Object Detection – Evaluation Measures

Priyank Thakkar
Associate Professor
CSE Department
Institute of Technology
Nirma University

Mean Average Precision (mAP) [1]

- True Positive (TP): A correct detection. Detection with IOU ≥ threshold
- False Positive (FP): A wrong detection. Detection with IOU < threshold
- False Negative (FN): A ground truth not detected
- True Negative (TN): Does not apply. It would represent a corrected misdetection. In the object detection task there are many possible bounding boxes that should not be detected within an image. Thus, TN would be all possible bounding boxes that were correctly not detected (so many possible boxes within an image). That's why it is not used by the metrics.

Mean Average Precision (mAP) [1]

- Precision and Recall:

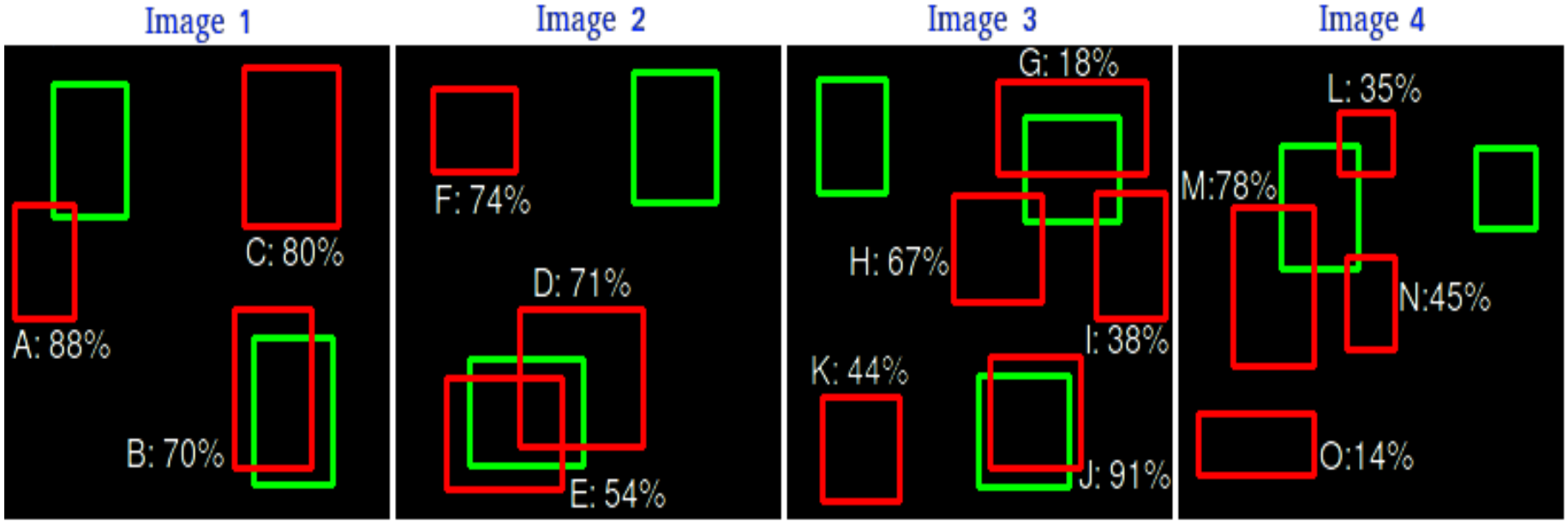
$$Precision = \frac{TP}{TP + FP} = \frac{TP}{\text{all detections}}$$

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{\text{all ground truths}}$$

- Average Precision
- mean Average Precision

Mean Average Precision (mAP) [1]

- TP or FP: (Note: Images with the detection and the ground truth of one specific class)

