Samples with Low Loss Curvature Improve Data Efficiency - Supplementary

Isha Garg
Purdue University
West Lafayette, USA
gargi@purdue.edu

Kaushik Roy
Purdue University
West Lafayette, USA
kaushik@purdue.edu

A. Curvature Histograms and Corresponding Images for More Datasets and Networks

A.1. CIFAR-100

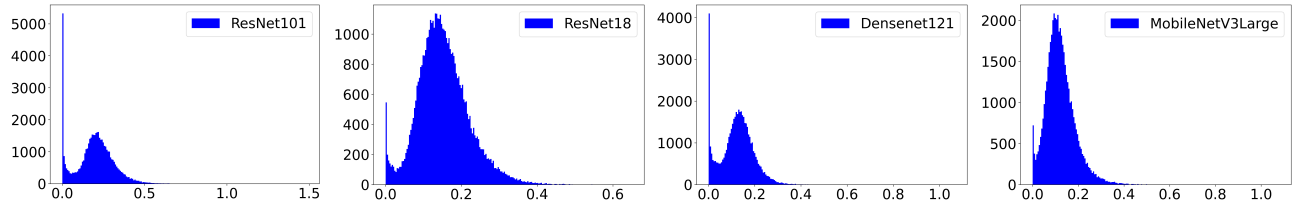
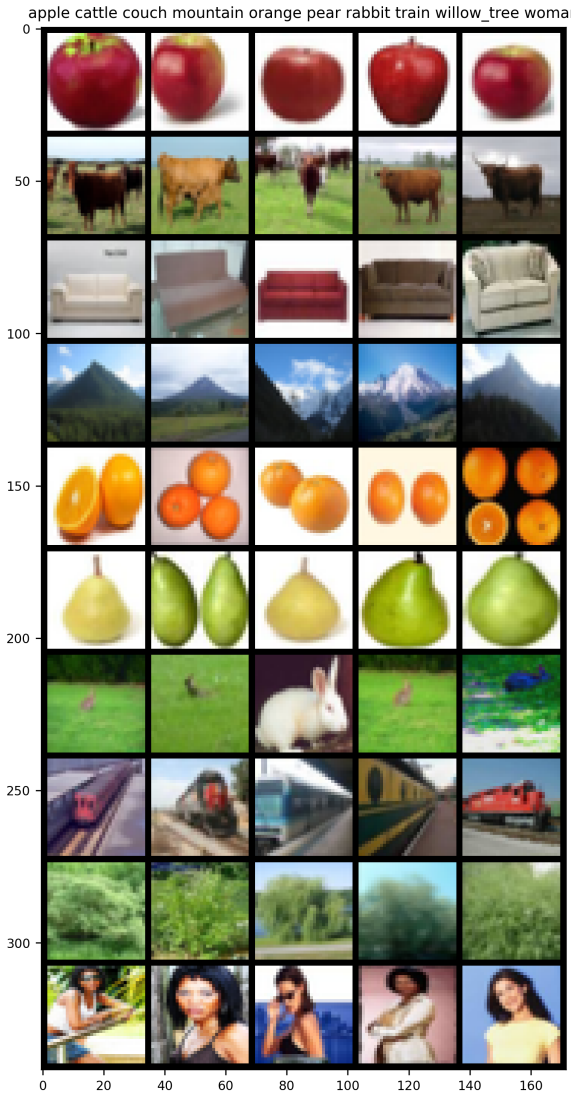
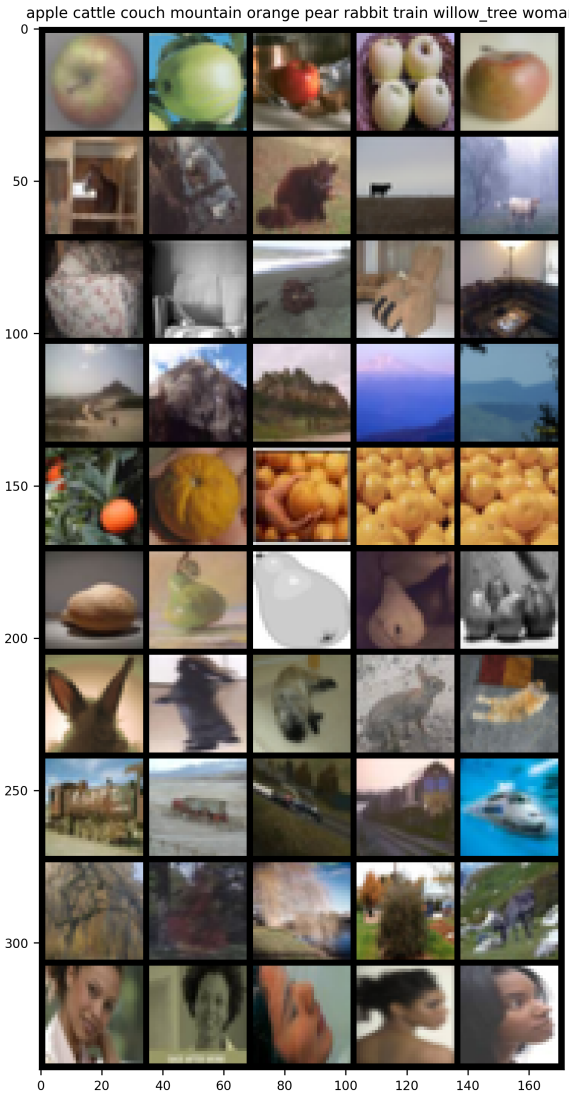


