Animal Detection in Man-made Environments Supplementary

Abhineet Singh [1], Marcin Pietrasik [2], Gabriell Natha [2], Nehla Ghouaiel [2], Ken Brizel [2], and Nilanjan Ray [1]

[1]Department of Computing Science, University of Alberta
[2]Alberta Centre for Advanced MNT Products (ACAMP)

I. Introduction

This supplementary provides details of the tracking (Sec. II) and multi-model pooling experiments (Sec. III) along with the corresponding results. Class-agnostic Recall-Precision (cRP) for training configurations #1 - #5 (Table II) along with class-specific results for #1 and #2 (Tables III - VI) are also included. Finally, links are provided for more synthetic data samples generated using the four mask generation methods detailed in the paper (Table I).

A video demonstrating the end-to-end real and synthetic data generation pipeline created using the labeling tool accompanies this document. A higher resolution version of the same is also available on YouTube [2] along with mask generation results produced by several instance and semantic segmentation methods [1].

II. YOLO+DASiamRPN

Algorithm 1 details the process used for combining YOLO with DASiamRPN tracker to reduce false negatives using temporal information in videos. The *cumulative_score* in line 46 is a composite measure of overall tracker fitness defined as:

$$\textit{cumulative\_score} = \textit{cumulative\_confidence} \times \frac{\textit{assoc\_count} + 1}{\textit{unassoc\_count} + 1} \tag{1}$$

where *assoc_count* is the number of frames that the tracker has been successfully associated in since its initialization, *unassoc_count* is the number of frames since it was last associated and *cumulative_confidence* is the product of tracker confidence in all tracked frames so far as well as the confidence of the detection on which it was initialized.

Removing trackers that have remained unassociated for too long (lines 30-34) can help to increase precision by reducing tracked boxes corresponding to false positive detections. However, this also leads to a significant drop in recall since false negatives often occur in several consecutive frames which causes trackers corresponding to real objects to be removed. Thus, the overall performance, shown in Fig. 1, is nearly identical to not using tracking at all. Experiments were done with $\textit{max\_unassoc} \in \{2, 3, 4, 5\}$ and best results were obtained with $\textit{max\_unassoc} = 2$, which is the only one included here.

Algorithm 1 YOLO+DASiamRPN

```
1:  min_conf, min_iou, max_unassoc, max_trackers, min_init_gap ← params
2:  trackers ← ∅
3:  for frame in video_sequence do
4:      detections ← YOLO (frame)
5:      associated_detections ← ∅
6:      associated_trackers ← ∅
7:      for t in trackers do
8:          t.update(frame)
9:          if t.confidence < min_conf then
10:             trackers ← trackers \ t
11:             continue
12:         end if
13:         if t in associated_trackers then
14:             continue
15:         end if
16:         for d in detections do
17:             if d in associated_detections then
18:                 continue
19:             end if
20:             if d.class = t.class and iou(d, t) > min_iou then
21:                 associated_detections ← associated_detections ∪ d
22:                 associated_trackers ← associated_trackers ∪ t
23:                 t.unassoc_count ← 0
24:                 t.assoc_count ← t.assoc_count + 1
25:                 continue
26:             end if
27:         end for
28:     end for
29:     unassociated_trackers ← trackers \ associated_trackers
30:     for t in unassociated_trackers do
31:         t.unassoc_count ← t.unassoc_count + 1
32:         if t.unassoc_count > max_unassoc then
33:             trackers ← trackers \ t
34:         end if
35:     end for
36:     unassociated_detections ← detections \ associated_detections
37:     tracker_init_gap ← tracker_init_gap + 1
38:     if tracker_init_gap > min_init_gap then
39:         tracker_init_gap ← 0
40:         for d in unassociated_detections do
41:             t ← initialize_new_tracker(frame, d)
42:             t.unassoc_count ← 0
43:             t.assoc_count ← 1
44:             trackers ← trackers ∪ t
45:         end for
46:     end if
47:     if | trackers | > max_trackers then
48:         trackers ← sort trackers by cumulative_score (Eq. 1)
49:         trackers ← {t_i | t_i ∈ trackers, 1 ≤ i ≤ max_trackers}
50:     end if
51: end for
```

