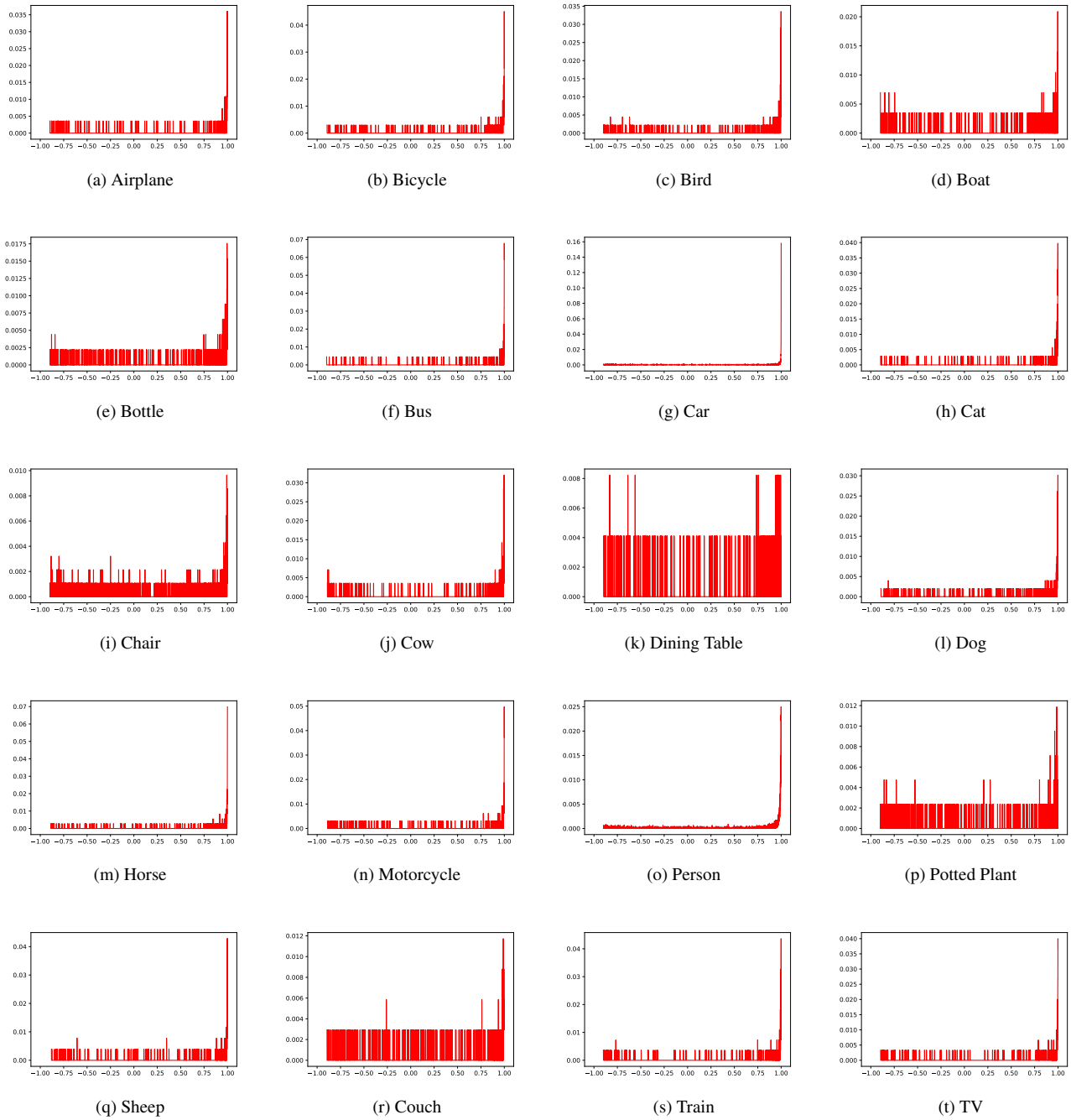
Figure 1: MISDETECTIONS FROM FASTER R-CNN *These results are from the Faster R-CNN network used in our paper. The network was trained on the PASCAL VOC 2007-2012 training and validation set while the above images belong to MSCOCO dataset. Ideally, the network was supposed to reject all the above detections as background.*

