

SUPPLEMENTARY MATERIAL FOR HYPERCORE: CORE-SET SELECTION UNDER NOISE VIA HYPERSPHERE MODELS

Anonymous authors

Paper under double-blind review

A HARDWARE AND SOFTWARE

All experiments were run on a workstation equipped with an NVIDIA RTX A6000 GPU (48 GB VRAM). Our implementation uses PyTorch 1.10.1 with torchvision 0.11.2, and we build upon the DeepCore library for coreset selection routines.

B DATASETS

We evaluated HyperCore on two widely used image-classification benchmarks:

- **CIFAR-10**: 50,000 training and 10,000 test images across 10 classes, each of size 32×32 pixels. Its small, balanced format makes it ideal for studying the behavior of coreset methods at extreme pruning levels.
- **ImageNet-1K**: 1,281,167 training and 50,000 validation images spanning 1,000 object categories. Following standard practice, we resize all images to 224×224 pixels before selection and training.

C ADDITIONAL CROSS-ARCHITECTURE EVALUATIONS

Table 1: Coreset selection performances on CIFAR-10 with five randomly initialized VGG-16 models. The best results are marked in bold.

Fraction ($1 - \alpha$)	0.1%	0.5%	1%	5%	10%	20%	30%	40%	50%	60%	90%	100%
Herding	11.6 $\pm$ 1.2	12.7 $\pm$ 0.6	16.0 $\pm$ 3.9	54.1 $\pm$ 3.1	66.0 $\pm$ 3.5	71.5 $\pm$ 1.5	78.3 $\pm$ 1.0	80.9 $\pm$ 0.5	85.7 $\pm$ 1.5	88.2 $\pm$ 0.7	93.2 $\pm$ 0.3	94.3 $\pm$ 0.0
k-Center Greedy	12.3 $\pm$ 1.0	14.5 $\pm$ 2.9	14.7 $\pm$ 1.3	38.7 $\pm$ 12.0	75.6$\pm$1.3	85.5 $\pm$ 0.3	89.0 $\pm$ 0.4	91.0 $\pm$ 0.2	91.8 $\pm$ 0.3	92.7 $\pm$ 0.2	94.1 $\pm$ 0.1	94.3 $\pm$ 0.0
Forgetting	12.5 $\pm$ 0.8	13.8 $\pm$ 2.8	25.6$\pm$4.3	55.9 $\pm$ 1.3	75.2 $\pm$ 1.4	86.2$\pm$0.2	89.0 $\pm$ 0.2	91.3 $\pm$ 0.2	91.9 $\pm$ 0.2	92.8 $\pm$ 0.0	94.1 $\pm$ 0.2	94.3 $\pm$ 0.0
GraNd	11.7 $\pm$ 0.6	12.5 $\pm$ 0.0	16.6 $\pm$ 3.0	28.1 $\pm$ 0.0	42.4 $\pm$ 0.9	73.4 $\pm$ 0.7	86.5 $\pm$ 0.7	91.2 $\pm$ 0.0	93.1 $\pm$ 0.4	93.6$\pm$0.2	94.1 $\pm$ 0.2	94.3 $\pm$ 0.0
CAL	12.9 $\pm$ 2.0	15.8 $\pm$ 4.1	19.2 $\pm$ 4.5	63.0$\pm$1.4	72.2 $\pm$ 0.2	78.1 $\pm$ 0.9	82.6 $\pm$ 0.5	85.0 $\pm$ 0.2	87.6 $\pm$ 0.7	89.6 $\pm$ 0.4	93.0 $\pm$ 0.2	94.3 $\pm$ 0.0
DeepFool	11.6 $\pm$ 0.7	14.6 $\pm$ 1.3	15.6 $\pm$ 4.1	47.3 $\pm$ 6.8	71.7 $\pm$ 1.1	84.2 $\pm$ 0.4	90.0$\pm$0.2	91.6$\pm$0.1	92.4 $\pm$ 0.2	93.2 $\pm$ 0.2	94.1 $\pm$ 0.1	94.3 $\pm$ 0.0
Craig	13.3 $\pm$ 1.4	14.2 $\pm$ 1.4	13.6 $\pm$ 0.0	46.6 $\pm$ 0.0	67.3 $\pm$ 3.1	82.4 $\pm$ 0.6	88.8 $\pm$ 0.3	91.4 $\pm$ 0.0	92.7 $\pm$ 0.2	93.4 $\pm$ 0.1	94.2$\pm$0.1	94.3 $\pm$ 0.0
GradMatch	13.4 $\pm$ 2.0	13.0 $\pm$ 2.3	13.0 $\pm$ 0.0	44.6 $\pm$ 3.7	53.9 $\pm$ 1.4	77.9 $\pm$ 0.4	87.2 $\pm$ 0.3	90.2 $\pm$ 0.6	92.6$\pm$0.0	93.4 $\pm$ 0.0	94.2$\pm$0.1	94.3 $\pm$ 0.0
Gliester	15.7$\pm$0.0	13.9 $\pm$ 3.0	18.9 $\pm$ 5.0	44.0 $\pm$ 4.0	53.5 $\pm$ 0.9	78.2 $\pm$ 2.2	86.4 $\pm$ 0.1	89.7 $\pm$ 0.0	92.5 $\pm$ 0.2	93.3 $\pm$ 0.2	94.2$\pm$0.1	94.3 $\pm$ 0.0
HyperCore (ours)	15.7$\pm$0.6	16.6$\pm$1.2	24.2 $\pm$ 2.2	59.7 $\pm$ 0.3	73.3 $\pm$ 0.4	83.5 $\pm$ 0.1	87.2 $\pm$ 0.2	89.3 $\pm$ 0.2	90.6 $\pm$ 0.2	91.5 $\pm$ 0.1	93.7 $\pm$ 0.1	94.3 $\pm$ 0.0