III. Muti-Model Pooling

Muti-model pooling can be an effective way to reduce false negatives only if the pooled models exhibit distinct patterns of missing objects. As shown in Sec. 4.2 of the main paper, model recall depends strongly on the range of backgrounds in the training set. It would thus seem that model separation can best be achieved by training them on images with significantly different backgrounds, e.g. with snowy winter and green summer scenes or under sunny and cloudy / rainy weather conditions. However, the existing labeled data does not contain sufficient

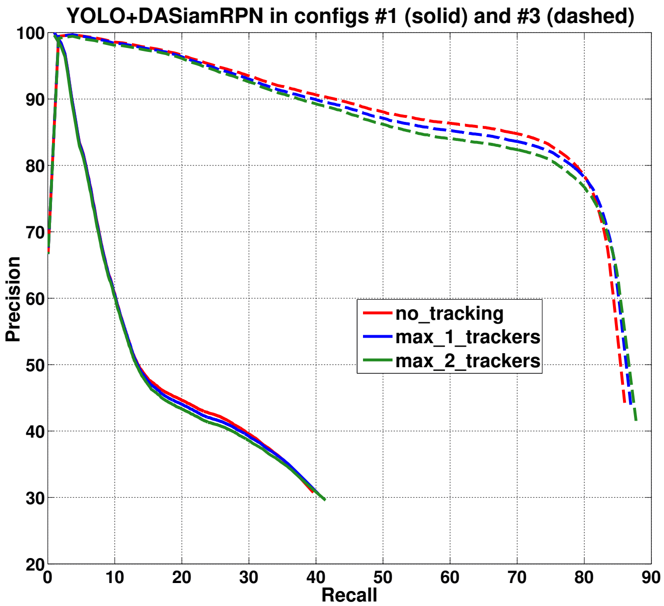

Fig. 1: Mean Recall vs Precision for DASiamRPN + YOLO in configurations #1 (solid) and #3 (dashed) using unassociated tracker removal with $max\_unassoc = 2$

background	animal			
airport	bear	coyote	deer	moose
highway	bear	coyote	deer	moose

TABLE I: Google Photos albums showing samples of synthetic data generated using all 4 types of mask generation methods - (left to right in each row) manual masks, Mask RCNN, SiamMask and no mask/Gaussian blending,

variety to divide it into class-balanced subsets that each have relatively homogeneous backgrounds differing significantly from those in other subsets. Model separation was instead achieved by training different architectures on the same data.

Two pooling strategies were employed:

1) Model Switching: All detections from the model with the highest confidence detection are retained in each frame while all others models' detections are discarded.
2) Model Aggregation: Detections from all models are pooled followed by class-agnostic non-maximum suppression applied to boxes from different models.

Both strategies require inter-model confidence comparison which is not straightforward to do since, as shown in Fig. 3 of the main paper (and Tables III - VI here), the meaning of confidence magnitude in terms of detection accuracy varies significantly between models. As a result, raw confidence values from different models need to be normalized before they can be compared. Several ways of doing this were explored but the one that worked best was to scale the confidence values of each model so that its mRP threshold becomes 0.5 while the range of values remains [0, 1].

Experiments were done by pooling YOLO and RETINA as well as YOLO, SSDIN and SSDMO but the former combination gave much better results so only these are included here. Similarly, configurations #1, #3 and #5 were all tested but only #5 is included as it produced the best pooling results relative to the individual detectors. Fig. 2 gives results for the two strategies in #5, both with and without confidence normalization. Somewhat surprisingly, none of the normalization techniques managed to outperform unnormalized confidence based pooling. Also, none of the pooling strategies managed to the increase recall over the better of the pooled detectors while always leading to a significant drop in precision. This shows that pooling is not a practicle way to improve recall, at least not with the relatively simple methods employed here so as not to compromise speed.

IV. Class-agnostic Recall-Precision (cRP)

Table II shows the cRPs for all models in configurations #1 - #5. It is interesting to note that the large relative increase in cRP exhibited by YOLO in #1 and #3 is not present in the other easier configurations. This might indicate that YOLO manifests its tendency to overfit to specific background-foreground combinations most strongly with very limited training data.

