Supplementary Materials

1. Dataset

the detailed characteristics of these datasets are shown in Table 1.

Table 1. Summary of datasets used in the experiments.

	# of training	# of test	# of class	size
MNIST	60,000	10,000	10	28×28
CIFAR-10	50,000	10,000	10	32×32
CIFAR-100	50,000	10,000	100	32×32
Clothing1M	1,000,000	10,000	14	224×224

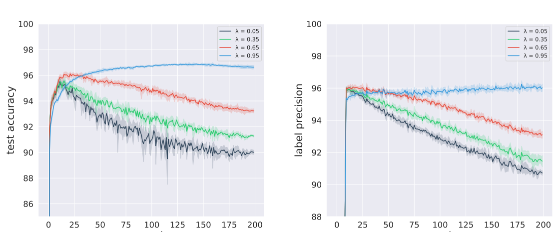
2. Network Architecture

The network architectures of the MLP and CNN models are shown in Table 2.

Table 2. The models used on *MNIST*, *CIFAR-10* and *CIFAR-100*

MLP on *MNIST*	CNN on *CIFAR-10*& *CIFAR-100*
28×28 Gray Image	32×32 RGB Image
Dense $28 \times 28 \longrightarrow 256$, ReLU	3×3, 64 BN, ReLU 3×3, 64 BN, ReLU 2×2 Max-pool
	3×3, 128 BN, ReLU 3×3, 128 BN, ReLU 2×2 Max-pool
	3×3, 196 BN, ReLU 3×3, 196 BN, ReLU 2×2 Max-pool
Dense $256 \longrightarrow 10$	Dense $256 \longrightarrow 100$

3. Parameter Sensitivity Analysis



(a) test accuracy (b) label precision

Figure 1. Results of JoCoR with different λ on *MNIST*

To conduct the sensitivity analysis on parameter λ, we set up the experiments above *MNIST* with symmetry-50% noise. Specifically, we compare λ in the range of [0.05, 0.35, 0.65, 0.95]. The larger the λ is, the less the divergence of two classifiers in JoCoR would be.

The test accuracy and label precision vs. number of epochs are in Figure 1. Obviously, As the λ increases, the performance of our algorithm on test accuracy gets better and better. When $\lambda = 0.95$, JoCoR achieves the best performance. We can see the same trends on label precision, which means that JoCoR can select clean example more precisely with a larger λ.