Figure 1. Curvature histograms for training set of CIFAR-100 for different trained networks. The same training hyperparameters were used as for CIFAR-10, mentioned in section 4.1 of the main text.



(a) Top 50 examples with the lowest loss curvature on CIFAR100



(b) Top 50 examples with the highest loss curvature on CIFAR100

Figure 2. Samples from 10 randomly chosen classes with lowest and highest curvature of loss identified over 3 different randomly initialized ResNet18 models trained on CIFAR-100. Each row corresponds to a class, consistent across both figures, and five ordered samples for each class are shown. The classes are apple, cattle, couch, mountain, orange, pear, rabbit, train, willow-tree and woman.

A.2. ImageNet

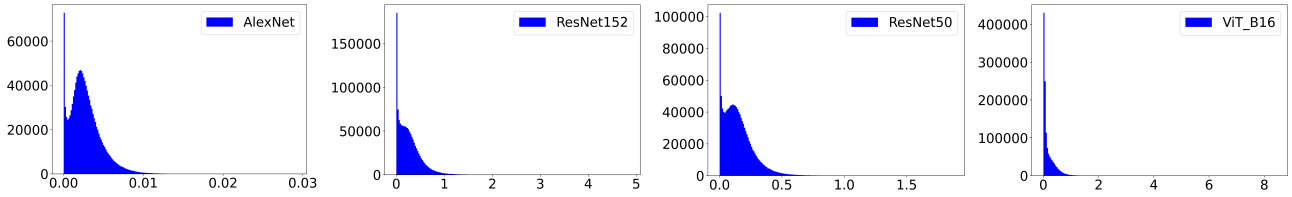


Figure 3. Curvature histograms for training set of ImageNet for different trained networks. The pretrained weights were selected from torchvision [**?**], and we selected the V1 weights for all models.

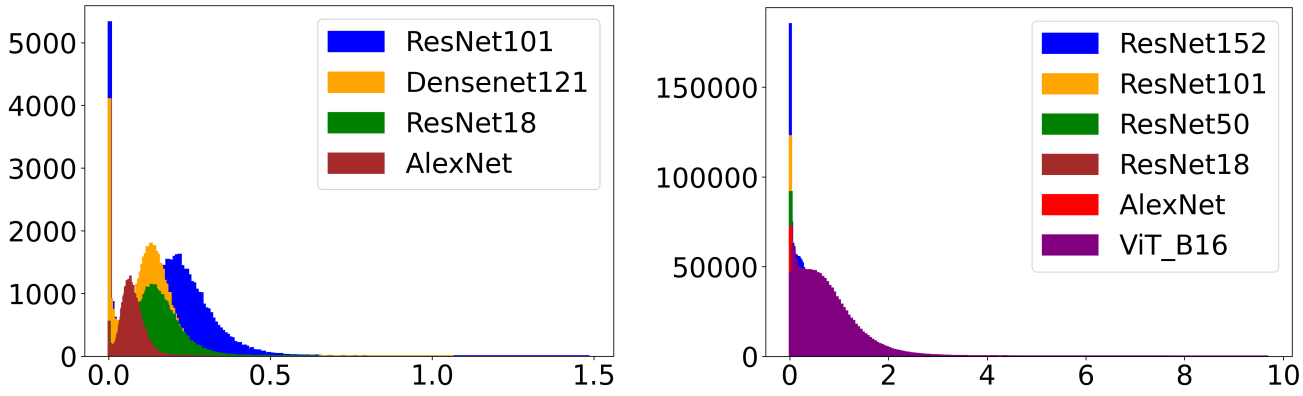
(a) Top 50 examples with the lowest loss curvature on ImageNet

