

Supplementary Materials

1 Hyperparameters for Pseudo-Label

All hyperparameters were chosen based on validation result. When tuning hyperparameters we set the data seed to 0 in all cases. Data seed only affects the split of labelled training data and unlabelled training data, it does not change the validation set or the test set. Table 1 and Table 2 show the hyperparameters used for Pseudo-Label in experiments of section 4. Table 3 and Table 4 show the hyperparameters used for Pseudo-Label in experiments of section 5.

Table 1: Hyperparameters used for Pseudo-Label when unlabelled data is clean.

Hyperparameter	MNIST	Fashion-MNIST	CIFAR-10
a	1	5	1
lr	0.1	0.1	0.001
γ	0.99	0.9	0.9

Table 2: Hyperparameters used for Pseudo-Label when unlabelled data is dirty.

Hyperparameter	MNIST	Fashion-MNIST	CIFAR-10
a	1	5	1
lr	0.1	0.1	0.001
γ	0.99	0.9	0.9

Table 3: Hyperparameters used for Pseudo-Label with weights calculated from raw images when unlabelled data is dirty.

Hyperparameter	MNIST	Fashion-MNIST	CIFAR-10
a	1	0.5	1
lr	0.1	0.1	0.001
γ	0.99	0.9	0.9
β	10	10	20

Table 4: Hyperparameters used for Pseudo-Label with weights calculated from encodings when unlabelled data is dirty.

Hyperparameter	MNIST	Fashion-MNIST	CIFAR-10
a	1	0.5	1
lr	0.1	0.1	0.001
γ	0.99	0.9	0.9
β	10	5	5

2 Hyperparameters for Mean Teacher

All hyperparameters were chosen based on validation result. When tuning hyperparameters we set the data seed to 0 in all cases. Data seed only affects the split of labelled training data and unlabelled training data, it does not change the validation set or the test set. Table 1 and Table 2 show the hyperparameters used for Mean Teacher in experiments of section 4. Table 3 and Table 4 show the hyperparameters used for Mean Teacher in experiments of section 5.

Table 5: Hyperparameters used for Mean Teacher when unlabelled data is clean.

Hyperparameter	MNIST	Fashion-MNIST	CIFAR-10
a	5	50	100
T	10	10	10
lr	0.3	0.1	0.001
γ	0.999	0.99	0.9
η	0.9	0.9	0.995

Table 6: Hyperparameters used for Mean Teacher when unlabelled data is dirty.

Hyperparameter	MNIST	Fashion-MNIST	CIFAR-10
a	5	50	50
T	10	10	10
lr	0.3	0.1	0.001
γ	0.999	0.99	0.9
η	0.9	0.9	0.995

Table 7: Hyperparameters used for Mean Teacher with weights calculated from raw images when unlabelled data is dirty.

Hyperparameter	MNIST	Fashion-MNIST	CIFAR-10
a	5	1	100
T	10	10	10
lr	0.3	0.1	0.001
γ	0.999	0.99	0.9
η	0.9	0.9	0.999
β	10	10	2

Table 8: Hyperparameters used for Mean Teacher with weights calculated from encodings when unlabelled data is dirty.

Hyperparameter	MNIST	Fashion-MNIST	CIFAR-10
a	50	50	100
T	10	10	10
lr	0.3	0.1	0.001
γ	0.999	0.99	0.9
η	0.9	0.9	0.999
β	10	3	2