Unsupervised Object Discovery With Masked Autoencoders

Essam Sleiman,[1] **Hamed Pirsiavash,**[1]

[1] University of California, Davis

Abstract

Object detectors are used in autonomous systems and typically have a particular focus on anomaly detection as wrong predictions can cost lives in safety-critical scenarios. To achieve a high level of performance, these systems need to train on vast amounts of costly annotated data. In this work, we investigate a method for Masked Autoencoders (MAEs) to detect unseen objects in unannotated images which can be used to find and later annotate anomalies.

MAEs are optimized to better predict masked image patches. To the model, in some cases, multiple different predictions for a masked patch can all be semantically true. For example, in an image of a dog in a field, if the model were to randomly mask the dog, it could predict the mask to be any number of objects it has seen before, such as a dog, rabbit, bird, etc. As a result, we suspect the model's learned prediction is influenced by a weighted average of these potential options. We hypothesize patches that are difficult to correctly predict are potential objects unseen by the model, however this previously mentioned scenario introduces false negatives. To diminish this effect, we propose to produce k ¿ 1 different prediction per masked patch. We believe the model will now learn to predict each of the different potential options in each of the k heads. To test this, we ran an experiment on the original MAE, creating heatmaps of the patch prediction losses, and consistently found false negatives (poor patch prediction performance for regions with objects seen before by the model). After applying the above k-prediction method, our preliminary results show these false negatives are reduced by 24%. These results show our method has the potential to perform well in object discovery.