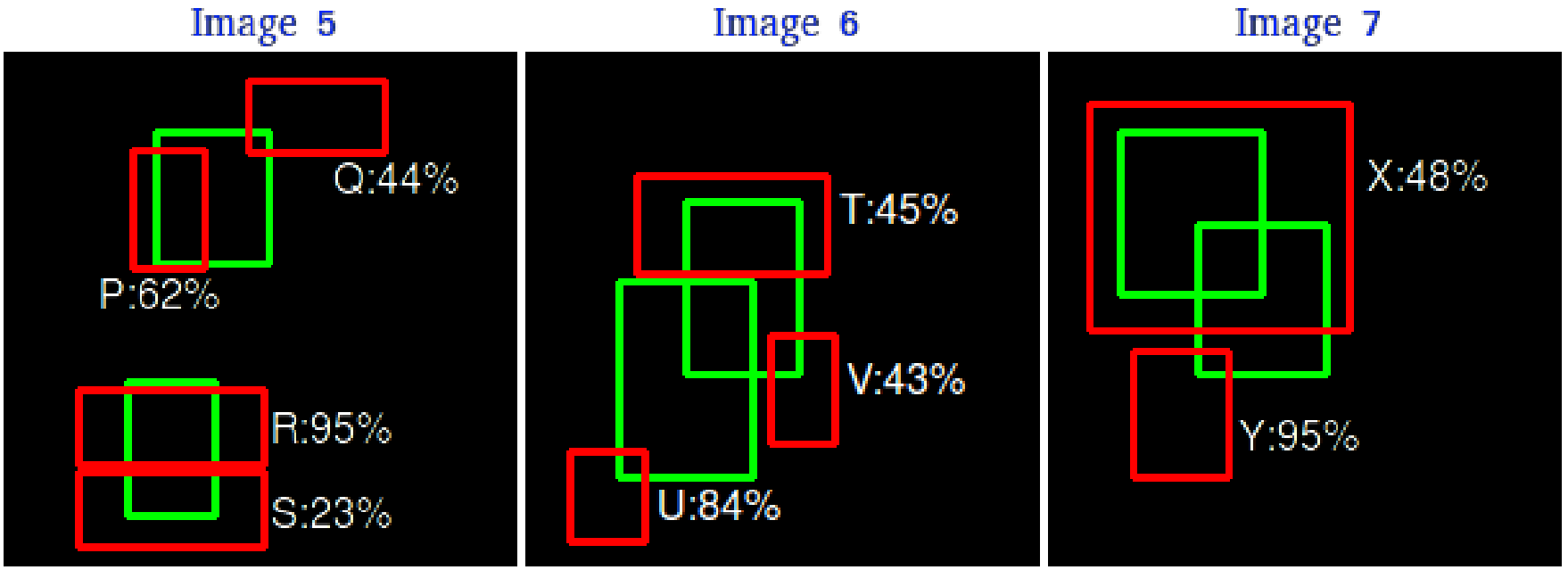


Images	Detections	Confidences	TP or FP
Image 1	A	88%	FP
Image 1	B	70%	TP
Image 1	C	80%	FP
Image 2	D	71%	FP
Image 2	E	54%	TP
Image 2	F	74%	FP
Image 3	G	18%	TP
Image 3	H	67%	FP

Images	Detections	Confidences	TP or FP
Image 3	I	38%	FP
Image 3	J	91%	TP
Image 3	K	44%	FP
Image 4	L	35%	FP
Image 4	M	78%	FP
Image 4	N	45%	FP
Image 4	O	14%	FP

Mean Average Precision (mAP) [1]

- Precision and Recall:



Images	Detections	Confidences	TP or FP
Image 5	P	62%	TP
Image 5	Q	44%	FP
Image 5	R	95%	TP
Image 5	S	23%	FP

Images	Detections	Confidences	TP or FP
Image 6	T	45%	FP
Image 6	U	84%	FP
Image 6	V	43%	FP
Image 7	X	48%	TP
Image 7	Y	95%	FP

Mean Average Precision (mAP) [1]

- Precision and Recall:

Images	Detections	Confidences	TP	FP	Acc TP	Acc FP	Precision	Recall
Image 5	R	95%	1	0	1	0	1	0.0666
Image 7	Y	95%	0	1	1	1	0.5	0.0666
Image 3	J	91%	1	0	2	1	0.6666	0.1333
Image 1	A	88%	0	1	2	2	0.5	0.1333
Image 6	U	84%	0	1	2	3	0.4	0.1333
Image 1	C	80%	0	1	2	4	0.3333	0.1333
Image 4	M	78%	0	1	2	5	0.2857	0.1333
Image 2	F	74%	0	1	2	6	0.25	0.1333
Image 2	D	71%	0	1	2	7	0.2222	0.1333
Image 1	B	70%	1	0	3	7	0.3	0.2
Image 3	H	67%	0	1	3	8	0.2727	0.2
Image 5	P	62%	1	0	4	8	0.3333	0.2666
Image 2	E	54%	1	0	5	8	0.3846	0.3333
Image 7	X	48%	1	0	6	8	0.4285	0.4

