ISCAS 2020
Virtual, October 10-21
IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS
DEPLOYING CAS KNOWLEDGE TOWARDS SOCIETY GRAND CHALLENGES
CAS IEEE CIRCUITS AND SYSTEMS SOCIETY

NRS: Non-Reliability Suppression for False Positives Elimination in YOLOv3

Ren Wang[1], Jin-Sung Kim[2]*, Hyuk-Jae Lee[1]
[1]Seoul National University, Korea [2]Sun Moon University, Korea

2020 IEEE International Symposium on Circuits and Systems
Virtual, October 10-21, 2020

Ren Wang

서울대학교 컴퓨터구조 및 병렬처리 연구실
Computer Architecture and Parallel Processing Lab.

Motivation

- Modern object detectors generally give redundant detections for a single object.
- Non-maximum suppression (NMS) is widely used as a post-processing step to select an accurate bounding box for an object among cluttered candidates.
- However, NMS cannot handle the redundant detections with different classification results.

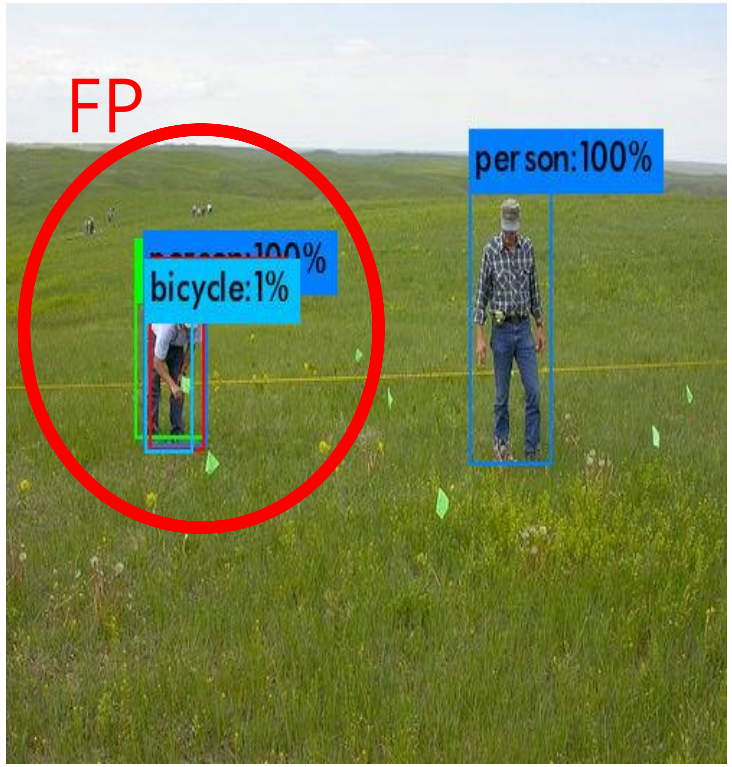
Redundant ‘bicycle’ and ‘person’ detections for the object ‘person’

NMS

Remaining redundant ‘bicycle’ detection for the object ‘person’

Problem Analysis

- Redundant detections are considered as false positives (FP). -> ***Accuracy degradation.***
- Two basic methods for eliminating potential FPs.
  - **Confidence Threshold**: Detections with low confidence scores.
  - **NMS**: Detections with identical classification results and high overlap ratios.
- Limitation of Confidence Threshold: loss of true positives (TP).