Table 2: Coreset selection performances on CIFAR-10 with five randomly initialized ResNet-50 models. The best results are marked in bold.

Fraction ($1 - \alpha$)	0.1%	0.5%	1%	5%	10%	20%	30%	40%	50%	60%	90%	100%
Herding	16.2 $\pm$ 2.4	23.0 $\pm$ 2.4	29.5 $\pm$ 2.8	46.3 $\pm$ 2.8	58.5 $\pm$ 5.2	73.1 $\pm$ 2.1	80.0 $\pm$ 0.9	84.1 $\pm$ 0.5	86.4 $\pm$ 0.9	89.6 $\pm$ 0.9	94.6 $\pm$ 0.3	95.4 $\pm$ 0.3
k-Center Greedy	15.4 $\pm$ 1.8	18.8 $\pm$ 3.1	25.0 $\pm$ 4.6	53.1 $\pm$ 1.9	68.8 $\pm$ 2.1	86.6$\pm$0.4	90.2 $\pm$ 0.6	91.7 $\pm$ 0.2	93.4 $\pm$ 0.3	94.2 $\pm$ 0.3	95.3 $\pm$ 0.3	95.4 $\pm$ 0.3
Forgetting	17.0 $\pm$ 2.2	26.9 $\pm$ 2.7	34.0 $\pm$ 1.0	49.5 $\pm$ 2.9	67.5 $\pm$ 2.6	86.2 $\pm$ 0.7	89.9 $\pm$ 0.4	92.0 $\pm$ 0.3	92.8 $\pm$ 0.2	93.6 $\pm$ 0.1	95.2 $\pm$ 0.1	95.4 $\pm$ 0.3
GraNd	15.1 $\pm$ 1.7	17.2 $\pm$ 1.4	19.7 $\pm$ 1.2	23.6 $\pm$ 1.6	34.9 $\pm$ 2.4	70.3 $\pm$ 5.0	85.5 $\pm$ 1.4	92.0 $\pm$ 0.3	94.2$\pm$0.4	94.9$\pm$0.4	95.4$\pm$0.2	95.4 $\pm$ 0.3
CAL	19.9 $\pm$ 1.3	27.5$\pm$0.6	35.9 $\pm$ 1.2	56.4$\pm$1.4	66.2 $\pm$ 1.1	77.6 $\pm$ 0.8	83.0 $\pm$ 0.4	85.5 $\pm$ 0.5	88.2 $\pm$ 0.5	90.5 $\pm$ 0.6	94.2 $\pm$ 0.2	95.4 $\pm$ 0.3
DeepFool	14.4 $\pm$ 2.0	19.8 $\pm$ 1.4	26.4 $\pm$ 1.1	42.7 $\pm$ 2.9	65.7 $\pm$ 2.8	86.3 $\pm$ 0.9	91.1$\pm$0.5	92.9$\pm$0.3	93.9 $\pm$ 0.2	94.5 $\pm$ 0.2	95.3 $\pm$ 0.2	95.4 $\pm$ 0.3
Craig	17.0 $\pm$ 1.1	21.5 $\pm$ 1.4	28.1 $\pm$ 2.6	42.0 $\pm$ 1.5	55.1 $\pm$ 2.7	80.8 $\pm$ 2.8	89.9 $\pm$ 0.7	92.5 $\pm$ 0.7	93.5 $\pm$ 0.2	94.7 $\pm$ 0.1	95.4$\pm$0.2	95.4 $\pm$ 0.3
GradMatch	14.3 $\pm$ 2.8	21.7 $\pm$ 0.9	28.2 $\pm$ 0.9	37.4 $\pm$ 2.6	50.5 $\pm$ 3.5	74.0 $\pm$ 2.3	87.6 $\pm$ 1.1	92.3 $\pm$ 0.4	93.8 $\pm$ 0.3	94.7 $\pm$ 0.5	95.2 $\pm$ 0.2	95.4 $\pm$ 0.3
Glister	14.9 $\pm$ 0.9	20.4 $\pm$ 1.5	24.2 $\pm$ 3.3	37.0 $\pm$ 1.9	50.1 $\pm$ 3.0	79.6 $\pm$ 2.5	88.6 $\pm$ 0.8	92.0 $\pm$ 0.3	93.5 $\pm$ 0.2	94.5 $\pm$ 0.3	95.2 $\pm$ 0.2	95.4 $\pm$ 0.3
HyperCore (ours)	20.3$\pm$1.8	23.0 $\pm$ 1.6	36.9$\pm$0.5	56.4$\pm$2.7	69.2$\pm$1.4	82.4 $\pm$ 1.0	88.0 $\pm$ 0.4	90.4 $\pm$ 0.2	91.6 $\pm$ 0.2	92.9 $\pm$ 0.3	94.8 $\pm$ 0.2	95.4 $\pm$ 0.3

