Table I. The classification accuracy (mean±std %) of each comparing algorithm on corrupted benchmark dataset of CIFAR-10 (The backbone is instantiated with ResNet-18). The number of labeled instances per class is set to $v = 400$. The number of false positive labels is set to $r \in \{3,5,7\}$. The best results among methods are highlighted in bold.

Dataset	Method	$r = 3$	$r = 5$	$r = 7$
CIFAR-10 ($v = 400$)	ConCont	$88.74 \pm 0.14\%$	$88.32 \pm 0.17\%$	$80.16 \pm 0.38\%$
	SPMI	$92.54 \pm 0.22\%$	$90.72 \pm 0.28\%$	$82.13 \pm 0.19\%$
	FairMatch	$92.85 \pm 0.09\%$	$92.37 \pm 0.33\%$	$81.12 \pm 0.17\%$
	Ours	$\mathbf{94.32 \pm 0.16\%}$	$\mathbf{93.26 \pm 0.21\%}$	$\mathbf{86.11 \pm 0.33\%}$

Table II. The classification accuracy (mean±std %) of each comparing algorithm on corrupted benchmark dataset of CIFAR-100 (The backbone is instantiated with ResNet-18). The number of labeled instances per class is set to $v = 100$. The number of false positive labels is set to $r \in \{5,10,15,20\}$. The best results among methods are highlighted in bold.

Dataset	Method	$r = 5$	$r = 10$	$r = 15$	$r = 20$
CIFAR-100 ($v = 100$)	ConCont	$62.71 \pm 0.11\%$	$61.89 \pm 0.23\%$	$56.32 \pm 0.17\%$	$50.28 \pm 0.37\%$
	SPMI	$60.19 \pm 0.45\%$	$58.87 \pm 0.33\%$	$54.94 \pm 0.53\%$	$47.98 \pm 0.43\%$
	FairMatch	$63.74 \pm 0.32\%$	$61.12 \pm 0.18\%$	$57.55 \pm 0.43\%$	$53.69 \pm 0.31\%$
	Ours	$\mathbf{71.15 \pm 0.12\%}$	$\mathbf{69.76 \pm 0.19\%}$	$\mathbf{64.24 \pm 0.41\%}$	$\mathbf{60.81 \pm 0.26\%}$