V. Class-Specific Results

Results for configuration #1 are in tables III and IV while those for #2 are in tables V and VI. The second column gives the recall – precision (RP) value corresponding to the score threshold where the two are equal while the threshold itself is given in the third column. The next three columns provide the recall, precision and their average for the score threshold corresponding to the overall mRP, i.e. where the mean recall and mean precision are equal. The last column gives the total number of ground truth boxes available for each class. Note that many images had multiple objects so that this count is more than the number of images in the corresponding test set.

It can be seen that the large inter-class variation of RP thresholds holds for all the models as well as for both low and high performance scenarios, represented by #1 and #2 respectively. AP and RP show considerable variation too, though only in #1. This is probably because animals with the larger ranges of backgrounds in the training set are often falsely detected in test images of remaining animals, thus leading to significant biases in their recall and precision.

VI. Synthetic data samples

Table I provides links to Google Photos albums showing synthetic data samples generated using all 4 methods - (left to right in each row) manual masks, Mask RCNN, SiamMask and no masks.

References

[1] A. Singh. Animal mask generation with segmentation methods. online: https://www.youtube.com/playlist?list=PLYIOwkVj6L8hErKjWOGtEDpbWZVq1BuRb, January 2020.
[2] A. Singh. Labeling tool demo. online: https://youtu.be/ZkjcP8s0QVQ, January 2020.

