

In this paper, the authors investigate effect of background on the efficiency of object detection models and in turn find interesting relationship between background and foreground of an image. Interestingly enough they find models to be surprisingly dependant on background while classifying objects and can show considerably good predictions by relying on background alone! More alarming yet, models tend to missclassify correctly labeled foreground images if not coupled with the known background. The third observation being that more accurate models tend to depend less on background.

The authors curated a dataset called Imagenet-9 where they separate background from foreground and then evaluate the performance of state-of-the-art models.

The dataset is further divided into seven parts differeing in how they process the background and foreground.

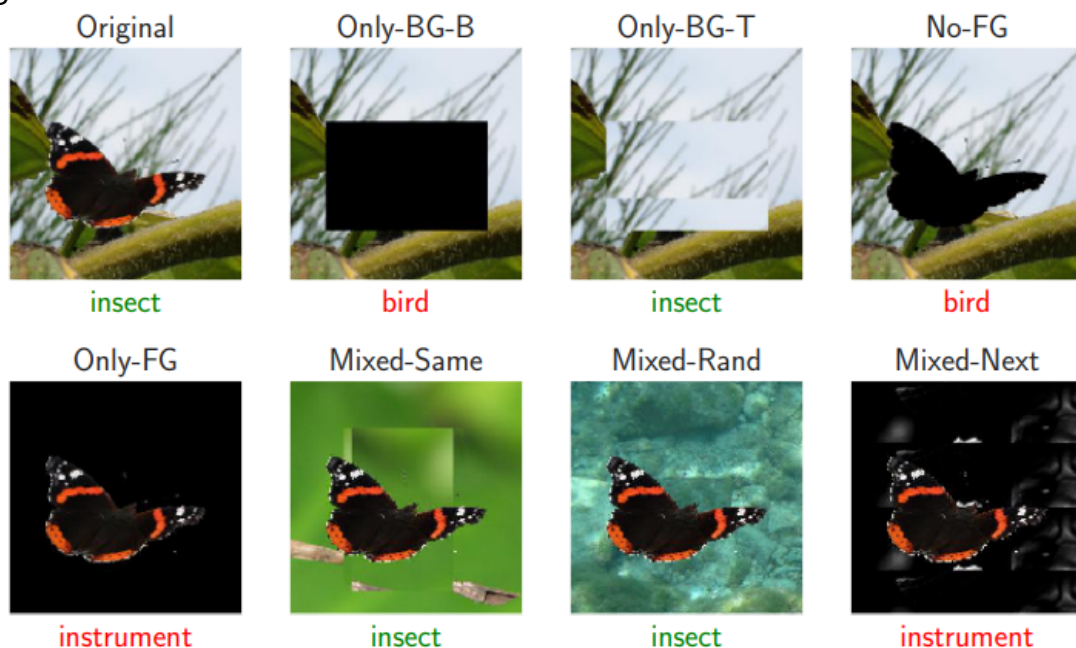


Figure 1: Variations of the synthetic dataset ImageNet-9, as described in Table 1. We label each image with its pre-trained ResNet-50 classification—green, if corresponding with the original label; red, if not. The model correctly classifies the image as “insect” when given: the original image, only the background, and two cases where the original foreground is present but the background changes. Note that, in particular, the model fails in two cases when the original foreground is present but the background changes (as in MIXED-NEXT or ONLY-FG).

Image source: <https://arxiv.org/abs/2006.09994>

Salient findings from the experiment:

1. Models can make apt classification based on background alone
2. Models take into account background signal to make decision
3. Models perform poorly when presented with the same data but with adversarial background

Some measures to take to mitigate such problems

1. Train models with images where background is decoupled from labels
2. perform specific image augmentation techniques to reduce such correlation

3. use training algorithms like distributionally robust optimization [Sag+20] and model-based robust learning [RHP20]

My takeaways

Machine learning models don't necessarily work the way we want to even if they seem to produce desired outputs. The fact that images utilize background and foreground may or may not be a good thing as humans also use contextual information to recognise objects. However humans are also good at recognising those objects when they are detached from the familiar context as well.