Learning Scalable Discriminative Dictionary with Sample Relatedness – Supplementary Material

Jiashi Feng[1,2], Stefanie Jegelka[2], Shuicheng Yan[1], Trevor Darrell[2]
[1]Department of ECE, National University of Singapore, Singapore
[2]Department of EECS & ICSI, UC Berkeley, USA
[1]{a0066331,eleyans}@nus.edu.sg, [2]{stefje,trevor}@eecs.berkeley.edu

1. Convergence

The convergence curve of the objective function in Eqn. (9) is given in Figure S.1, on the AwA dataset with $2,000$ training samples. The parameters are set as $\lambda = 1 \times 10^{-3}$ and $\beta = 0.1$. It can be seen that Algorithm 1 converges after about 100 iterations.

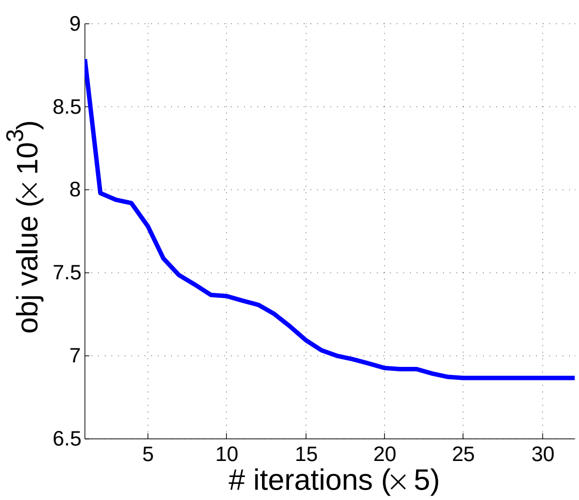

Figure S.1. The curve of objective function value vs. # iterations. The optimization process converges after around 150 iterations.

2. Component-wise Evaluation of the Model

Here, we aim to give some insights on the effectiveness of each component of our proposed model in Eqn. (9). In particular, we evaluate the discriminative component (Eqn. (6)) and the generative component (Eqn. (5)) of the proposed model individually. The evaluations are conducted for the task of classifying the 40 known categories and 10 novel categories on the AwA dataset.

First, we implement only the generative component of the model with parameter $\lambda = 1 \times 10^{-3}$ and obtain the estimated number of dictionary basis, *e.g.*, 196 basis are discovered from the $2,000$ training samples. Then we implement the discriminative component solely with the same basis number as the discovered by generative component. In this way, we can fairly compare the discriminative and generalization performance of these two components, using the binary representation of the same dimensionality. The evaluation results are shown in Table S.1.

Not surprisingly, for classifying the known categories, the generative component performs much worse than the discriminative component for classifying known categories. When sufficient training samples are provided, *e.g.*, 50 training images per category, discriminative component outperforms the generative component with a large margin of 10%. In addition, providing more training samples does not improve the performance of generative component much, in stark contrast to the discriminative component. Such significant difference roots in that the discriminative component fully utilize the supervision information with the training samples and thus can discover more discriminative dictionary. We also evaluate the performance of the complete model and the results demonstrate that combining the generative component into the discriminative component brings additional $2\% \sim 3\%$ accuracy enhancement.

Table S.1. Classification performance comparison on known and novel categories between the discriminative component (Eqn. (6)) and the generative component (Eqn. (5)). We show the performance on known categories classification with varying numbers of training samples per category. The performance for classifying novel categories based on 50 training samples per category is also reported. The numbers of learned dictionary basis are displayed in parentheses, in the column for generative component.

# labeled	Discriminative	Generative	Whole Model
15	22.21	18.67 (173)	24.25
20	24.68	19.23 (170)	26.25
25	27.32	20.45 (175)	29.38
30	30.56	20.06 (187)	32.13
50	30.74	20.48 (196)	33.00
Novel (50)	62.35	60.58 (196)	70.00

However, when classifying the novel categories, the case is quite different. The discriminative and generative components achieve comparable performance. Combining them together boosts the performance significantly (larger than 7% enhancement). This demonstrates that the generative component can automatically discover some general dictionary basis which are useful for describing novel categories.