Table 3: Coreset selection performances on CIFAR-10 with five randomly initialized InceptionNetV3 models. The best results are marked in bold.

Fraction ($1 - \alpha$)	0.1%	0.5%	1%	5%	10%	20%	30%	40%	50%	60%	90%	100%
Herding	14.6 $\pm$ 0.7	21.8 $\pm$ 1.7	28.5 $\pm$ 1.4	42.9 $\pm$ 4.0	61.1 $\pm$ 2.2	74.0 $\pm$ 1.3	80.5 $\pm$ 1.0	84.7 $\pm$ 0.4	87.7 $\pm$ 0.9	90.3 $\pm$ 0.8	94.8 $\pm$ 0.2	95.6 $\pm$ 0.1
k-Center Greedy	15.1 $\pm$ 2.0	22.1 $\pm$ 3.0	26.6 $\pm$ 1.4	51.2 $\pm$ 2.3	72.9$\pm$1.7	85.8 $\pm$ 0.4	89.8 $\pm$ 0.4	92.2 $\pm$ 0.3	93.6 $\pm$ 0.3	94.4 $\pm$ 0.4	95.4 $\pm$ 0.2	95.6 $\pm$ 0.1
Forgetting	18.5 $\pm$ 0.6	28.5 $\pm$ 1.0	32.2 $\pm$ 0.9	50.9 $\pm$ 1.5	69.3 $\pm$ 1.3	85.4$\pm$0.9	90.4$\pm$0.4	92.4 $\pm$ 0.1	93.4 $\pm$ 0.2	94.3 $\pm$ 0.3	95.4 $\pm$ 0.2	95.6 $\pm$ 0.1
GraNd	14.8 $\pm$ 2.3	17.4 $\pm$ 1.2	20.1 $\pm$ 0.8	27.1 $\pm$ 1.1	38.1 $\pm$ 1.4	70.0 $\pm$ 2.0	86.8 $\pm$ 0.6	92.9$\pm$0.4	94.4$\pm$0.3	95.0$\pm$0.3	95.5 $\pm$ 0.1	95.6 $\pm$ 0.1
CAL	17.9 $\pm$ 1.2	31.6 $\pm$ 1.5	36.3 $\pm$ 1.1	60.2$\pm$2.0	70.6 $\pm$ 0.6	78.7 $\pm$ 0.7	83.1 $\pm$ 0.5	86.3 $\pm$ 0.8	88.9 $\pm$ 0.4	90.3 $\pm$ 0.3	94.3 $\pm$ 0.1	95.6 $\pm$ 0.1
DeepFool	14.5 $\pm$ 1.6	22.3 $\pm$ 1.1	26.8 $\pm$ 1.6	47.7 $\pm$ 3.0	70.2 $\pm$ 2.1	83.6 $\pm$ 1.2	90.2 $\pm$ 0.4	92.8 $\pm$ 0.2	94.0 $\pm$ 0.2	94.6 $\pm$ 0.3	95.6$\pm$0.2	95.6 $\pm$ 0.1
Craig	16.5 $\pm$ 2.5	24.7 $\pm$ 1.3	27.8 $\pm$ 2.0	44.5 $\pm$ 1.8	58.8 $\pm$ 2.5	79.5 $\pm$ 1.4	88.0 $\pm$ 1.1	92.3 $\pm$ 0.2	94.0 $\pm$ 0.2	95.0$\pm$0.2	95.6$\pm$0.1	95.6 $\pm$ 0.1
GradMatch	16.5 $\pm$ 1.2	24.2 $\pm$ 0.8	29.1 $\pm$ 1.0	40.1 $\pm$ 2.7	53.1 $\pm$ 2.2	76.9 $\pm$ 2.6	87.1 $\pm$ 0.3	91.8 $\pm$ 0.7	94.2 $\pm$ 0.5	94.9 $\pm$ 0.3	95.4 $\pm$ 0.2	95.6 $\pm$ 0.1
Glister	14.8 $\pm$ 1.5	22.9 $\pm$ 1.8	29.3 $\pm$ 1.8	41.3 $\pm$ 1.8	52.4 $\pm$ 3.1	77.5 $\pm$ 1.7	87.8 $\pm$ 0.7	91.8 $\pm$ 0.4	94.0 $\pm$ 0.1	94.9 $\pm$ 0.1	95.6$\pm$0.1	95.6 $\pm$ 0.1
HyperCore (ours)	20.8$\pm$0.7	32.6$\pm$0.8	37.9$\pm$0.6	57.2 $\pm$ 1.6	71.0 $\pm$ 0.8	83.2 $\pm$ 0.9	87.5 $\pm$ 0.4	90.2 $\pm$ 0.2	92.4 $\pm$ 0.2	93.4 $\pm$ 0.3	95.2 $\pm$ 0.1	95.6 $\pm$ 0.1