(b) Top 50 examples with the highest loss curvature on ImageNet

Figure 4. Samples from 10 randomly chosen classes with lowest and highest curvature of loss identified over 4 different ResNet models of different sizes: ResNet18, ResNet50, ResNet101, ResNet152 trained on ImageNet. Each row corresponds to a class, consistent across both figures, and 5 ordered samples for each class are shown. The network was trained with standard random rescaled crops for augmentation, which was turned off to visualize these images. The classes are indian-cobra, gibbon, aircraft-carrier, binoculars, convertible, crutch, odometer, tile-roof, tripod and washer.

A.3. Overlap of Samples Across Architectures

Here, we show the histograms of curvature accumulated over different pretrained networks to show that these networks share many low curvature samples in common. The method for collecting these is described in Section 3.2 of the main text.



(a) CIFAR-100 (b) ImageNet

Figure 5. Cross architecture results to show commonality of low curvature samples across different architectures.

B. Mean and Variance Tables

In this section, we tabulate the results shown graphically in Section 4 of the main text. These numbers refer to the mean and variance of 5 randomly seeded runs.

B.1. CIFAR10: Comparison with Other Coreset Selection Methods

Table 1. Table corresponding to Figure 5a), Section 4.2 of main text.

NImPerC	Random	Cal	GraphCut	Glister	Grand	Forgetting	Herding	CRAIG	Margin	SLo-Curves
1	14.97 ±1.2	15.75 ±2.1	15.92 ±3.2	14.09 ±2.0	14.58 ±2.1	13.91 ±1.7	15.47 ±1.8	17.25 ±1.2	15.27 ±1.7	**18.54 ±1.7**
2	16.6 ±0.7	20.48 ±2.2	18.46 ±1.3	16.37 ±2.3	16.14 ±1.6	15.72 ±1.1	16.42 ±1.3	17.23 ±1.7	15.0 ±1.9	**21.45 ±1.1**
3	20.38 ±1.9	21.43 ±2.4	21.17 ±1.6	18.28 ±2.0	16.94 ±1.7	18.58 ±1.8	20.22 ±2.4	19.18 ±1.0	17.73 ±1.7	**22.35 ±0.6**
4	19.46 ±0.7	22.08 ±1.8	23.02 ±1.9	19.99 ±2.3	18.46 ±1.0	17.61 ±2.0	19.91 ±2.3	20.11 ±2.0	18.05 ±2.0	**23.35 ±1.3**