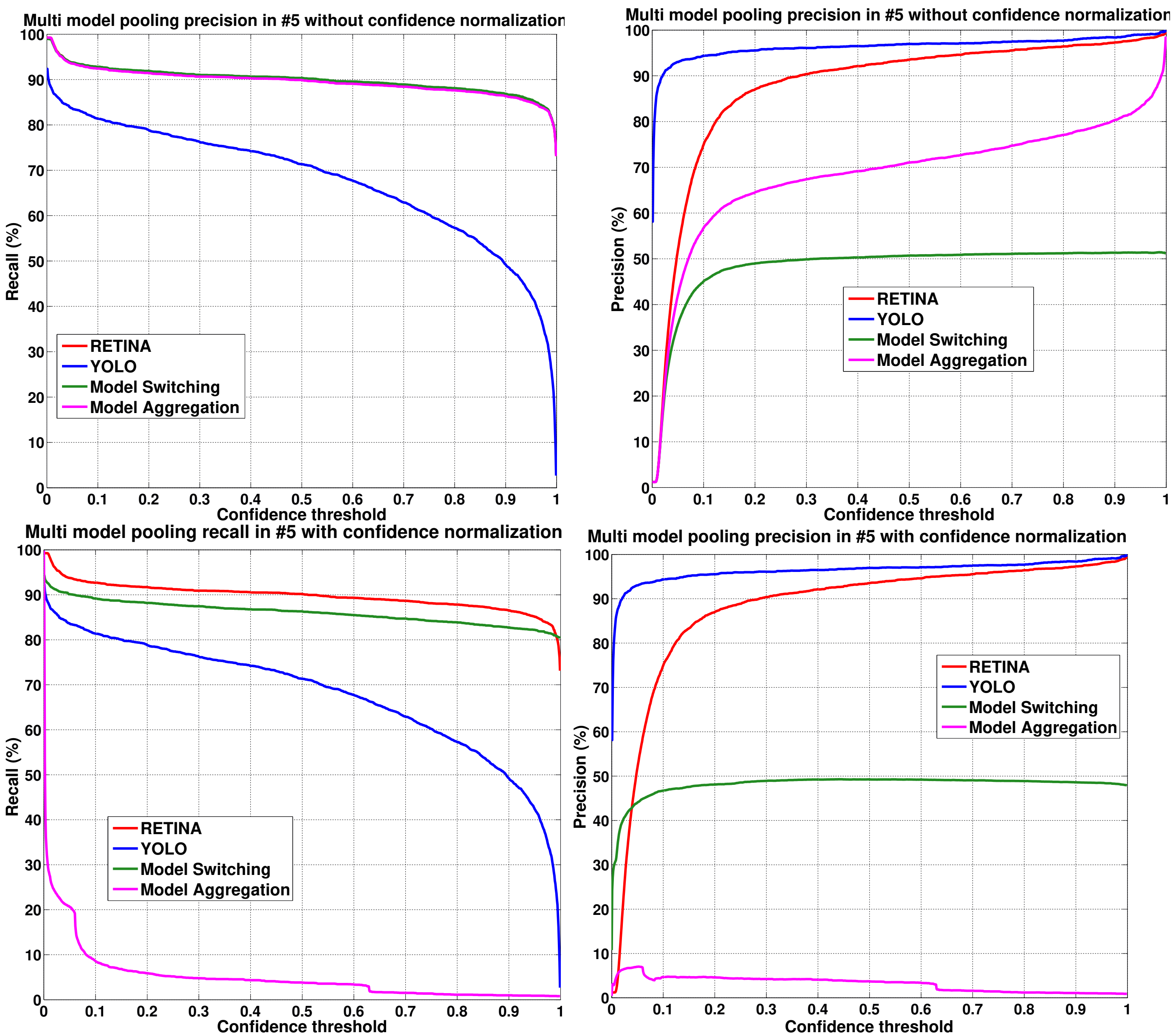

Fig. 2: Mean Recall and Precision for RETINA and YOLO pooling in #5 with and without confidence normalization

Detection cRP (%) for configuration #1 - #5					
Model	**1: 1K/Seq/8**	**2: 1K/Even/8**	**3: 10K/Seq/6**	**4: 5%/Start/6**	**5: 500/Static/3**
NAS	87.17	96.42	93.8	92.98	95.83
INRES	84.64	96.13	95.08	94.32	92.73
RES101	89.71	99.65	95.16	97.15	95.93
RFCN	88.93	99.54	94.64	97.27	93.88
RETINA	74.62	99.62	90.18	95.27	92.01
SSDIN	69.56	98.62	84.75	92	89.56
SSDMO	64.89	98.42	86.28	92.84	90.62
YOLO	80.76	96.67	91.7	94.16	90.36

TABLE II: Class agnostic Recall-Precision (cRP) for training configurations #1 - #5

Training Configuration #1: 1K/Seq/8							
NAS							
	Class Specific			**mRP threshold 62.00 %**			
class	**AP(%)**	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	**GT**
bear	79.01	72.81	86.84	78.55	60.97	69.76	25320
bison	72.07	65.47	31.08	55.69	80.55	68.12	31612
cow	42.89	42.99	98.6	60.16	10.73	35.45	4222
coyote	75.33	67.87	16.66	52.53	94.45	73.49	22337
deer	63.89	59.88	30.88	49.29	67.3	58.3	24465
elk	65.77	57.19	3.45	36.53	89.99	63.26	25353
horse	71.33	65.58	94.89	80.64	41.41	61.03	4840
moose	50.31	47.36	97.2	65.56	33.14	49.35	25033
average	**65.07**	**59.89**	**57.45**	**59.87**	**59.82**	**59.84**	**163182**
INRES							
	Class Specific			**mRP threshold 38.70 %**			
class	**AP(%)**	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	**GT**
bear	68.02	67.09	16	59.49	75.72	67.61	25320
bison	61.51	57.97	15.8	52.58	66.44	59.51	31612
cow	26.19	29.58	98.77	48.48	6.24	27.36	4222
coyote	67.31	61.38	3.49	50.83	90.74	70.79	22337
deer	67.34	64.49	58.18	68.49	61.02	64.75	24465
elk	59.12	54.26	4.5	42.42	65.98	54.2	25353
horse	50.11	49.13	96.45	62.95	22.22	42.59	4840
moose	49.53	49.04	47.33	50.46	47.36	48.91	25033
average	**56.14**	**54.12**	**42.57**	**54.47**	**54.46**	**54.46**	**163182**
RES101							
	Class Specific			**mRP threshold 65.80 %**			
class	**AP(%)**	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	**GT**
bear	70.95	67.76	98.52	84.5	42.27	63.38	25320
bison	56.19	55.56	49.13	48.64	58.46	53.55	31612
cow	44.79	45.33	99.54	63.26	14.59	38.93	4222
coyote	69.19	63.72	35.17	55.63	75.72	65.68	22337
deer	74.95	69.86	64.2	69.45	70.32	69.88	24465
elk	63.21	53.34	6.36	42.46	76.87	59.67	25353
horse	52.22	47.54	33.88	42.02	62.34	52.18	4840
moose	57.1	52.86	57.77	50.45	55.46	52.96	25033
average	**61.07**	**57**	**55.57**	**57.05**	**57**	**57.03**	**163182**
RFCN							
	Class Specific			**mRP threshold 60.10 %**			
class	**AP(%)**	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	**GT**
bear	66.95	67.8	73.86	73.36	62.64	68	25320
bison	59.79	57.16	42.37	50.14	62.48	56.31	31612
cow	39.19	43.98	99.91	50.09	7.04	28.57	4222
coyote	57.38	52.65	16.13	35.78	86.97	61.38	22337
deer	68.22	65.28	72.45	69.97	59.87	64.92	24465
elk	56.99	52.48	21.35	42.43	72.3	57.36	25353
horse	48.14	43.39	94.48	59.15	24.95	42.05	4840
moose	48.24	47.53	54.37	45.48	49.81	47.64	25033
average	**55.61**	**53.78**	**59.36**	**53.3**	**53.26**	**53.28**	**163182**