Table 4: Coreset selection performances on CIFAR-10 with five randomly initialized WRN-16-8 models. The best results are marked in bold.

Fraction ($1 - \alpha$)	0.1%	0.5%	1%	5%	10%	20%	30%	40%	50%	60%	90%	100%
Herding	23.8 $\pm$ 1.4	31.3 $\pm$ 3.5	39.6 $\pm$ 2.8	60.2 $\pm$ 1.2	66.9 $\pm$ 2.3	77.3 $\pm$ 1.0	81.9 $\pm$ 0.7	86.4 $\pm$ 1.3	89.0 $\pm$ 0.6	92.3 $\pm$ 1.0	95.2 $\pm$ 0.1	96.0 $\pm$ 0.1
k-Center Greedy	19.2 $\pm$ 0.6	27.7 $\pm$ 0.6	34.6 $\pm$ 0.2	67.9 $\pm$ 0.8	81.6$\pm$0.5	89.4$\pm$0.2	92.1 $\pm$ 0.3	93.6 $\pm$ 0.2	94.3 $\pm$ 0.1	94.9 $\pm$ 0.1	95.9 $\pm$ 0.1	96.0 $\pm$ 0.1
Forgetting	21.5 $\pm$ 1.0	30.1 $\pm$ 0.9	36.3 $\pm$ 0.6	58.5 $\pm$ 1.0	74.9 $\pm$ 0.8	88.8 $\pm$ 0.5	93.3$\pm$0.1	94.5$\pm$0.1	95.0 $\pm$ 0.1	95.4 $\pm$ 0.1	95.9 $\pm$ 0.1	96.0 $\pm$ 0.1
GraNd	15.0 $\pm$ 1.8	19.5 $\pm$ 0.7	21.4 $\pm$ 0.3	37.8 $\pm$ 1.0	58.7 $\pm$ 1.0	81.8 $\pm$ 0.7	92.1 $\pm$ 0.1	94.3 $\pm$ 0.2	95.3$\pm$0.1	95.7$\pm$0.2	96.0$\pm$0.2	96.0 $\pm$ 0.1
CAL	22.2 $\pm$ 2.4	37.1 $\pm$ 2.5	45.5 $\pm$ 1.4	66.2 $\pm$ 0.6	74.2 $\pm$ 0.9	82.7 $\pm$ 0.5	86.3 $\pm$ 0.8	89.2 $\pm$ 0.6	91.0 $\pm$ 0.3	92.4 $\pm$ 0.2	95.3 $\pm$ 0.1	96.0 $\pm$ 0.1
DeepFool	18.9 $\pm$ 1.9	29.2 $\pm$ 1.0	35.0 $\pm$ 1.8	57.4 $\pm$ 3.2	74.0 $\pm$ 1.8	87.5 $\pm$ 0.4	92.0 $\pm$ 0.3	93.5 $\pm$ 0.2	94.6 $\pm$ 0.1	95.2 $\pm$ 0.1	96.0$\pm$0.1	96.0 $\pm$ 0.1
Craig	21.9 $\pm$ 1.5	30.0 $\pm$ 0.8	38.1 $\pm$ 1.4	60.1 $\pm$ 1.4	69.2 $\pm$ 0.8	86.5 $\pm$ 0.4	91.4 $\pm$ 0.2	93.8 $\pm$ 0.1	94.8 $\pm$ 0.1	95.4 $\pm$ 0.1	95.9 $\pm$ 0.1	96.0 $\pm$ 0.1
GradMatch	20.1 $\pm$ 1.8	27.8 $\pm$ 1.2	31.4 $\pm$ 2.1	54.4 $\pm$ 2.6	70.5 $\pm$ 2.0	84.6 $\pm$ 0.7	90.7 $\pm$ 0.5	93.3 $\pm$ 0.3	94.7 $\pm$ 0.4	95.3 $\pm$ 0.1	96.0$\pm$0.1	96.0 $\pm$ 0.1
Glister	20.1 $\pm$ 1.2	28.0 $\pm$ 1.4	31.6 $\pm$ 0.8	49.3 $\pm$ 3.3	67.9 $\pm$ 1.7	83.6 $\pm$ 0.8	90.5 $\pm$ 0.9	93.6 $\pm$ 0.1	94.6 $\pm$ 0.1	95.2 $\pm$ 0.2	96.0$\pm$0.2	96.0 $\pm$ 0.1