5	20.17 ±0.6	24.4 ±0.6	22.6 ±1.0	20.44 ±1.4	18.97 ±1.1	21.72 ±1.7	18.5 ±2.4	24.3 ±2.8	18.67 ±1.2	**25.69 ±1.7**
6	22.94 ±1.0	23.71 ±2.3	22.73 ±1.1	22.52 ±1.4	20.48 ±2.1	20.26 ±1.1	20.44 ±2.3	23.24 ±2.3	19.1 ±2.1	**25.86 ±0.3**
7	26.31 ±1.0	**27.02 ±1.1**	24.53 ±1.8	22.08 ±1.1	19.78 ±1.0	21.74 ±1.4	22.11 ±3.5	21.84 ±1.7	19.31 ±1.0	27.01 ±1.2
8	23.8 ±0.7	26.3 ±1.7	27.89 ±2.3	22.03 ±1.1	20.9 ±2.0	21.11 ±0.9	22.81 ±1.6	24.21 ±2.5	20.1 ±0.6	**28.58 ±1.3**
9	23.88 ±0.5	26.69 ±2.0	**28.88 ±1.4**	22.47 ±1.6	20.63 ±0.8	22.01 ±0.2	22.78 ±2.1	24.31 ±1.5	20.86 ±1.9	28.08 ±1.0
10	25.78 ±0.5	**29.53 ±1.8**	28.12 ±0.9	24.56 ±1.0	23.61 ±0.9	23.47 ±1.8	21.38 ±3.2	23.94 ±0.7	20.43 ±0.5	28.78 ±0.9
15	30.42 ±1.1	31.57 ±0.8	32.01 ±1.1	26.39 ±1.5	26.01 ±2.1	28.47 ±1.5	21.98 ±2.5	26.73 ±0.8	21.62 ±1.3	**35.14 ±0.9**
20	32.42 ±0.8	35.15 ±2.8	35.38 ±0.7	28.84 ±1.3	26.9 ±1.2	28.56 ±1.0	22.02 ±4.3	29.73 ±1.0	23.98 ±1.1	**38.35 ±1.7**
25	34.04 ±0.2	35.52 ±1.0	36.67 ±1.1	31.44 ±1.6	27.28 ±1.9	29.57 ±1.2	25.89 ±3.8	30.87 ±1.5	25.19 ±1.5	**38.06 ±0.5**
30	34.83 ±0.4	37.61 ±1.9	39.05 ±1.1	32.18 ±0.4	29.12 ±0.8	33.19 ±0.4	24.93 ±3.9	31.13 ±2.8	27.36 ±2.1	**42.21 ±0.9**
35	36.61 ±1.0	41.17 ±1.8	40.29 ±1.5	33.2 ±1.3	28.43 ±0.9	34.1 ±1.2	31.09 ±3.3	30.74 ±2.8	28.41 ±1.2	**42.98 ±0.7**
40	38.22 ±0.6	40.61 ±1.3	42.61 ±1.6	32.68 ±0.5	30.13 ±1.7	34.57 ±0.5	29.69 ±3.1	30.63 ±3.0	28.0 ±1.2	**44.38 ±0.5**
45	41.05 ±1.2	42.42 ±1.0	44.02 ±0.7	32.85 ±1.3	31.22 ±1.0	33.62 ±1.2	25.86 ±3.2	34.7 ±1.3	27.03 ±1.8	**47.55 ±1.0**
50	39.32 ±0.5	44.98 ±1.2	47.36 ±0.8	35.24 ±1.6	31.69 ±1.3	34.56 ±0.9	30.74 ±2.4	37.32 ±2.0	30.94 ±1.6	**48.26 ±0.8**
100	50.3 ±0.6	51.64 ±1.7	56.09 ±1.6	39.88 ±1.8	36.53 ±1.5	41.46 ±1.0	39.21 ±3.3	41.49 ±1.8	36.63 ±1.5	**57.07 ±1.3**

B.2. CIFAR100 : Comparison with Other Coreset Selection Methods

Table 2. Table corresponding to Figure 5b), Section 4.2 of main text.