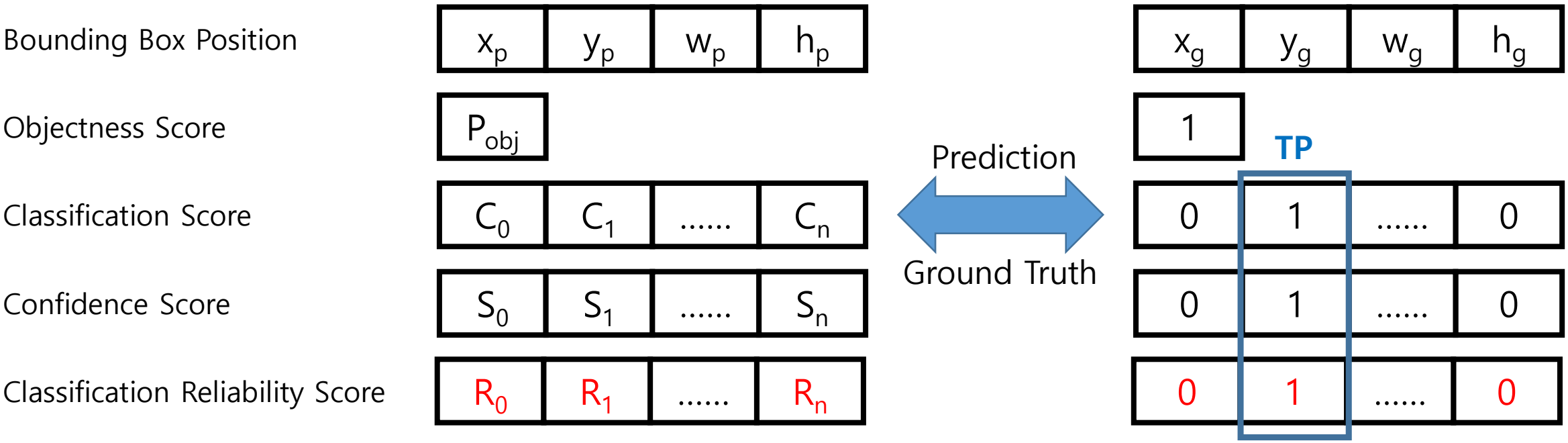
Redundant 'bicycle' detection for the object 'person'.

Same Confidence Score

Correct 'bicycle' detection for the object 'bicycle'.

Proposed Metric: Classification Reliability Score (R)

- What is a perfect detection?
  - Condition 1: Confidence score of true positive is **1**. ($S_n = C_n * P_{obj}$)
  - Condition 2: IoU between the predicted bounding box and ground truth box is **1**.
  - **Condition 3**: The sum of false positives' classification scores is **0**.
- A new metric: Classification Reliability Score. -> $R_n = C_n / sum(C_0, C_1, \ldots C_n)$

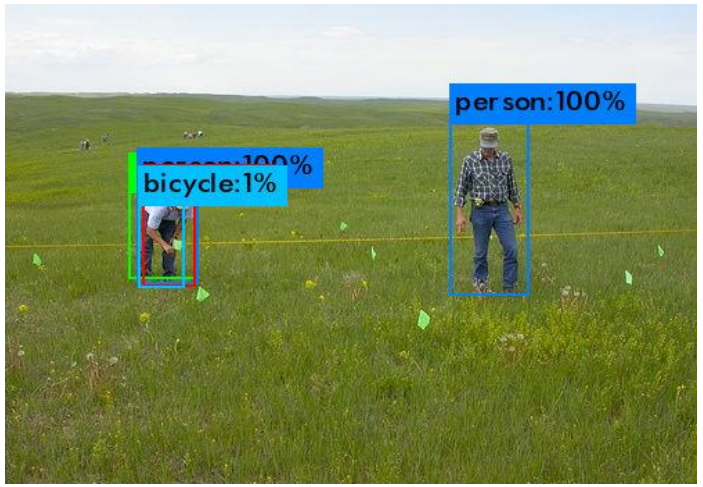


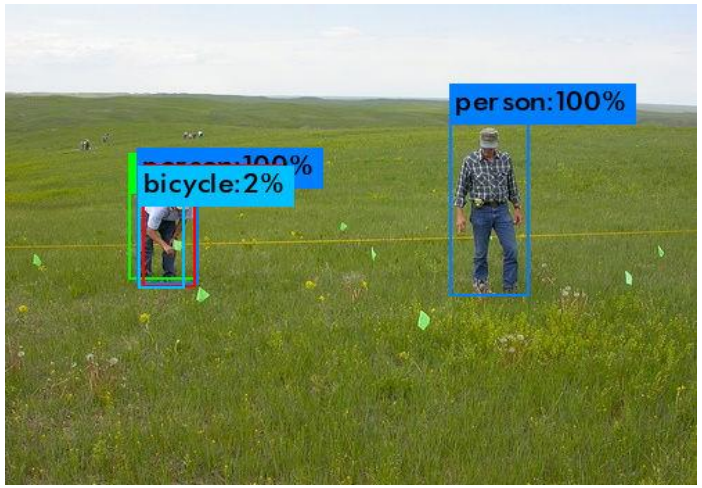
Components in the predicted bounding box of YOLOv3