HyperCore (ours)	26.9$\pm$0.7	39.3$\pm$0.2	45.9$\pm$0.2	67.4$\pm$0.3	79.1 $\pm$ 0.3	86.5 $\pm$ 0.2	90.3 $\pm$ 0.1	92.2 $\pm$ 0.2	93.2 $\pm$ 0.2	94.0 $\pm$ 0.2	95.6 $\pm$ 0.1	96.0 $\pm$ 0.1

Table 5: Fixed coreset selection performance of five randomly initialized VGG-16 under label noise on CIFAR-10, where 10% of the training labels are randomly corrupted by assigning them to incorrect classes. Bold entries indicate the highest performance at each data fraction.

Fraction ($1 - \alpha$)	0.1%	0.5%	1%	5%	10%	20%	30%	40%	50%	60%	90%	100%
Herding	12.5±0.7	11.7±1.6	11.8±1.4	10.5±0.8	14.6±2.8	33.7±4.3	57.2±2.0	71.9±1.8	81.1±0.7	83.3±0.4	88.6±0.3	88.8±0.2
k-Center Greedy	12.0±1.4	12.2±1.3	12.6±1.2	17.0±4.7	40.6±3.8	68.2±1.7	76.0±0.6	79.6±0.5	82.6±0.3	85.1±0.3	88.3±0.2	88.8±0.2
Forgetting	12.6±0.9	14.4±1.2	21.5±5.0	55.8±1.0	72.5±2.0	83.0±0.8	86.0±0.8	87.8±0.2	88.1±0.4	88.2±0.3	88.6±0.1	88.8±0.2
GraNd	12.8±1.2	11.4±0.5	12.1±1.1	10.9±0.4	11.1±0.4	24.9±0.8	45.4±0.6	64.1±2.2	76.2±0.9	82.9±0.6	88.8±0.2	88.8±0.2
CAL	13.9±1.5	13.2±1.3	18.0±5.3	57.6±2.0	69.6±2.1	76.1±0.7	81.8±0.5	84.1±0.5	86.3±0.4	88.0±0.3	89.6±0.2	88.8±0.2
DeepFool	13.1±2.0	13.2±1.0	17.4±2.2	36.1±4.3	62.5±2.9	74.5±0.7	82.0±0.4	84.8±0.3	86.5±0.4	87.1±0.2	88.6±0.1	88.8±0.2
Craig	13.7±2.0	11.5±0.3	13.0±1.5	22.8±2.1	36.0±3.3	55.2±4.6	67.3±1.8	78.5±1.5	82.2±0.5	85.3±0.4	88.7±0.3	88.8±0.2
GradMatch	12.1±0.8	12.7±1.4	13.7±2.9	31.6±1.9	41.5±2.6	48.3±2.3	61.4±1.4	72.7±1.1	79.0±1.5	84.6±0.4	88.8±0.3	88.8±0.2
Glister	12.7±1.4	13.0±0.7	13.7±2.4	30.4±1.0	42.6±3.2	49.3±3.8	61.5±1.0	71.1±2.3	79.3±1.5	84.5±0.4	88.7±0.3	88.8±0.2
HyperCore (ours)	14.0±0.6	21.9±3.4	26.0±6.1	62.6±1.9	70.7±3.4	83.4±0.3	87.3±0.3	89.5±0.3	91.0±0.3	92.1±0.1	92.2±0.1	88.8±0.2