TABLE III: Training configuration #1 class-specific results for NAS, INRES, RES101 and RFCN

Training Configuration #1: 1K/Seq/8							
RETINA							
class	**Class Specific**			**mRP threshold 41.20 %**			**GT**
	AP(%)	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	
bear	50.16	50.86	36.15	47.53	52.61	50.07	25320
bison	53.61	51.01	28.55	45.93	59.69	52.81	31612
cow	14.22	18.95	97.88	46.31	6.66	26.48	4222
coyote	43.94	42.57	15.13	32.99	71.96	52.47	22337
deer	47.7	47.82	53.13	53.83	42.61	48.22	24465
elk	47.46	46.71	17.82	36.56	63.72	50.14	25353
horse	36.16	38.24	84.68	51.01	14.08	32.55	4840
moose	25.73	27	36.93	24.89	27.53	26.21	25033
average	**39.87**	**40.4**	**46.28**	**42.38**	**42.36**	**42.37**	**163182**
SSDIN							
class	Class Specific			mRP threshold 11.30 %			**GT**
	AP(%)	RP(%)	Score(%)	Recall(%)	Precision(%)	Average(%)	
bear	48.2	49.42	3.06	38.59	58.53	48.56	25320
bison	39.95	46.49	10.54	45.95	46.85	46.4	31612
cow	28.56	32.64	99.6	69.23	7.09	38.16	4222
coyote	40.68	44.75	1.73	35.81	66.03	50.92	22337
deer	37.36	41.02	2.63	34.08	54.78	44.43	24465
elk	38.71	42.8	1.13	30.12	65.25	47.69	25353
horse	46.31	46.28	85.14	59.52	16.62	38.07	4840
moose	23.46	27.77	15.25	28.87	26.63	27.75	25033
average	**37.9**	**41.4**	**27.39**	**42.77**	**42.72**	**42.75**	**163182**
SSDMO							
class	**Class Specific**			**mRP threshold 27.30 %**			**GT**
	AP(%)	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	
bear	54.58	54.29	22.15	52.66	56.2	54.43	25320
bison	39.8	43.17	30.05	43.85	42.56	43.2	31612
cow	28.22	33.73	98.99	52.7	12.92	32.81	4222
coyote	42.28	43.4	3.9	35.56	57.13	46.35	22337
deer	38.14	41.86	65.72	49.43	36.83	43.13	24465
elk	40.22	41.73	1.18	27.58	73.55	50.56	25353
horse	38.47	38.04	96.24	49.09	20.16	34.62	4840
moose	26.45	29.56	7.31	24.41	35.58	29.99	25033
average	**38.52**	**40.72**	**40.69**	**41.91**	**41.87**	**41.89**	**163182**
YOLO							
class	**Class Specific**			**mRP threshold 0.50 %**			**GT**
	AP(%)	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	
bear	19.3	31.12	1.93	34.36	29.69	32.02	25320
bison	18.07	39.39	0.1	27.62	52.2	39.91	31612
cow	13.39	16.32	87.34	45.1	5.01	25.05	4222
coyote	17.57	37.74	0.1	22	58.65	40.33	22337
deer	28.71	44.69	2.69	50.23	40.31	45.27	24465
elk	27.98	38.62	0.1	33.72	45.72	39.72	25353
horse	17.16	22.71	91.92	50.87	7.89	29.38	4840
moose	13.65	30.63	0.1	20.56	41.89	31.22	25033
average	**19.48**	**32.65**	**23.03**	**35.56**	**35.17**	**35.36**	**163182**

