-	82.82	82.90	82.95
TRADES	-	78.69	78.80	79.60
MART	-	71.42	72.21	73.43
RS	-	87.63	87.98	88.08
IBP	35.13	39.98	40.00	44.41
PRL	-	89.88	90.63	91.97
Ours	-	91.73	91.75	92.78

For $\kappa > 0$, α takes 10^{-2}.

APPENDIX

We carry out an ablation study to assess the effect of the hyper-parameters in our method.

Vicinity size ϵ. To investigate the impact of the vicinity size on certified robust accuracy, we evaluate the models with altered L^∞-norm radius ϵ on each dataset. Specifically, for MNIST, values of ϵ are selected from $\{0.1, 0.3\}$, while for the other three datasets, its values are chosen from $\{2/255, 8/255\}$. The results are shown in Table VI. We observe a trade-off between certified robust accuracy and the usefulness of certification, i.e., decreasing the vicinity radius increases certified robust accuracy. Our approach achieves high certified robust accuracy ($> 85\%$) within a reasonable range of the vicinity and experiences a 0.36% increase with a one-third reduction and a 2.98% average increase with a one-quarter reduction.

Percentage to certify. To investigate how the strictness of certification requirement influences the certified robust accuracy, we vary the acceptable level κ and significance level α. The certified robust accuracy with regard to different acceptable levels and significance levels is presented in Table VII and Table VIII, respectively. Note that $\kappa = 0$ means conducting deterministic robustness certification on the model, which can only be achieved by IBP. The remaining baselines and our method can only provide probabilistic robustness certification results for the model. It can be observed that the variation of both the acceptable level κ and significance level α does not have a significant impact on the certified robust accuracy, except ERM and DA. Specifically, for our method, when κ has changed from 10^{-3} to 10^{-1}, the certified robust accuracy has only improved by 1.05%; no increase in certified robust accuracy is observed when α varies from 10^{-3} to 5×10^{-2}.

TABLE VIII: Comparison of the influence of different α values on the certified robust accuracy of CIFAR-10 where $\kappa = 10^{-2}$.

Approach	$1-\alpha = 0.95$	$1-\alpha = 0.99$	$1-\alpha = 0.999$
ERM	2.55	1.25	1.25
DA	77.56	76.07	76.07
PGDT	82.90	82.90	82.90
TRADES	78.80	78.80	78.80
MART	72.21	72.21	72.21
RS	87.98	87.98	87.98
IBP	40.00	40.00	40.00
PRL	90.63	90.63	90.63
Ours	91.75	91.75	91.75