Table 6: Fixed coreset selection performance of five randomly initialized ResNet-50 under label noise on CIFAR-10, where 10% of the training labels are randomly corrupted by assigning them to incorrect classes. Bold entries indicate the highest performance at each data fraction.

Fraction ($1 - \alpha$)	0.1%	0.5%	1%	5%	10%	20%	30%	40%	50%	60%	90%	100%
Herding	11.4±1.1	12.4±1.9	10.6±0.7	11.4±0.8	11.3±0.7	36.5±3.8	56.6±2.1	71.5±2.6	79.3±1.0	84.5±0.8	89.7±0.3	89.8±0.4
k-Center Greedy	11.0±0.6	14.0±0.9	14.8±1.8	27.7±1.4	42.5±3.1	64.6±2.9	75.7±1.7	81.6±0.7	84.6±0.6	86.1±0.4	89.4±0.4	89.8±0.4
Forgetting	17.6±3.3	26.0±4.1	32.6±2.1	53.8±2.5	68.9±3.2	82.4±0.8	87.1±0.7	88.5±0.3	89.9±0.3	90.2±0.3	90.1±0.4	89.8±0.4
GraNd	13.5±1.3	13.5±0.8	12.8±1.7	11.3±1.1	12.0±1.2	22.4±1.3	44.8±1.5	62.9±1.4	77.1±0.2	84.6±1.0	89.5±0.4	89.8±0.4
CAL	16.9±4.1	27.0±2.2	34.5±2.4	52.2±2.9	65.2±0.3	76.8±1.2	81.7±1.0	84.5±0.5	86.9±0.4	88.5±0.5	90.5±0.1	89.8±0.4
DeepFool	16.0±3.4	19.8±1.2	24.5±1.6	38.0±1.2	52.4±4.1	75.0±2.6	84.1±0.5	86.2±0.5	87.8±0.5	88.7±0.4	90.1±0.2	89.8±0.4
Craig	17.0±1.0	15.5±2.3	20.6±2.1	31.9±0.9	33.8±3.0	46.7±2.4	65.9±3.5	75.6±2.1	83.3±0.3	85.6±0.3	89.7±0.2	89.8±0.4
GradMatch	13.9±1.7	17.7±1.8	24.1±2.0	29.2±0.5	35.7±2.7	46.9±6.1	59.3±2.6	73.1±1.1	80.9±1.0	84.4±0.3	89.6±0.4	89.8±0.4
Glister	13.6±0.9	18.9±1.9	22.0±2.1	28.6±2.0	34.5±2.1	47.4±4.1	58.6±0.9	71.5±2.6	80.3±0.6	84.6±1.2	89.2±0.5	89.8±0.4
HyperCore (ours)	20.3±1.4	28.8±1.7	36.4±1.6	56.1±2.4	68.5±0.7	83.5±0.7	87.8±0.3	90.7±0.3	92.1±0.1	93.1±0.2	93.2±0.3	89.8±0.4

Table 7: Fixed coreset selection performance of five randomly initialized InceptionNetV3 under label noise on CIFAR-10, where 10% of the training labels are randomly corrupted by assigning them to incorrect classes. Bold entries indicate the highest performance at each data fraction.