Figure 2: MISDETECTION FROM RETINANET *These results are from the RetinaNet network used in our paper. The network was trained on the PASCAL VOC 2007-2012 training and validation set while the above images belong to MSCOCO dataset. Ideally the network was supposed to reject all the above detections as background.*



(a) Airplane (b) Bicycle (c) Bird (d) Boat

(e) Bottle (f) Bus (g) Car (h) Cat

(i) Chair (j) Cow (k) Dining Table (l) Dog

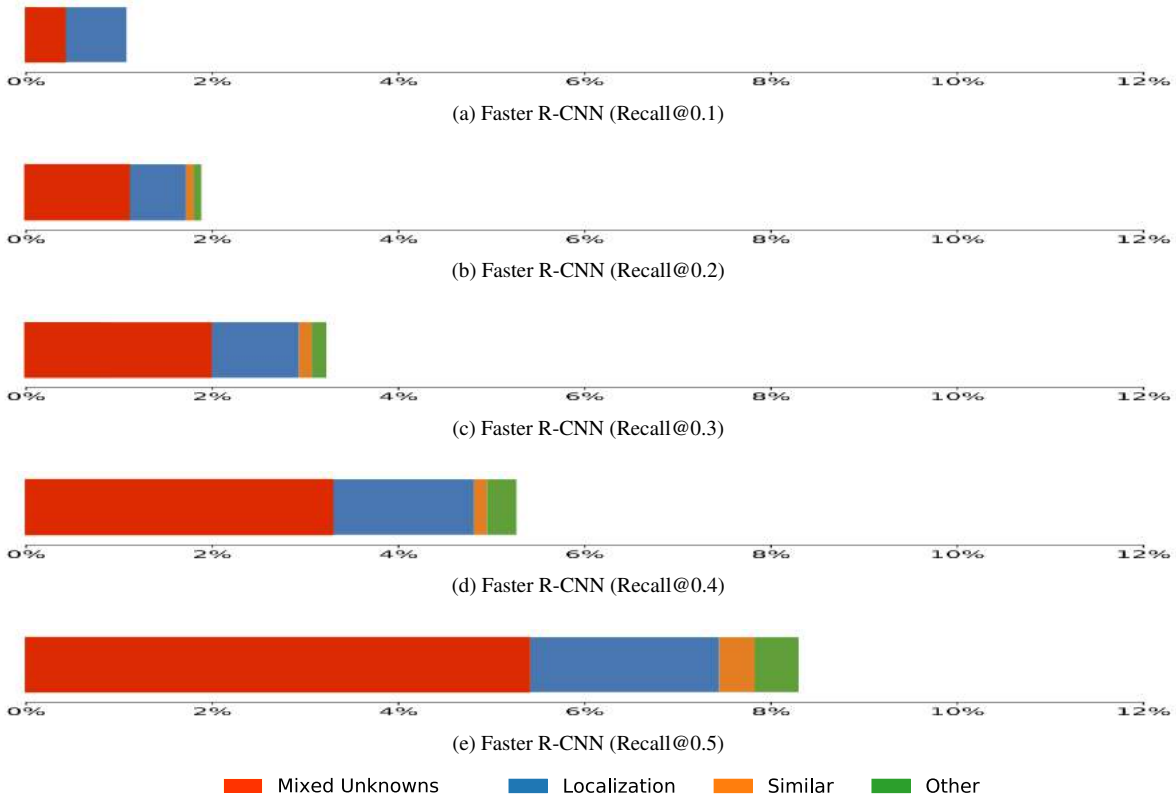
(m) Horse (n) Motorcycle (o) Person (p) Potted Plant

