EFFECTIVE RELATIONSHIP BETWEEN CHARACTERISTICS OF TRAINING DATA AND LEARNING PROGRESS ON KNOWLEDGE DISTILLATION

– SUPPLEMENTARY MATERIAL –

Anonymous ICIP submission

Anonymous Affiliation

1. EXPEIMENTAL DATA

Table 1 shows details of the experimental dataset used in our evaluation experiments in Section 4 of our main paper.

Dataset	Class	Train images	Test images
Flowers102	102	1,020	6,149
Pets37	37	3,312	3,669
Food101	101	68,175	25,250

Table 1. Datasets used in our experiments. Numbers of classes, training images and test images.

2. EXPERIMENTS USING COMBINATION OF RANGE RESTRICTION AND LEARNING DEGREE CONTROL WITH MORE EPOCHS

2.1. Cosine curve decrease in latter half

Table 2 gives concrete values of the mean and the standard deviation on the combination of restricting sampling range and decreasing upper bound for mix ratios with the trend of the cosine curve in the latter half plotted in Fig. 6 with each number of epochs in our main paper.

2.2. Linear decrease in latter half

Fig. 1 graphs experimental results using the combination of the sampling range restriction and the learning degree control using linear descent in the latter half of the learning process. Also, concrete values used in Fig. 1 are tabulated in Table 3. Fig. 1 and Table 3 provide that the more epochs, the narrower sampling range gives better improvement similar to those shown in Section 4.2.3 in our main paper. This means that the diversity based on the combination of classes is over that given by the random sampling of mix ratios; then, the high output entropy derived from sampling range restriction improves generalization performance.

Range	Flowers102		Pets37			Food101	
	1,000	10,000	300	1,000	3,000	100	1,000
normal	74.31±0.65	79.03±0.59	74.27±0.17	77.73±0.50	81.96±0.54	80.60±0.07	83.46±0.08
1/2	76.97±0.12	81.83±0.99	75.93±0.64	80.51±0.54	84.60±0.26	81.04±0.17	84.39±0.17
1/4	78.41±0.67	83.10±0.62	74.89±0.54	80.94±0.33	85.62±0.15	81.16±0.03	84.70±0.13
zero	77.59±0.47	82.95±0.20	73.36±0.55	79.75±0.26	85.19±0.29	80.60±0.17	84.79±0.14

Table 2. Classification accuracies [%] on combination of restricting sampling range and decreasing upper bound for mix ratios with trend of cosine curve in latter half for each dataset. Second row expresses epochs number.

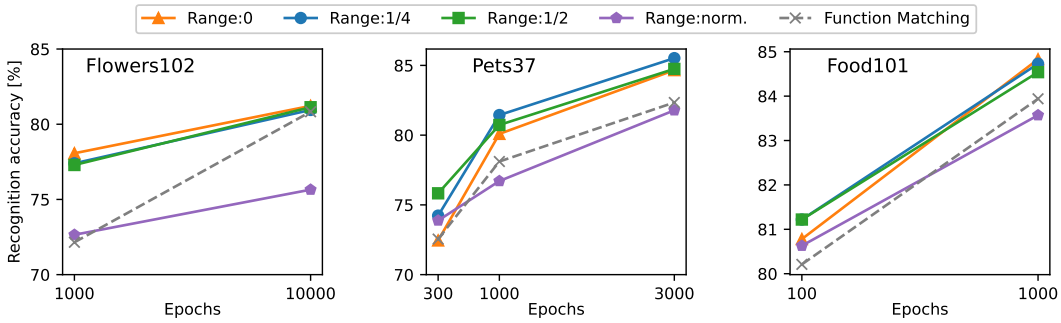

Fig. 1. Accuracies [%] for each dataset using the combination of our proposals with the decrease using linear descent. "Lin" means linear descent. "Range:norm, $1/2$, $1/4$, and zero" express normal (default) sampling range, a half sampling range, a quarter range for sampling, and no range for sampling, respectively. "FunMatch" means Function Matching.

Range	Flowers102		Pets37			Food101	
	1,000	10,000	300	1,000	3,000	100	1,000
normal	72.65±0.79	75.65±0.64	73.88±0.27	76.71±0.55	81.78±0.11	80.63±0.13	83.57±0.03
1/2	77.29±0.32	81.13±0.86	75.83±0.24	80.73±0.75	84.75±0.44	81.08±0.14	84.54±0.07
1/4	77.40±0.47	80.95±0.60	74.23±0.32	81.44±0.42	85.52±0.14	81.22±0.11	84.73±0.03
zero	78.07±0.29	81.22±0.84	72.45±0.16	80.09±0.69	84.66±0.82	80.78±0.03	84.83±0.07

Table 3. Classification accuracies [%] on combination of restricting sampling range and decreasing upper bound for mix ratios with trend of linear decline in latter half for each dataset. Second row expresses epochs number.