Properties of Classification Reliability Score

- Range: [0, 1]
- High R score means high reliability of classification result.
- Example: Two ‘bicycle’ detections have the same confidence scores but totally different R scores.

Confidence Score



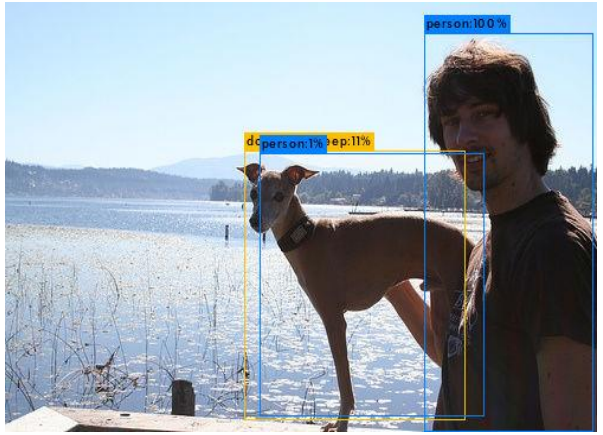


Classification Reliability Score

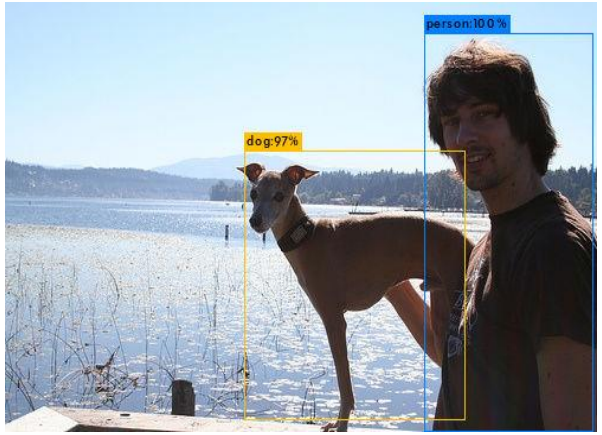
Non-Reliability Suppression (NRS)

- Detections with low R scores are likely to be FPs.
- NRS: Eliminate detections with R scores lower than predefined thresholds.
- R threshold: category specific & confidence score specific
  - Confidence regions: [1, 0.5], [0.5, 0.1], [0.1, 0.05], [0.05, 0.01], [0.01, 0.005]
- Applied after NMS.