(q) Sheep (r) Couch (s) Train (t) TV

Figure 3: OBJECT VS BACKGROUND PROBABILITY DISTRIBUTION *One might think that for open set detections it would be sufficient to reject when the background class is greater than any other class. Unfortunately, this fails miserably. To see why, we compare the distribution of probability scores of correct detections (p_i) to background probability (p_b) for Faster R-CNN on the PASCAL VOC 2007 test set. This figure shows the histogram of ($p_i - p_b$), with the x-axis ranging from -1 ($p_i = 0, p_b = 1$) to 1 ($p_i = 1, p_b = 0$). As seen in the histograms, for a significant number of detections the difference is negative, i.e., the probability score for an ROI being background is higher than the probability for the actual object in the ROI. We also summarize the Mean and Standard Deviation values for each of the classes in Table 1. As explained in Section 2 of the main paper (Multi-Class Classifiers with Background), this skew indicates that we should not perform the evaluation as for a classification problem.*

Class	Mean	Standard Deviation
aeroplane	0.66	0.598
bicycle	0.6842	0.539
bird	0.6129	0.6284
boat	0.3873	0.6842
bottle	0.4747	0.6718
bus	0.6737	0.5411
car	0.6555	0.5947
cat	0.7638	0.4953
chair	0.2317	0.6881
cow	0.6331	0.6023
diningtable	0.264	0.6867
dog	0.7326	0.4882
horse	0.7319	0.5088
motorbike	0.6125	0.6161
person	0.6157	0.611
pottedplant	0.2321	0.7348
sheep	0.5874	0.6021
sofa	0.3421	0.6321
train	0.6825	0.5756
tvmonitor	0.653	0.5939

Table 1: OBJECT VS BACKGROUND PROBABILITY DISTRIBUTION



(a) Faster R-CNN (Recall@0.1)

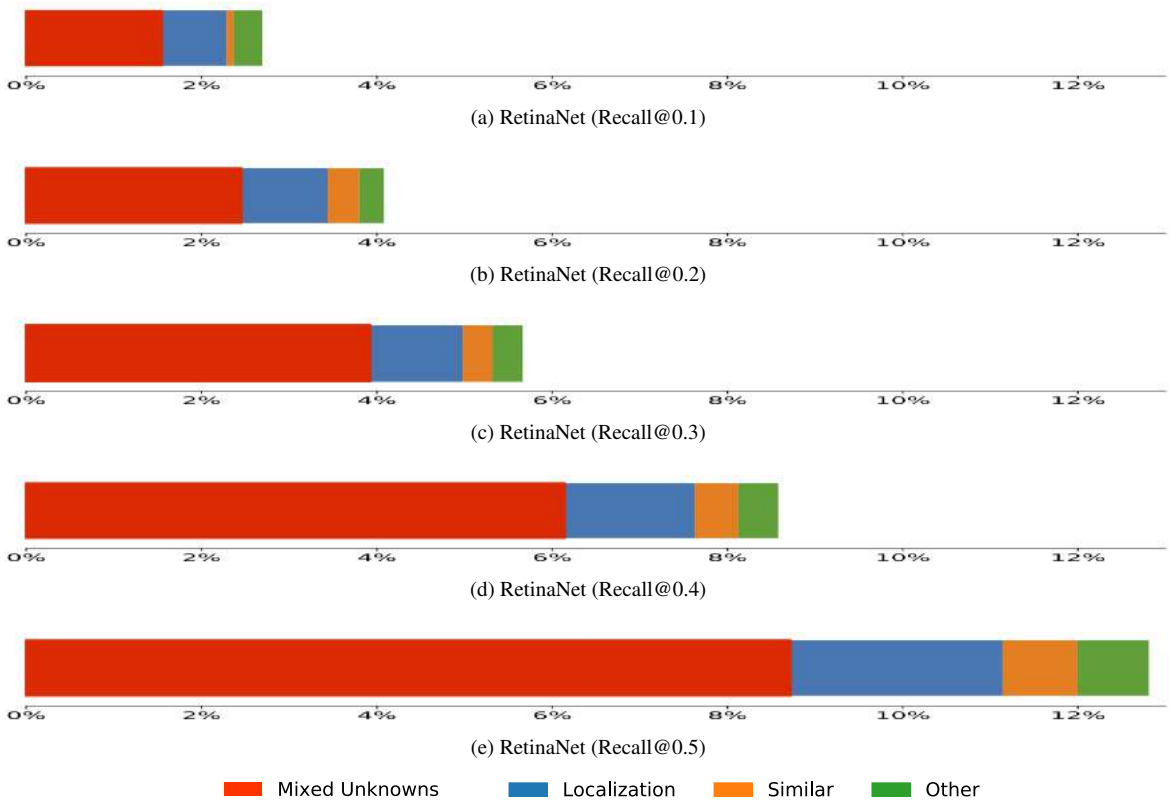
(b) Faster R-CNN (Recall@0.2)