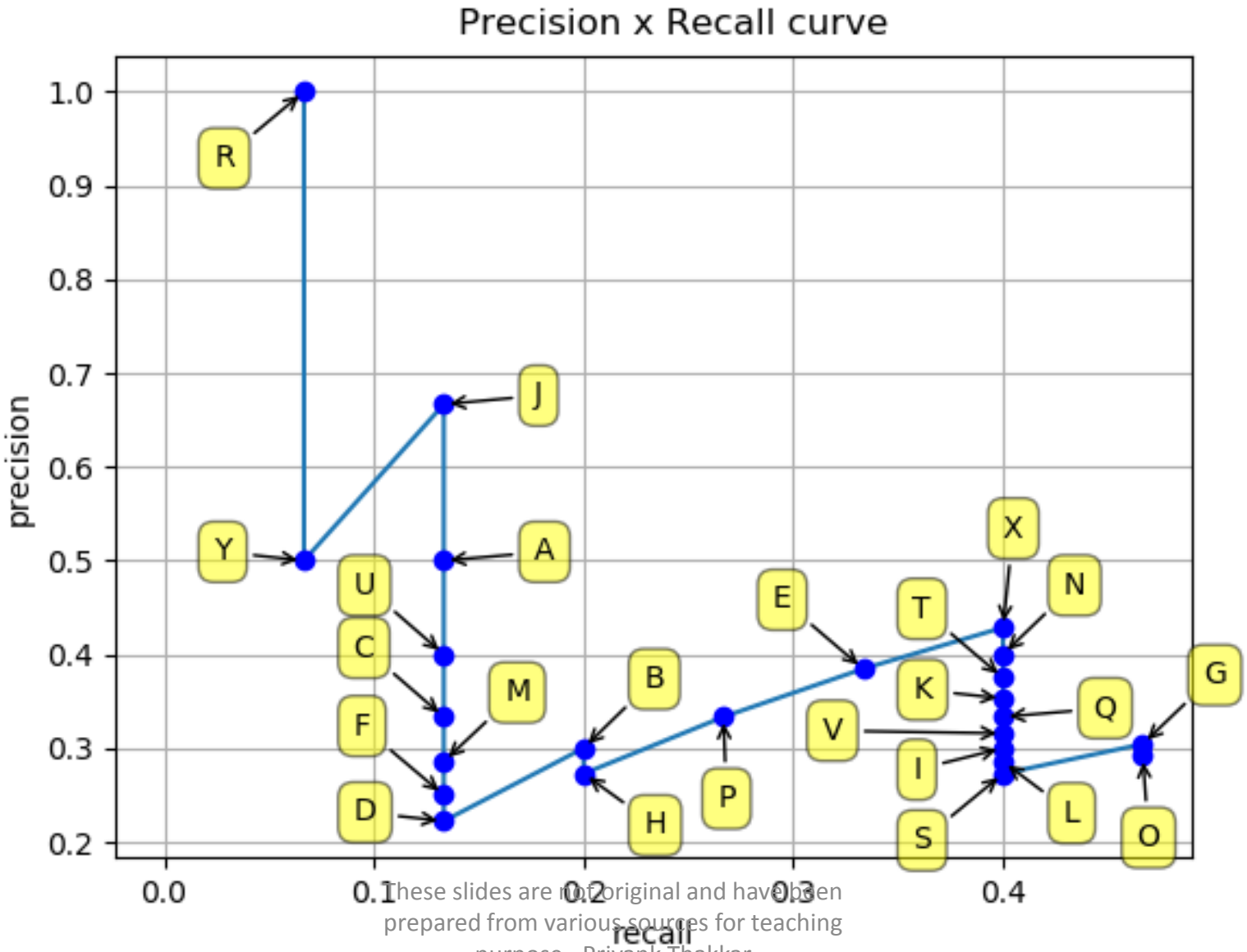
Mean Average Precision (mAP) [1]

- Precision and Recall:

Images	Detections	Confidences	TP	FP	Acc TP	Acc FP	Precision	Recall
Image 4	N	45%	0	1	6	9	0.4	0.4
Image 6	T	45%	0	1	6	10	0.375	0.4
Image 3	K	44%	0	1	6	11	0.3529	0.4
Image 5	Q	44%	0	1	6	12	0.3333	0.4
Image 6	V	43%	0	1	6	13	0.3157	0.4
Image 3	I	38%	0	1	6	14	0.3	0.4
Image 4	L	35%	0	1	6	15	0.2857	0.4
Image 5	S	23%	0	1	6	16	0.2727	0.4
Image 3	G	18%	1	0	7	16	0.3043	0.4666
Image 4	O	14%	0	1	7	17	0.2916	0.4666

