Learning Class Prototypes via Structure Alignment for Zero-Shot Recognition

Huajie Jiang[1,2,3,4][0000−0002−1158−6321], Ruiping Wang[1,4][0000−0003−1830−2595], Shiguang Shan[1,4][0000−0002−8348−392X], and Xilin Chen[1,4][0000−0003−3024−4404]

[1]Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, 100190, China
[2]Shanghai Institute of Microsystem and Information Technology, CAS, Shanghai, 200050, China
[3]ShanghaiTech University, Shanghai, 200031, China
[4]University of Chinese Academy of Sciences, Beijing, 100049, China
huajie.jiang@vipl.ict.ac.cn, {wangruiping, sgshan, xlchen}@ict.ac.cn

This document gives more details about the parameter settings and the visualization results of the unseen-class prototypes.

1 Parameter Settings

In this part, we analyze some parameters used in our framework. The parameters λ, α, β in the objective function Eq.(5) control the relative importance of semantic information, domain adaptation and prototype learning. γ in Eq.(9) is a regularization parameter. The values of these four parameters used in our paper are shown in Table 1. In our experiments, we find that λ is important in the learning process. When it is set as 1, the learning process is more stable though the best performance is obtained by 10 on SUNA. For parameter γ, we simply set it as 1 for all datasets. Other parameters are slightly different according to the characters of different datasets.

Table 1: Parameters used on the four datasets: aPY , AwA , CUB and SUNA.

Dataset	λ	α	β	γ
aPY	1	0.01	0.001	1
AwA	1	1	0.1	1
CUB	1	0.1	0.1	1
SUNA	10	0.001	0.01	1

2 Visualization Results

The proposed framework automatically learns the prototypes of all class. To understand what the model has learned, we visualize the unseen-class prototypes

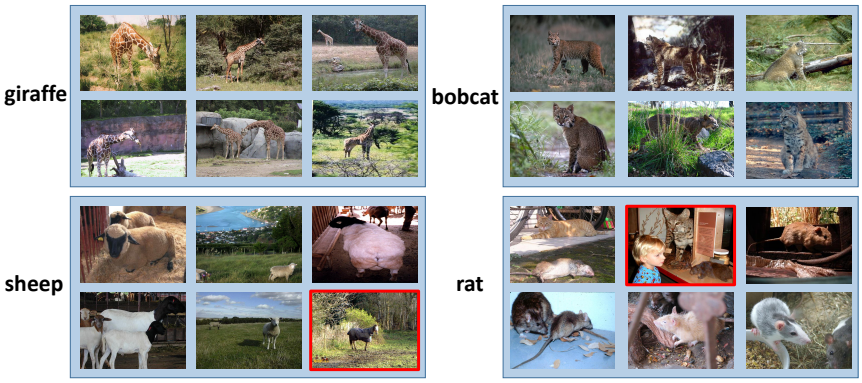

Fig. 1: Visualization of the unseen-class prototypes with the six most similar images in the visual space on AwA. Images with red rectangles are inconsistent with the corresponding classes.

by showing their nearest neighbours in the visual space. Figure 1 shows some prototype visualization results on AwA. It can be figured out that the prototypes are really representations of corresponding classes though some outliers may exist.