TABLE IV: Training configuration #1 class-specific results for RETINA, SSDIN, SSDMO and YOLO

Training Configuration #2: 1K/Even/8							
NAS							
class	Class Specific			mRP threshold 83.10 %			GT
	AP(%)	RP(%)	Score(%)	Recall(%)	Precision(%)	Average(%)	
bear	97.78	93.94	93.93	95.12	90.44	92.78	25268
bison	97.53	94.44	59.38	92.85	97.57	95.21	31294
cow	97.03	94.93	45.74	92.76	98.7	95.73	4340
coyote	96.96	91.61	90.7	93.11	87.62	90.36	22300
deer	97.89	93.54	66.8	91.72	95.9	93.81	24374
elk	98.5	95.17	86.24	95.51	94.66	95.08	25292
horse	97.82	95.71	91.63	96.45	92.95	94.7	4759
moose	98.36	95.15	84.48	95.25	94.91	95.08	24943
average	**97.73**	**94.31**	**77.36**	**94.1**	**94.09**	**94.09**	**162570**
INRES							
class	Class Specific			mRP threshold 12.20 %			GT
	AP(%)	RP(%)	Score(%)	Recall(%)	Precision(%)	Average(%)	
bear	97.37	96.37	3.03	95.7	99.05	97.37	25268
bison	92.13	90.73	3.37	90.31	96.07	93.19	31294
cow	98.62	98.23	20.62	98.32	96.3	97.31	4340
coyote	98.67	96.75	24.83	97.47	94.66	96.07	22300
deer	98.53	97.03	38.24	97.92	93.09	95.5	24374
elk	98.13	97.2	10.21	97.13	97.69	97.41	25292
horse	90.18	90.1	3.77	90	97.45	93.72	4759
moose	99.5	98.27	69.91	99.21	91.52	95.36	24943
average	**96.64**	**95.59**	**21.75**	**95.76**	**95.73**	**95.74**	**162570**
RES101							
class	Class Specific			mRP threshold 90.80 %			GT
	AP(%)	RP(%)	Score(%)	Recall(%)	Precision(%)	Average(%)	
bear	99.72	99.53	88.64	99.52	99.59	99.55	25268
bison	99.79	99.42	86.91	99.35	99.53	99.44	31294
cow	99.86	99.65	89.18	99.65	99.72	99.69	4340
coyote	99.92	99.34	92.52	99.39	99.31	99.35	22300
deer	99.87	99.5	96.2	99.57	99.39	99.48	24374
elk	99.94	99.84	75.47	99.81	99.92	99.86	25292
horse	99.77	99.62	98.57	99.66	99.23	99.44	4759
moose	99.91	99.6	75.28	99.49	99.75	99.62	24943
average	**99.85**	**99.56**	**87.85**	**99.56**	**99.55**	**99.56**	**162570**
RFCN							
class	Class Specific			mRP threshold 84.30 %			GT
	AP(%)	RP(%)	Score(%)	Recall(%)	Precision(%)	Average(%)	
bear	99.65	99.44	79.47	99.39	99.58	99.49	25268
bison	99.63	99.09	79.33	98.99	99.23	99.11	31294
cow	99.88	99.72	49.54	99.65	99.86	99.76	4340
coyote	99.95	99.48	89.86	99.57	99.31	99.44	22300
deer	99.86	99.61	73.09	99.48	99.71	99.6	24374
elk	99.86	99.75	91.57	99.79	99.53	99.66	25292
horse	99.74	99.58	95.96	99.66	99.23	99.44	4759
moose	99.86	99.56	81.89	99.52	99.59	99.56	24943
average	**99.8**	**99.53**	**80.09**	**99.51**	**99.5**	**99.51**	**162570**

