**1.1. Examine the classifier from the classifier results provided**

There are total 43 classes in the training set; in the classifier results there are only 22 classes detected. There is no detection for any of the remaining classes.

In Figure 1: 4491 detected objects found (blue), classified in only 22 classes. However, the number true objects (1370) is less than 1/3 of detected objects and distributed quite evenly among all 43 classes. The number of true detections is higher on the right side, where no detection found: 731/1370 = 53%. That means the machine learning model **missed out at least 53%** of the objects 🡪 not very good model.

**1.2. Which objects classified well and which more poorly by the machine learning algorithm?**

**The objects classified poorly (23 → 43 in Fig. 1): Zero detections**

For each of the 21 classes with zero detections:

* True Positive = 0, False Positive = 0
* False Negative ≥ 1

Hence, Precision = 0 and Recall = 0: These objects classified poorly.

We can consider penalty for classifier that missed out more objects: e.g. rainbowPens missed out the biggest number of objects (86) — that is the worst classifier, followed by turtle (60).

**The objects classified better (1 → 22 in Fig. 1): Detected objects**

For each class that has objects detected, we use Average Precision (AP) — precision averaged across all recall values between 0 and 1, to evaluate the performance. Then we compare AP between different object detectors to rank them.

By interpolating all points, the Average Precision (AP) can be interpreted as an approximated area under the curve (AUC) of the Precision x Recall curve (PASCAL VOC challenge). Higher value means higher accuracy on detection.

The process for getting AP is in Figure 2.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group: Classifiers can detect objects | | |  | Group: Classifiers cannot detect objects | | | |
| Rank | **Class name** | **AP (%)** |  | **Rank** | **Class name** | **AP (%)** | **Penalty** |
| 1 | blueSunglasses | 36.24 |  | **23** | voiceRecorder | 0 | 9 |
| 2 | paperPlate | 30.7 |  | **24** | cowbell | 0 | 15 |
| 3 | yellowBalloon | 28.7 |  | **25** | redDart | 0 | 16 |
| 4 | pumkinNotes | 25.01 |  | **26** | gyroscope | 0 | 18 |
| 5 | metalKey | 20.97 |  | **27** | carabiner | 0 | 19 |
| 6 | stickerBox | 20.46 |  | **28** | spiderRing | 0 | 23 |
| 7 | cupcakePaper | 17.88 |  | **29** | cactusPaper | 0 | 23 |
| 8 | partyFavor | 17.3 |  | **30** | giftBag | 0 | 26 |
| 9 | vancouverCards | 16.88 |  | **31** | hairRoller | 0 | 27 |
| 10 | noisemaker | 14.93 |  | **32** | pinkEraser | 0 | 30 |
| 11 | canadaPencil | 12.09 |  | **33** | glassBead | 0 | 30 |
| 12 | redWhistle | 11.86 |  | **34** | redBow | 0 | 33 |
| 13 | eyeball | 10.88 |  | **35** | plaidPencil | 0 | 34 |
| 14 | sign | 9.4 |  | **36** | foamDart | 0 | 35 |
| 15 | silverStraw | 8.64 |  | **37** | gClamp | 0 | 39 |
| 16 | cloudSign | 8.6 |  | **38** | brownDie | 0 | 43 |
| 17 | lavenderDie | 8.21 |  | **39** | legoBracelet | 0 | 52 |
| 18 | yellowBag | 7.57 |  | **40** | rubiksCube | 0 | 55 |
| 19 | hairClip | 6 |  | **41** | miniCards | 0 | 58 |
| 20 | birdCall | 5.75 |  | **42** | turtle | 0 | 60 |
| 21 | pinkCandle | 4.37 |  | **43** | rainbowPens | 0 | 86 |
| 22 | trophy | 0.41 |  |  |  |  |  |

**OG:**

**The objects classified poorly (23 → 43 in Fig. 1)**

For each of the 21 classes with zero detections:

* True Positive TP = 0, False Positive FP = 0
* False Negative ≥ 1

Hence, Precision = 0 and Recall = 0: These objects classified poorly.

We can consider penalty for classifier that miss out more objects: rainbowPens missed out the biggest number of objects (86) — that is the worst classifier, followed by turtle (60).

**The objects classified well (1 → 22 in Fig. 1)**

For the each class that has objects detected, we use Average Precision (AP) — precision averaged across all recall values between 0 and 1, to evaluate the performance. Then we compare AP between different object detectors to rank them. We will use the interpolation performed by PASCAL VOC challenge uses all data points. By interpolating all points, the Average Precision (AP) can be interpreted as an approximated area under the curve (AUC) of the Precision x Recall curve. Higher value means higher accuracy on detection.

The definition of True Positive and False Positive is subject to evaluator. In our case, TP equals to there is an object belong to that class within the bounding box, otherwise it is FP.

/// end of Q1