Fraction ($1 - \alpha$)	0.1%	0.5%	1%	5%	10%	20%	30%	40%	50%	60%	90%	100%
Herding	12.4±1.2	12.1±0.6	10.8±0.7	10.1±0.2	10.6±0.6	34.5±4.0	54.5±2.3	69.3±1.8	77.9±1.0	83.8±1.1	90.9±0.4	90.9±0.4
	-2.2	-9.7	-17.7	-32.9	-50.5	-39.5	-26.1	-15.4	-9.9	-6.5	-3.9	-4.6
k-Center Greedy	12.7±1.8	13.6±0.8	14.2±0.8	27.6±1.9	45.7±3.8	66.1±1.4	73.1±1.0	81.0±0.5	83.7±0.8	86.4±0.7	90.2±0.3	90.9±0.4
	-2.4	-8.5	-12.4	-23.6	-27.1	-19.7	-16.6	-11.2	-9.9	-7.9	-5.2	-4.6
Forgetting	17.1±2.4	28.6±0.7	33.1±1.7	50.0±1.0	71.2±0.8	83.1±1.2	87.3±0.7	89.4±0.1	90.2±0.3	90.6±0.3	90.8±0.3	90.9±0.4
	-1.4	0.0	+0.9	-1.0	+1.9	-2.3	-3.1	-2.9	-3.1	-3.7	-4.6	-4.6
GraNd	12.9±1.0	12.5±1.0	11.2±0.6	10.9±0.5	11.3±0.3	25.0±1.0	44.9±1.0	61.8±1.9	77.0±0.9	83.3±0.9	90.6±0.3	90.9±0.4
	-1.9	-4.9	-8.9	-16.2	-26.8	-45.0	-41.9	-31.0	-17.4	-11.6	-5.0	-4.6
CAL	15.9±1.5	26.8±1.9	30.9±1.6	56.3±3.5	67.8±2.7	77.0±0.7	82.3±0.5	84.5±0.5	87.1±0.3	89.1±0.4	91.5±0.3	90.9±0.4
	-2.1	-4.8	-5.4	-3.9	-2.7	-1.7	-0.7	-1.8	-1.8	-1.2	-2.8	-4.6
DeepFool	17.0±1.0	19.9±1.5	24.7±2.2	38.1±1.0	54.7±5.2	76.1±0.7	83.2±0.6	87.2±0.5	89.1±0.3	89.9±0.4	91.3±0.2	90.9±0.4
	+2.4	-2.4	-2.1	-9.7	-15.5	-7.5	-7.0	-5.5	-5.0	-4.7	-4.2	-4.6
Craig	17.8±2.5	21.0±1.1	23.4±1.4	29.8±2.0	36.2±3.2	45.0±3.5	58.6±2.2	71.4±1.1	78.5±1.1	84.3±0.8	90.7±0.4	90.9±0.4
	+1.3	-3.7	-4.4	-14.7	-22.6	-34.4	-29.4	-20.9	-15.5	-10.6	-4.9	-4.6
GradMatch	14.4±1.6	21.5±1.0	22.2±1.0	29.3±2.4	38.3±3.3	42.7±3.3	57.4±0.9	68.3±0.8	75.3±2.0	83.1±0.8	90.8±0.2	90.9±0.4
	-2.1	-2.7	-6.9	-10.8	-14.7	-34.2	-29.7	-23.5	-19.0	-11.8	-5.0	-4.6
Glister	15.0±1.4	20.2±1.4	22.9±0.7	32.3±1.7	38.3±3.4	43.1±2.3	54.8±3.6	67.4±2.9	77.0±1.0	82.9±0.6	90.9±0.2	90.9±0.4
	+0.2	-2.7	-6.3	-9.0	-14.0	-34.4	-33.0	-24.4	-17.1	-12.0	-4.7	-4.6
HyperCore (ours)	20.1±1.8	31.2±0.2	37.4±0.7	57.3±1.8	72.8±1.6	83.5±0.3	87.6±0.3	90.4±0.2	92.3±0.3	93.5±0.3	93.6±0.3	90.9±0.4
	-0.7	-1.3	-0.5	0.0	+1.8	+0.2	0.0	+0.3	-0.1	+0.1	-1.6	-4.6

Table 8: Fixed coreset selection performance of five randomly initialized WRN-16-8 under label noise on CIFAR-10, where 10% of the training labels are randomly corrupted by assigning them to incorrect classes. Bold entries indicate the highest performance at each data fraction.