NImPerC	Random	Cal	GraphCut	Glister	Grand	Forgetting	Herding	CRAIG	Margin	SLo-Curves
1	3.66 ±0.1	5.47 ±0.4	5.27 ±0.4	4.22 ±0.5	4.18 ±0.3	4.39 ±0.2	2.57 ±0.2	4.85 ±0.4	2.87 ±0.2	**6.03 ±0.1**
2	6.01 ±0.2	7.59 ±0.4	7.06 ±0.3	6.27 ±0.6	5.5 ±0.4	6.13 ±0.3	3.42 ±0.3	6.81 ±0.5	4.06 ±0.4	**7.97 ±0.3**
3	6.8 ±0.1	**10.08 ±0.5**	8.79 ±0.3	7.51 ±0.2	6.81 ±0.3	6.8 ±0.3	4.07 ±0.3	8.31 ±0.5	4.31 ±0.5	9.44 ±0.2
4	7.96 ±0.4	10.17 ±0.5	9.51 ±0.5	8.26 ±0.5	8.14 ±0.6	8.04 ±0.6	4.59 ±0.6	9.04 ±0.7	5.53 ±0.5	**10.52 ±0.6**
5	9.38 ±0.3	12.16 ±0.7	11.58 ±0.4	9.82 ±0.4	9.79 ±0.4	9.91 ±0.6	5.36 ±0.6	11.19 ±0.9	6.37 ±0.4	**12.7 ±0.3**
10	12.74 ±0.4	16.37 ±0.6	16.62 ±0.2	14.58 ±1.1	14.21 ±0.8	15.27 ±0.8	8.29 ±0.8	15.11 ±1.0	8.64 ±0.7	**17.52 ±0.4**
15	16.58 ±0.4	20.01 ±0.3	**20.43 ±0.7**	17.99 ±0.9	17.62 ±0.9	17.83 ±0.9	9.24 ±0.9	18.4 ±0.5	10.4 ±0.8	20.32 ±0.5
20	19.99 ±0.8	22.47 ±1.1	**24.33 ±0.6**	23.08 ±1.3	21.84 ±0.7	23.34 ±0.5	11.57 ±0.5	21.17 ±0.1	12.71 ±0.5	23.97 ±0.5
25	21.82 ±0.4	24.03 ±0.7	**26.95 ±0.5**	25.56 ±1.0	24.92 ±1.3	26.18 ±1.2	13.66 ±1.2	23.07 ±1.1	14.38 ±0.5	26.15 ±0.4
30	23.66 ±0.5	26.38 ±0.8	**30.75 ±0.6**	29.17 ±0.5	27.88 ±1.4	28.16 ±1.1	15.49 ±1.1	24.57 ±0.7	17.09 ±0.7	26.98 ±0.8
35	28.36 ±0.6	27.63 ±0.4	**33.58 ±0.8**	31.24 ±0.8	32.13 ±1.2	31.95 ±0.8	17.2 ±0.8	25.5 ±1.2	18.85 ±1.5	29.82 ±0.4
40	30.75 ±0.8	30.68 ±0.7	**34.48 ±0.8**	33.98 ±1.0	34.08 ±0.9	33.7 ±1.3	18.52 ±1.3	26.39 ±0.5	21.16 ±2.0	32.52 ±0.7
45	32.12 ±0.7	32.02 ±1.3	**37.08 ±0.4**	36.03 ±0.8	35.95 ±0.6	36.79 ±0.7	19.86 ±0.7	28.63 ±1.4	23.35 ±1.4	33.49 ±0.4
50	32.75 ±1.6	34.02 ±0.6	**38.77 ±0.7**	35.0 ±1.0	37.84 ±1.0	38.03 ±1.0	20.54 ±1.0	28.28 ±1.0	24.97 ±0.7	35.7 ±0.5

B.3. Cross Architecture Results for CIFAR-10

Table 3. Table corresponding to Figure 6a), Section 4.3 of main text.

NImPerC	Random (10%)	ResNet18 (diff seed) (91.8%)	ResNet101 (94.7%)	AlexNet (87.1%)	MobileNetV3 (79%)	DenseNet121 (93.8%)	VGG19 (93.3%)
1	14.97 + 1.2	18.54 + 1.7	16.49 + 0.5	18.82 + 1.4	17.1 + 1.0	18.0 + 1.1	18.14 + 1.1
2	16.6 + 0.7	21.45 + 1.1	19.78 + 1.2	21.6 + 1.2	19.74 + 0.9	22.45 + 0.7	20.25 + 0.5
3	20.38 + 1.9	22.35 + 0.6	23.82 + 0.8	23.28 + 0.5	22.45 + 1.6	22.39 + 1.0	22.8 + 1.2
4	19.46 + 0.7	23.35 + 1.3	25.23 + 1.4	24.35 + 1.2	26.28 + 0.8	24.59 + 0.9	24.62 + 1.1
5	20.17 + 0.6	25.69 + 1.7	25.64 + 0.3	27.33 + 0.3	26.53 + 1.2	27.83 + 1.1	25.26 + 1.3
10	25.78 + 0.5	28.78 + 0.9	31.94 + 0.7	30.46 + 0.3	29.71 + 0.8	30.49 + 1.2	28.92 + 0.8
15	30.42 + 1.1	35.14 + 0.9	34.24 + 0.5	35.82 + 0.8	37.09 + 0.2	33.75 + 0.7	33.83 + 0.5
20	32.42 + 0.8	38.35 + 1.7	37.39 + 1.0	37.75 + 1.9	37.96 + 0.5	37.69 + 0.8	36.84 + 0.6
30	34.83 + 0.4	42.21 + 0.9	43.29 + 0.9	43.23 + 0.4	45.66 + 0.7	40.81 + 2.7	40.01 + 0.5
40	38.22 + 0.6	44.38 + 0.5	45.98 + 1.2	44.29 + 1.0	46.95 + 0.5	42.81 + 0.7	43.8 + 0.4
50	39.32 + 0.5	48.26 + 0.8	48.35 + 0.6	46.58 + 1.7	50.32 + 0.8	45.87 + 0.6	46.01 + 1.3
100	50.3 + 0.6	57.07 + 1.3	57.05 + 0.6	55.34 + 0.2	55.55 + 0.5	53.85 + 0.3	55.69 + 0.8

