



Fig. 1. Result comparison of TagOOD and baseline on several confusing OOD samples. The texts in black represent IND categories corresponding to the image. ReAct produces scores in yellow, and our TagOOD generates the green one.

Supplementary Materials: TagOOD: A Novel Approach to Out-of-Distribution Detection via Vision-Language Representations and Class Center Learning

ANONYMOUS AUTHORS

1 IMAGE FEATURE DECOMPOSITION IN HANDLING CONFUSING OOD SAMPLES

To further illustrate TagOOD’s robust OOD detection capabilities, Figure 1 presents examples of challenging scenarios. Images with black box labels represent in-distribution (IND) data from ImageNet-1K, while others are out-of-distribution (OOD) data from Places. Each group shares common tags, displayed within the red boxes. Additionally, two scores displayed above each OOD image represent the OOD scores for ReAct (yellow) and TagOOD (green), respectively. The OOD scores range from 0 (low probability of being IND) to 1 (high probability of being IND).

The first group on the lefttop showcases two IND images, categorized as "swimming trunks" and "paddle." However, both images contain a surfboard, an object not labeled in ImageNet-1K. This suggests that models trained on ImageNet-1K might inadvertently learn features related to surfboards, potentially leading to misclassification by ReAct. In contrast, TagOOD’s tagging model successfully identifies the surfboard and recognizes it as absent from the IND vocabulary T^{in} . Consequently, TagOOD discards the surfboard feature and relies solely on IND object features for evaluation. This approach enables TagOOD to accurately detect the OOD sample, as demonstrated by its higher score. Similar scenarios are reflected in other groups within Figure 1.