Fraction ($1 - \alpha$)	0.1%	0.5%	1%	5%	10%	20%	30%	40%	50%	60%	90%	100%
Herding	11.3±0.8	10.7±1.6	8.9±1.7	8.6±1.4	12.9±2.5	32.8±1.7	56.0±1.5	71.7±1.6	79.5±1.0	85.1±0.8	91.3±0.3	91.6±0.2
	-12.5	-20.5	-30.7	-51.6	-53.9	-44.5	-25.9	-14.7	-9.6	-7.2	-3.9	-4.4
k-Center Greedy	12.6±1.0	15.0±0.9	18.3±0.7	31.4±1.6	47.8±3.9	68.8±2.3	78.3±0.9	82.4±0.6	85.5±0.3	87.8±0.2	91.1±0.2	91.6±0.2
	-6.6	-12.7	-16.3	-36.5	-33.8	-20.6	-13.8	-11.2	-8.8	-7.2	-4.8	-4.4
Forgetting	21.7±2.2	31.7±1.6	36.0±1.5	62.0±0.7	77.4±0.7	88.1±0.2	91.0±0.1	91.5±0.3	91.8±0.3	91.7±0.2	91.5±0.3	91.6±0.2
	+0.2	+1.6	-0.4	+3.5	+2.6	-1.7	-2.4	-3.0	-3.2	-3.7	-4.4	-4.4
GraNd	12.7±1.7	10.8±0.5	8.7±0.9	8.6±0.7	10.5±0.8	29.2±1.1	50.5±1.2	67.1±0.7	79.5±0.6	84.9±0.4	91.4±0.2	91.6±0.2
	-2.3	-8.8	-12.8	-29.1	-48.2	-52.5	-41.6	-27.2	-15.8	-10.7	-4.6	-4.4
CAL	24.9±2.0	36.2±1.3	44.2±2.2	67.4±1.3	73.9±0.9	82.5±0.5	85.5±0.6	88.7±0.3	90.6±0.3	91.6±0.3	92.8±0.2	91.6±0.2
	+2.6	-1.0	-1.3	+1.2	-0.2	-0.2	-0.7	-0.5	-0.4	-0.9	-2.6	-4.4
DeepFool	17.2±1.7	22.2±1.4	27.3±2.3	43.2±1.7	54.7±4.1	70.4±1.7	80.4±1.5	84.7±0.6	87.7±0.4	89.3±0.2	91.3±0.2	91.6±0.2
	-1.7	-7.0	-7.7	-14.2	-19.4	-17.1	-11.6	-8.8	-6.9	-5.9	-4.7	-4.4
Craig	20.6±0.9	24.6±2.3	27.0±2.5	28.5±1.8	33.2±1.8	46.8±4.0	62.5±1.8	74.6±0.6	82.2±0.5	86.5±1.0	91.2±0.2	91.6±0.2
	-1.3	-5.4	-11.1	-31.6	-36.0	-39.6	-28.8	-19.2	-12.5	-8.9	-4.7	-4.4
GradMatch	16.4±1.7	23.6±1.8	24.9±2.6	29.9±2.6	36.9±5.3	49.7±3.7	63.9±4.2	75.7±2.0	80.9±0.9	86.1±0.3	91.1±0.1	91.6±0.2
	-3.8	-4.2	-6.6	-24.5	-33.5	-34.9	-26.8	-17.6	-13.8	-9.3	-4.8	-4.4
Glister	17.1±0.7	24.8±1.4	28.8±2.6	28.3±2.9	39.3±2.7	47.8±3.8	63.4±5.0	74.0±1.2	81.5±0.4	86.5±0.2	91.0±0.1	91.6±0.2
	-3.0	-3.3	-2.8	-21.0	-28.5	-27.1	-19.5	-13.2	-8.7	-4.9	-4.4	-4.4
HyperCore (ours)	22.7±0.8	39.6±0.8	47.4±0.5	70.2±0.7	79.2±0.5	86.7±0.2	90.0±0.3	92.2±0.1	93.4±0.1	94.2±0.2	94.1±0.1	91.6±0.2
	-4.3	+0.4	+1.5	+2.8	+0.1	+0.2	-0.2	0.0	+0.2	+0.2	-1.5	-4.4