(c) Faster R-CNN (Recall@0.3)

(d) Faster R-CNN (Recall@0.4)

(e) Faster R-CNN (Recall@0.5)

Figure 4: IMPACT OF UNKNOWNS AT WR1 ON FASTER RCNN *The errors are plot upto 12% the white region upto 100% representing correct detections. From the magnitude of the unknown unknowns it may be observed that the detectors rapidly confuse objects they were not trained to identify with known objects with an IOU* ≥ 0.1.



(a) RetinaNet (Recall@0.1)

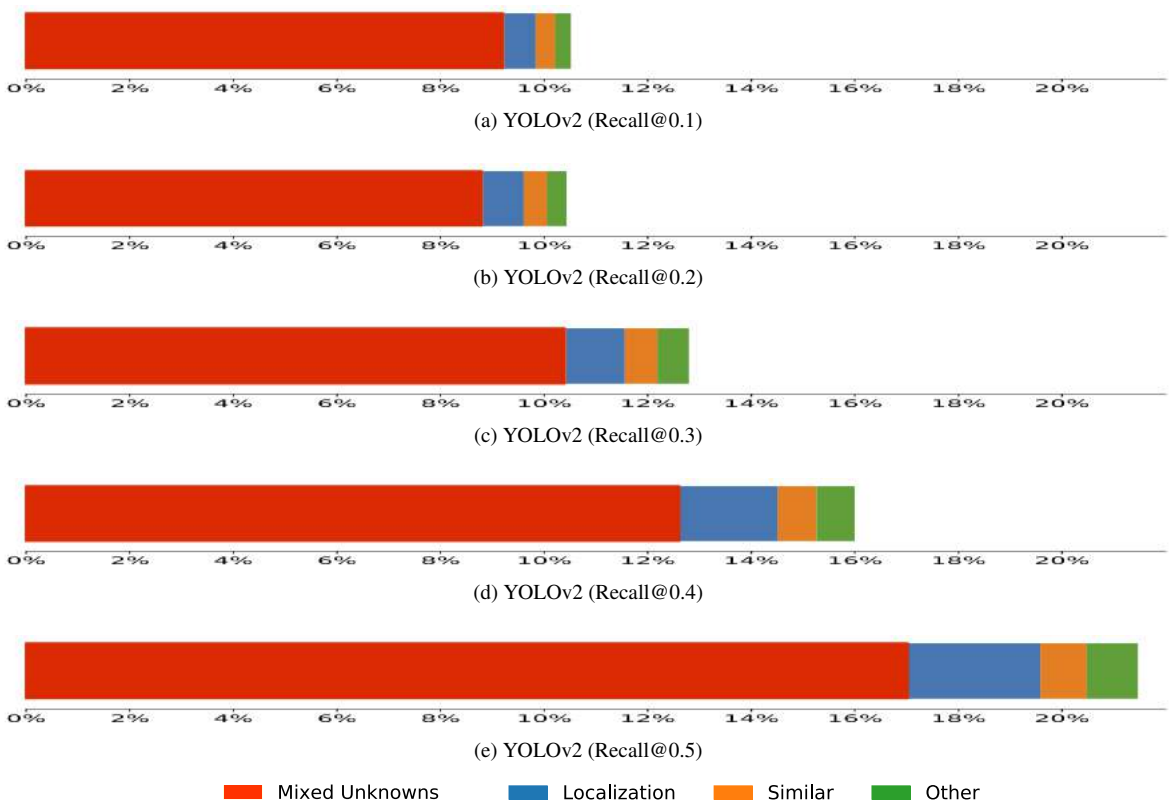
(b) RetinaNet (Recall@0.2)