B.4. Ablation Study for Regularization for ResNet18, trained on CIFAR-10

Table 4. Table corresponding to Figure 6a), Section 4.4 of main text.

NImPerC	Random Selection	SLo-Curves Selection without regularization	SLo-Curves Selection with best regularization
1	14.97 + 1.2	16.27 + 1.0	18.54 + 1.7
2	16.6 + 0.7	18.85 + 0.7	21.45 + 1.1
3	20.38 + 1.9	19.14 + 0.8	22.35 + 0.6
4	19.46 + 0.7	19.71 + 1.0	23.35 + 1.3
5	20.17 + 0.6	24.49 + 0.9	25.69 + 1.7
6	22.94 + 1.0	22.73 + 1.1	25.86 + 0.3
7	26.31 + 1.0	24.07 + 0.8	27.01 + 1.2
8	23.8 + 0.7	23.76 + 0.5	28.58 + 1.3
9	23.88 + 0.5	27.65 + 1.4	28.08 + 1.0
10	25.78 + 0.5	26.75 + 0.9	28.78 + 0.9
15	30.42 + 1.1	32.19 + 0.6	35.14 + 0.9
20	32.42 + 0.8	34.13 + 1.2	38.35 + 1.7
25	34.04 + 0.2	35.04 + 1.0	38.06 + 0.5
30	34.83 + 0.4	37.36 + 0.8	42.21 + 0.9
35	36.61 + 1.0	41.99 + 0.4	42.98 + 0.7
40	38.22 + 0.6	40.97 + 0.5	44.38 + 0.5
45	41.05 + 1.2	44.6 + 0.6	47.55 + 1.0
50	39.32 + 0.5	45.37 + 1.2	48.26 + 0.8
100	50.3 + 0.6	54.84 + 0.6	57.07 + 1.3

B.5. Early Emergence of Low Curvature Samples, for DenseNet121 trained on CIFAR-10

Table 5. Table corresponding to Figure 7b), Section 4.5 of main text.

NImPerC	Random (10%)	DenseNet RandomInit (10%)	DenseNet Epoch 10 (84.7%)	DenseNet Converged (93.8%)
1	14.97 + 1.2	16.71 + 1.6	16.49 + 0.5	18.0 + 1.1
2	16.6 + 0.7	18.27 + 0.5	19.78 + 1.2	22.45 + 0.7
3	20.38 + 1.9	19.96 + 0.8	23.82 + 0.8	22.39 + 1.0
4	19.46 + 0.7	20.45 + 0.4	25.23 + 1.4	24.59 + 0.9
5	20.17 + 0.6	22.31 + 0.4	25.64 + 0.3	27.83 + 1.1
10	25.78 + 0.5	23.21 + 1.1	31.94 + 0.7	30.49 + 1.2
15	30.42 + 1.1	25.48 + 0.3	34.24 + 0.5	33.75 + 0.7
20	32.42 + 0.8	26.88 + 0.9	37.39 + 1.0	37.69 + 0.8
30	34.83 + 0.4	29.46 + 1.0	41.34 + 1.3	40.81 + 2.7
40	38.22 + 0.6	30.38 + 0.9	43.25 + 0.9	42.81 + 0.7
50	39.32 + 0.5	33.94 + 0.4	45.71 + 0.3	45.87 + 0.6
100	50.3 + 0.6	41.51 + 0.8	54.8 + 1.0	53.85 + 0.3