# (1)

## (a)

The issue is that the model detects only blue bottles more accurately than bottles of other colors. It probably occurred because of “Class Imbalance” .

It means there are other similar objects that are actually not the “bottle” but due to their structure the model also detects them. Now before training a model we have to give it datasets and set the classes. Now here we actually set the class to “bottle” but I think in the image there were other objects similar in structure to the bottle or there may be more “blue bottles” than other colored bottles. That’s why the “blue” got prioritized over other colors.

This is a problem related to the “training dataset”. I think During training most of the bottle images in the dataset were likely blue-colored. So the model has learned to associate the *“*bottle*”* class with the blue color feature instead of the shape or structure of the bottle.
The model actually overfitted to color specific bottles instead of focusing on color variations.

## (b)

**Solution:**

1) I need to balance the dataset by adding more bottle images of different colors such as green, red, transparent, etc so that the model can learn color invariance.

2) I need to check the data augmentation also by using color jittering, hue/saturation shifts, or brightness randomization to simulate color diversity. Also to make the images more clearer to understand.

3) Then I could apply techniques like dropout or early stopping to prevent overfitting on color specific patterns.

4) I need to fine tune the model on a balanced dataset with equal representation of all bottle colors. It means taking a pre-trained model (one that already learned general features from a large dataset) and then training it a bit more on my specific dataset so it adapts to the bottle detection task.

# (2)

## (a)

The formal name of the issue that I can think of  is  dataset bias or biased training data. It is a form of overfitting to color features .

## (b)

The most likely cause is that in the training dataset every single or most of the mallets were orange and no other orange objects were present in the background or as negative examples. As a result the model learns that “Orange” means mallet. But there can be other orange objects right? And so the model also detects them as a mallet which is not.

It actually fails to learn the structure of the mallet. It actually learns the color. Because here color is a dominant factor but we want the model to detect the shape,texture,contexts all together.

And I think training dataset problem is a dominant factor than training problem. Because training problem will occur if the training dataset has problems.

## ( c )

## Solution:

1. I need to improve the data by adding diversity in mallets datasets by Including blue, green, wooden and metallic mallets. Also to make the datasets more realistic I may have to Include negative examples such as adding orange cones, fruits, toys labeled as "not mallet" etc.

2. For data Augmentation I may have to add a color jittler such as randomly changing hue, saturation and brightness. I will force the model to rely on shape instead of color and remove color reliance features.

3. For validation I need to test with colored non-mallets to check for false positives and verify detection of non orange mallets to check for false negatives

# (3)

## (a)

Lack of negative samples or absence of background  examples in the dataset is causing this problem.

The model has only seen images that always contain bottle or mallet during training.
 So it learns that every image must contain one of those objects but  it never learned what an empty background looks like.
 As a result during testing on open landscapes it forces itself to detect on random areas.

## (b)

## Solution:

1) **I** need to add negative samples as background images by Including many images with no objects at all just Martian desert . Also I need to label them as having no bounding boxes. This will teach the model that sometimes there are no targets present in that scene.

2) I need to Increase dataset diversity by Including  backgrounds with different lighting, textures etc for augmentation purpose. This will prevent the model from confusing random patterns with real objects.

3) For regularization purpose I may need to raise the detection confidence threshold to reduce random low-confidence detections.