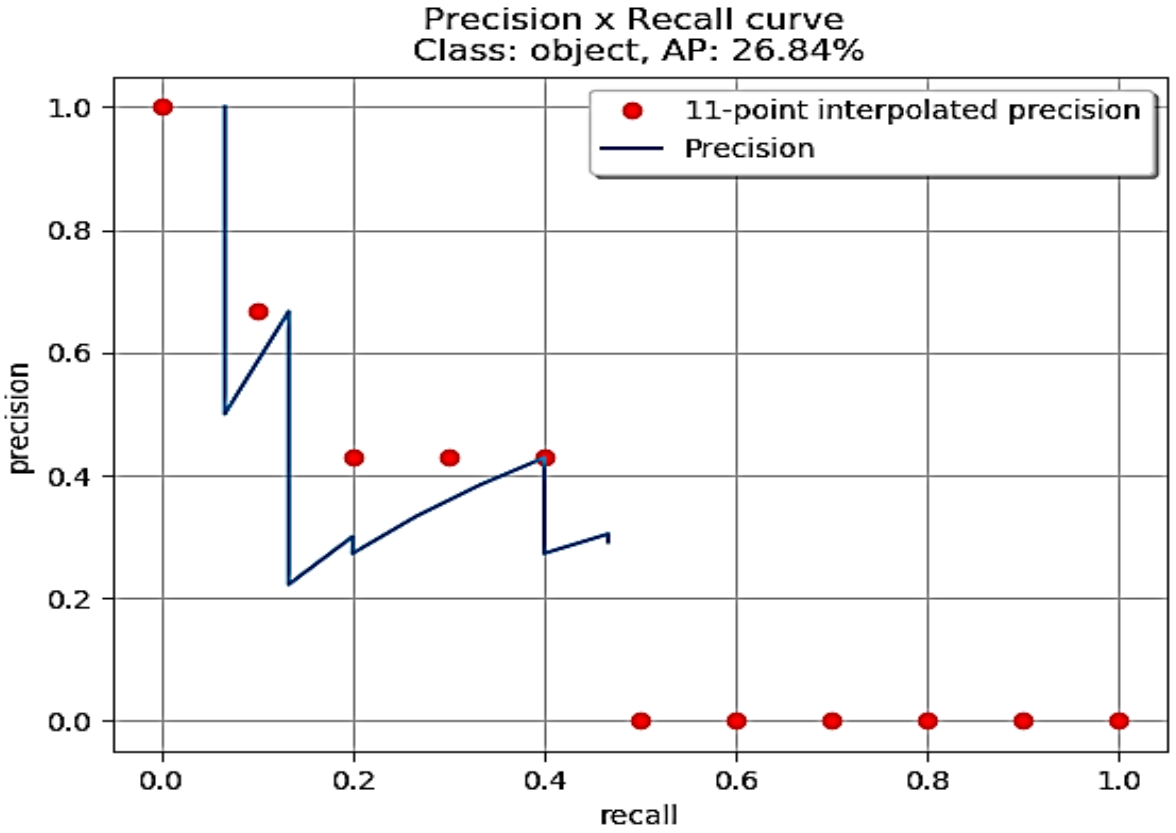
Mean Average Precision (mAP) [1]

- Precision vs. Recall Curve



Mean Average Precision (mAP) [1]

- Calculating the 11-point interpolation



By applying the 11-point interpolation, we have:

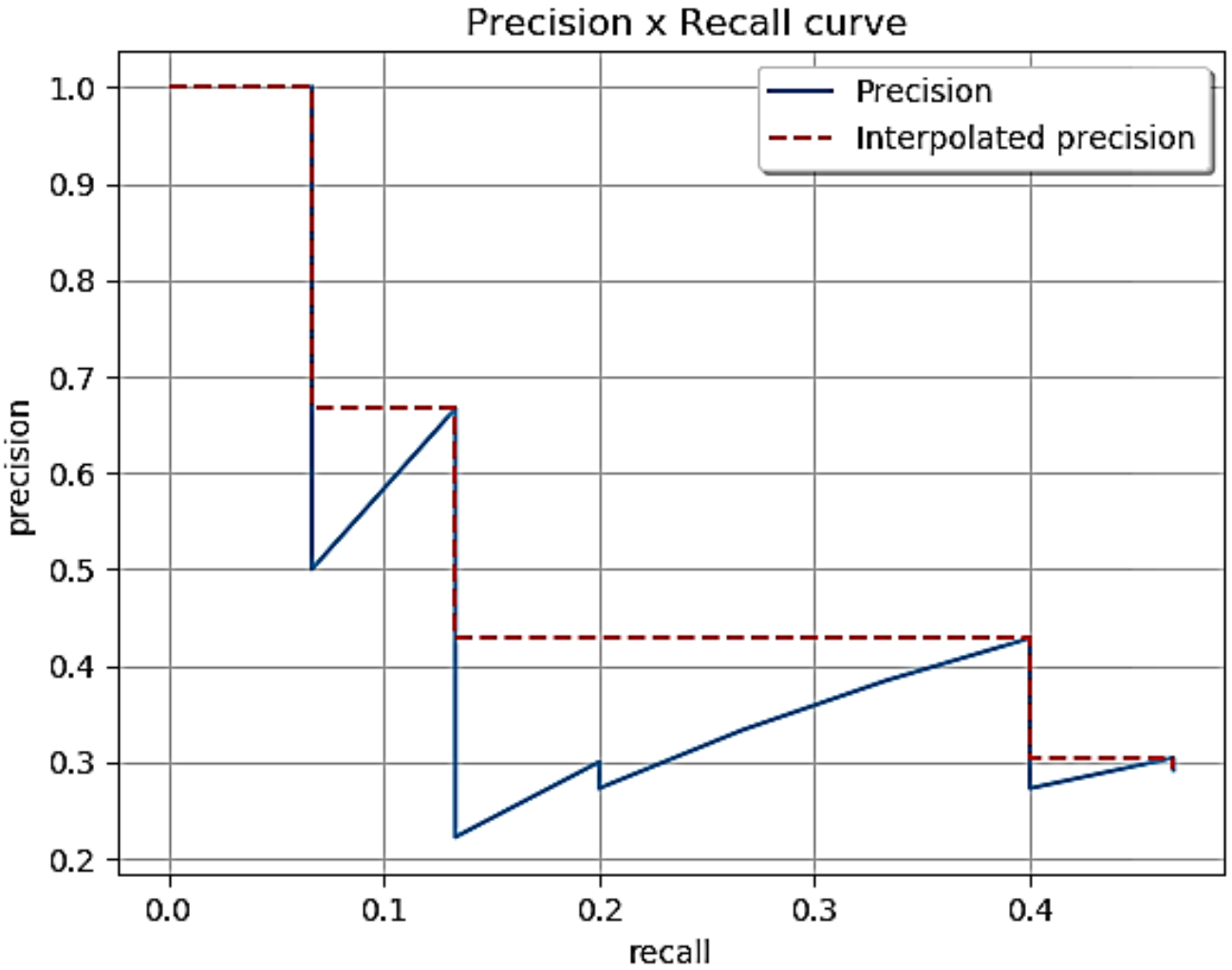
$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \ldots, 1\}} \rho_{\text{interp}(r)}$$

$$AP = \frac{1}{11} (1 + 0.6666 + 0.4285 + 0.4285 + 0.4285 + 0 + 0 + 0 + 0 + 0 + 0)$$