(c) RetinaNet (Recall@0.3)

(d) RetinaNet (Recall@0.4)

(e) RetinaNet (Recall@0.5)

Figure 5: IMPACT OF UNKNOWNS AT WR1 ON RETINANET *The errors are plot upto 13% the white region upto 100% representing correct detections.*



(a) YOLOv2 (Recall@0.1)

(b) YOLOv2 (Recall@0.2)

(c) YOLOv2 (Recall@0.3)

(d) YOLOv2 (Recall@0.4)

(e) YOLOv2 (Recall@0.5)

Figure 6: IMPACT OF UNKNOWNS AT WR1 ON YOLOV2 *The errors are plot upto 22% the white region upto 100% representing correct detections.*

WR	β=0.5				β=1.0				β=1.5			
	F Score	Prec.	Recall	Thre.	F Score	Prec.	Recall	Thre.	F Score	Prec.	Recall	Thre.
0.0	0.85	87.91	77.46	0.62	0.83	79.25	86.07	0.49	0.84	78.52	86.89	0.48
0.25	0.81	87.38	73.77	0.65	0.80	87.38	73.77	0.65	0.81	69.51	86.89	0.48
0.5	0.75	84.39	70.90	0.69	0.77	81.08	73.77	0.65	0.77	65.19	84.43	0.51
0.75	0.68	82.23	66.39	0.72	0.74	76.89	70.90	0.69	0.75	68.98	77.46	0.62
1.0	0.65	70.87	73.77	0.65	0.72	70.87	73.77	0.65	0.73	65.85	77.46	0.62
1.25	0.63	71.19	70.90	0.69	0.71	71.19	70.90	0.69	0.72	62.17	77.46	0.62
1.5	0.61	75.36	65.16	0.73	0.70	75.36	65.16	0.73	0.71	64.75	73.77	0.65
1.75	0.58	71.30	65.16	0.73	0.68	71.30	65.16	0.73	0.69	60.40	73.77	0.65
2.0	0.57	69.43	65.16	0.73	0.67	69.43	65.16	0.73	0.68	57.88	73.77	0.65
2.25	0.54	72.06	60.25	0.77	0.66	67.83	63.93	0.74	0.67	59.45	70.90	0.69
2.5	0.52	69.34	60.25	0.77	0.64	69.34	60.25	0.77	0.65	55.45	70.90	0.69
2.75	0.49	65.04	60.25	0.77	0.63	65.04	60.25	0.77	0.64	51.95	70.90	0.69
3.0	0.49	64.47	60.25	0.77	0.62	64.47	60.25	0.77	0.63	50.44	70.90	0.69
3.25	0.47	62.82	60.25	0.77	0.62	62.82	60.25	0.77	0.62	57.45	64.75	0.74
3.5	0.46	61.51	60.25	0.77	0.61	61.51	60.25	0.77	0.62	56.23	64.75	0.74
3.75	0.45	59.51	60.25	0.77	0.60	59.51	60.25	0.77	0.61	54.30	64.75	0.74
4.0	0.44	58.57	60.25	0.77	0.59	58.57	60.25	0.77	0.61	52.84	64.75	0.74
4.25	0.42	55.89	60.25	0.77	0.58	56.64	59.43	0.77	0.60	50.64	64.75	0.74

Table 2: USING F_β TO SELECT OPERATING POINT FOR COW *We show various values of β that might be used to choose an operating point for RetinaNet in order to detect the cow class. For a provided set of detections, the threshold that provides the maximum value of F_β is chosen. For each such operating point we provide the threshold, F_β score, precision and recall values. The operating point may also be chosen at various levels of open-set conditions or wilderness ratios (WR). It may be observed from the above table that the thresholds selected for a specific value of β at a certain wilderness ratio do not generalize to other levels of wilderness.*