TABLE V: Training configuration #2 class-specific results for NAS, INRES, RES101 and RFCN

Training Configuration #2: 1K/Even/8							
RETINA							
class	**Class Specific**			**mRP threshold 28.30 %**			**GT**
	AP(%)	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	
bear	99.88	99.66	29.67	99.68	99.64	99.66	25268
bison	99.91	99.34	25.21	99.13	99.44	99.28	31294
cow	100	99.82	52.52	99.95	99.54	99.75	4340
coyote	99.99	99.64	42.03	99.96	99.39	99.67	22300
deer	99.98	99.67	30.56	99.72	99.64	99.68	24374
elk	99.99	99.8	38.48	99.89	99.75	99.82	25292
horse	99.89	99.41	26.15	98.76	99.6	99.18	4759
moose	99.96	99.72	27.8	99.7	99.74	99.72	24943
average	**99.95**	**99.63**	**34.05**	**99.6**	**99.59**	**99.6**	**162570**
SSDIN							
class	**Class Specific**			**mRP threshold 22.30 %**			**GT**
	AP(%)	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	
bear	99.63	98.36	34.83	98.97	97.3	98.14	25268
bison	99.79	99.18	13.43	98.87	99.6	99.23	31294
cow	99.82	98.43	27.94	98.71	97.92	98.31	4340
coyote	99.65	97.54	32.92	98.37	96.24	97.3	22300
deer	99.88	98.89	13.25	98.23	99.46	98.84	24374
elk	99.96	99.36	13.82	99.1	99.79	99.45	25292
horse	99.78	98.72	19.4	98.59	99.09	98.84	4759
moose	99.76	98.55	13.25	97.67	99.03	98.35	24943
average	**99.78**	**98.63**	**21.1**	**98.56**	**98.55**	**98.56**	**162570**
SSDMO							
class	**Class Specific**			**mRP threshold 34.10 %**			**GT**
	AP(%)	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	
bear	99.64	98.56	19.39	97.7	99.26	98.48	25268
bison	99.72	98.37	49.9	98.83	97.1	97.96	31294
cow	99.88	98.55	46.9	99.24	96.55	97.89	4340
coyote	99.87	98.52	35.81	98.67	98.41	98.54	22300
deer	99.9	98.81	27.96	98.35	99.11	98.73	24374
elk	99.93	99.09	19.37	98.49	99.66	99.07	25292
horse	99.69	99.01	23.16	97.48	99.34	98.41	4759
moose	99.73	97.47	37.34	97.85	97.18	97.52	24943
average	**99.79**	**98.55**	**32.48**	**98.33**	**98.32**	**98.33**	**162570**
YOLO							
class	**Class Specific**			**mRP threshold 6.80 %**			**GT**
	AP(%)	**RP(%)**	**Score(%)**	**Recall(%)**	**Precision(%)**	**Average(%)**	
bear	98.92	96.74	11.85	97.33	94.12	95.72	25268
bison	98.18	96.17	5.81	95.98	96.56	96.27	31294
cow	98.56	95.18	7.31	95.35	94.84	95.09	4340
coyote	96.97	94.57	5.5	94.24	95.37	94.8	22300
deer	99.04	96.04	14.66	96.94	94.09	95.52	24374
elk	97.97	97.75	0.38	97.22	99.97	98.59	25292
horse	97.5	95.15	5.73	94.94	96.39	95.67	4759
moose	99.85	99.22	3.63	98.96	99.58	99.27	24943
average	**98.37**	**96.35**	**6.86**	**96.37**	**96.37**	**96.37**	**162570**

TABLE VI: Training configuration #2 class-specific results for RETINA, SSDIN, SSDMO and YOLO