$$AP = 26.84\%$$

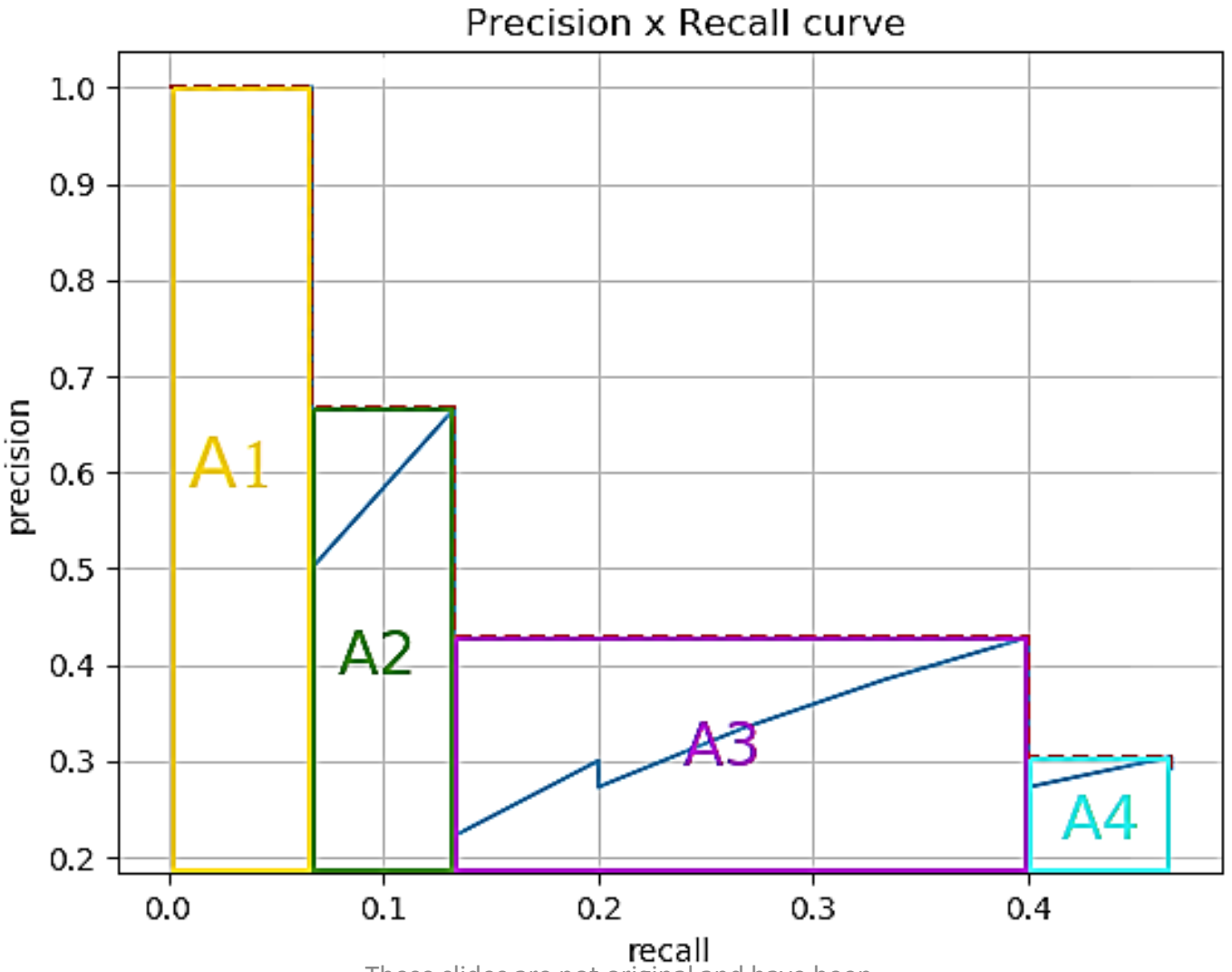
Mean Average Precision (mAP) [1]

- Calculating the interpolation performed in all points



Mean Average Precision (mAP) [1]

- Calculating the interpolation performed in all points



Mean Average Precision (mAP) [1]

- Calculating the interpolation performed in all points

Calculating the total area, we have the AP:

$$AP = A1 + A2 + A3 + A4$$

with:

$$A1 = (0.0666 - 0) \times 1 = \mathbf{0.0666}$$
$$A2 = (0.1333 - 0.0666) \times 0.6666 = \mathbf{0.04446222}$$
$$A3 = (0.4 - 0.1333) \times 0.4285 = \mathbf{0.11428095}$$
$$A4 = (0.4666 - 0.4) \times 0.3043 = \mathbf{0.02026638}$$

$$AP = 0.0666 + 0.04446222 + 0.11428095 + 0.02026638$$
$$AP = 0.24560955$$
$$AP = \mathbf{24.56\%}$$

Mean Average Precision (mAP) [1]

- Putting it all together
  - Now that we've defined Average Precision (AP) and seen how the IoU threshold affects it, the mean Average Precision or mAP score is calculated by taking the mean AP over all classes and/or over all IoU thresholds, depending on the competition. For example:
    - PASCAL VOC2007 challenge only 1 IoU threshold was considered: 0.5 so the mAP was averaged over all 20 object classes.
    - For the COCO 2017 challenge, the mAP was averaged over all 80 object categories and all 10 IoU thresholds.
  - Averaging over the 10 IoU thresholds rather than only considering one generous threshold of IoU ≥ 0.5 tends to reward models that are better at precise localization.

References

1. https://github.com/rafaelpadilla/Object-Detection-Metrics

2. https://medium.com/@timothycarlen/understanding-the-map-evaluation-metric-for-object-detection-a07fe6962cf3

Disclaimer

- These slides are not original and have been prepared from various sources for teaching purpose.