

Out-of-distribution detection using feature encoder

Juseung Yun, and Junmo Kim, EE, KAIST, Daejeon Korea

Abstract— Nowadays machine learning based Deep Neural Networks show high performances in classification problems. Despite high performances, there is the problem that even well-trained networks can not distinguish In-Distribution(ID) samples and Out-of-Distribution(OOD) samples. In this paper, we propose a simple yet effective method for detecting out-of-distribution samples. We created an additional module that receives the intermediate features of the main network and outputs different encoded values for each class. In the case of in-distribution it was easy to fit the encoded code, but in the case of out-of-distribution it was more difficult to fit the code. Thus we distinguished between in and out-of-distribution by how well the features fit this code. The proposed method is evaluated on CIFAR-10 as in-distribution and showed better performance compared to the baseline method.