***Solution* 1 :**

1. Prove that the IoU metric between any two pairs of bounding boxes is always a non-negative real number in [0,1]

Let us consider we have two boxes A and B

$$IoUmetric = \frac{A \cap B}{A \cup B} \tag{1}$$

From the properties of sets we know that, $A \cup B = A + B - A \cap B$

$$\frac{A \cup B}{A \cup B} = \frac{A + B - A \cap B}{A \cup B}$$

$$\frac{A \cap B}{A \cup B} = \frac{A + B}{A \cup B} - 1 \tag{2}$$

$\boldsymbol{Case\ 1: \ When\ A\ and\ B\ are\ disjoint}$:

$$A \cap B = 0\ (Null).\ Hence,\ IoU = 0 \tag{3}$$

$\boldsymbol{Case\ 2: \ When\ A\ and\ B\ are\ slightly\ overlapping}$:

$$A + B > A \cup B$$

$$\frac{A + B}{A \cup B} > 1$$

$$\frac{A + B}{A \cup B} - 1 > 0$$

$$Substituting\ in\ (2),\ \frac{A \cap B}{A \cup B} > 0\ or\ IoU > 0 \tag{4}$$

$\boldsymbol{Case\ 3: \ When\ A \in B\ or\ B \in A}$:

$When\ A \in B,\ A \cup B = B$ and $A \cap B = A$

$$Since A < B, \frac{A \cap B}{A \cup B} < 1 \tag{5}$$

Similarly when, $B \in A,\ A \cup B = A$ and $A \cap B = B$

$$Since\ B < A, \frac{A \cap B}{A \cup B} < 1 \tag{6}$$

$\boldsymbol{Case\ 4: \ When\ A\ and\ B\ are\ totally\ overlapping}$:

In this case, if A = B, $A \cup B = A$ and $A \cap B = A$

$$\frac{A \cap B}{A \cup B} = 1 \tag{7}$$

Combining, (3), (4), (5), (6), (7),

$$0 <= IoU\ metric <= 1 \tag{8}$$

Hence, proved.

---

2. If we represent each bounding box as a function of the top-left and bottom-right coordinates (assume all coordinates are real numbers) then argue that the IoU metric is non-differentiable and hence cannot be directly optimized by gradient descent.

In order for a function to be differentiable, it must be continuous, and its derivative must be continuous too. Considering a case for 2 bounding boxes, where B is fixed and A is variable, we can have 4 different possibilities as mentioned in Part 1 of the Solution 1.

The IoU varies significantly and is not continuous.
Case 1: When A and B are disjoint - IoU = 0
Case 2: When A and B are slightly overlapping - IoU > 0
Case 3: When $A \in B$ or $B \in A$ - IoU < 1
Case 4: W hen A and B are totally overlapping - IoU = 1

Therefore, we can see that the function is not continuous (smooth) and hence is not differentiable and hence cannot be directly optimized by gradient descent.

***Solution* 2 :**

1. Calculate A, B, C, D, E, F

***A* = 30** - Since padding = same, image size remains constant
***B* = 30** - Since padding = same, image size remains constant
***C* = 16** - Same as number of filters applied
***D* = 30** - Dimensions must be similar to regular conv layer and also as padding = same, image size remains constant
***E* = 30** - Dimensions must be similar to regular conv layer and also as padding = same, image size remains constant
***F* = 64** - Same as number of filters applied

---

2. Computational cost of the regular conv layer -

$$(30 \times 30 \times 64) \times (5 \times 5 \times 200) = 288,000,000$$

Computational cost of bottleneck path -

$$(30 \times 30 \times 16) \times (1 \times 1 \times 200) + (30 \times 30 \times 64) \times (5 \times 5 \times 16) = 25,920,000$$

Ratio = 25,920,000/288,000,000 = 9/100 = 0.09 = 9%

Hence, using by using bottleneck, we reduce the computations to 9% of the regular conv layer

---

***Solution* 3 :**

In attached ipynb file.

Observations with COCO Keypoint Person Detector model with a ResNet50 − FPN base and following thresholds −

Threshold = 0.1 - The number of silhouettes observed were maximum, but it also misclassified other items as persons (like a couple of cotton candies on top). This can be reduced by increasing the threshold we can avoid these misclassifications.

Threshold = 0.2 - The observed misclassification had a confidence of 15% and hence by increasing the threshold to 20% we were able to omit such cases (for the provided picture).

Threshold = 0.5 - The number of silhouettes were same as before and the misclassifications were avoided as the threshold is high.

Threshold = 0.7 - The number of silhouettes were lower than before.

Threshold = 0.9 - The number of silhouettes were even lower than before.

Hence, I chose the threshold of 0.5 for the final output

Observations with Mask R − CNN model with ResNet50 − FPN and following thresholds

Threshold = 0.1 - The number of misclassifications were quite high like the cotton candies misclassified as umbrella and lamp as traffic light.

Threshold = 0.3 - The number of misclassifications were lower than before but still quite high like the cotton candies misclassified as umbrella.

Threshold = 0.5 - The number of misclassifications were still quite high but lost on correct classification like cycle

Threshold = 0.55 - The number of misclassifications were lower.

Threshold = 0.9 - The number of misclassifications were even lower than before but missed out on some correct classifications.

Hence, I chose the threshold of 0.55 for the final output although misclassification of balloons could not be avoided.

After training the balloons –

At threshold of 0.5, most of the balloons were classified and classified correctly.

References -

https://github.com/facebookresearch/detectron2
https://gilberttanner.com/blog/detectron-2-object-detection-with-pytorch
https://colab.research.google.com/drive/16jcaJoc6bCFAQ96jDe2HwtXj7BMD_-m5#scrollTo=ZyAvNCJMmvFF