dog: 97%
sheep: 11%
person: 10.0%
person: 1%

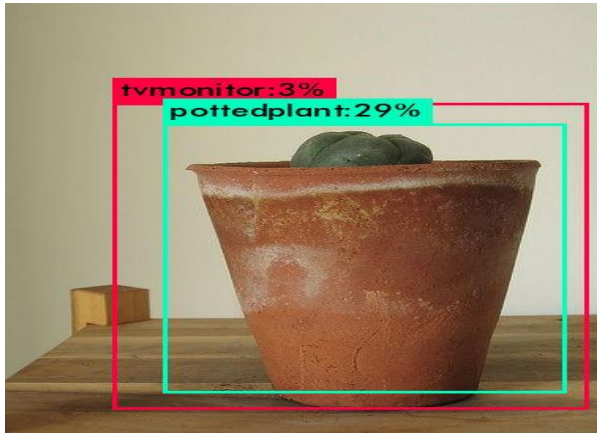


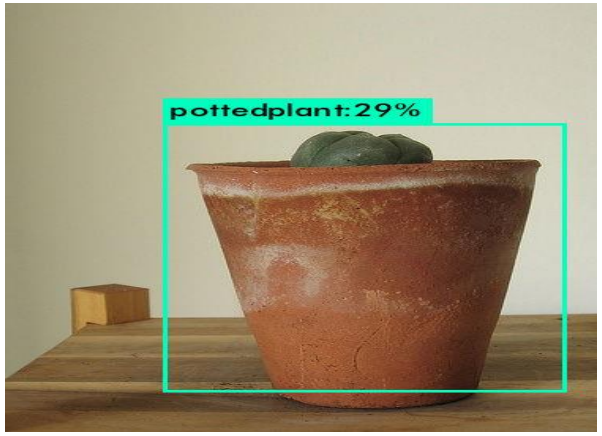
NRS



dog: 97%
person: 10.0%

tvmonitor: 3%
pottedplant: 29%





pottedplant: 29%

Experiments & Results

- Dataset: PASCAL VOC detection benchmark.
- Performance on different categories.

Class	NMS			NMS + NRS			# of GT
	# of TP	*# of FP*	AP_{50}	*# of TP*	*# of FP*	AP_{50}	
aeroplane	256	196	87.2	254	95	87.1	285
bicycle	292	313	84.0	289	109	83.9	337
bird	361	288	75.4	360	259	75.4	459
boat	206	421	70.4	205	230	**71.0**	263
bottle	323	449	61.8	322	316	**62.1**	469
bus	187	237	84.8	187	149	84.8	213
car	1081	655	87.9	1079	434	**88.0**	1201
cat	324	287	88.3	323	179	88.3	358
chair	543	1300	60.9	539	1022	**61.0**	756
cow	222	343	81.3	221	330	81.2	244
diningtable	170	373	74.0	168	152	**74.4**	206
dog	440	595	85.4	439	465	85.4	489
horse	311	387	86.9	311	371	**87.1**	348
motorbike	289	342	86.1	287	129	**86.2**	325
person	3823	2227	81.4	3819	1958	81.4	4528
pottedplant	282	534	48.1	282	415	**48.8**	480
sheep	203	396	76.1	202	246	**76.2**	242
sofa	211	580	78.4	211	435	**78.5**	239
train	249	287	84.7	248	96	**84.8**	282
tv/monitor	247	280	75.7	246	122	**76.3**	308

3.8% FP ↓

0.7 AP ↑

66.6% FP ↓

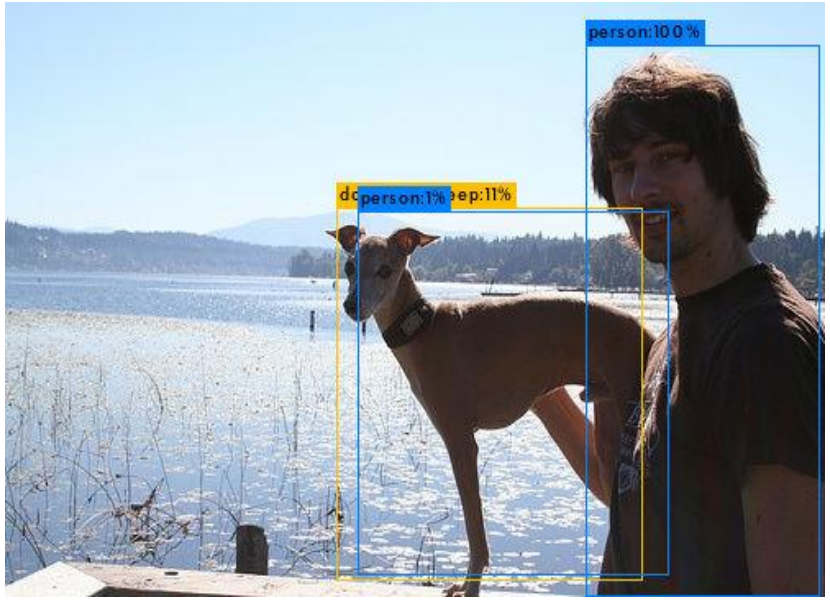
Experiments & Results

- Performance on different confidence thresholds.
  - 0.005: 28.4% FP ↓ & 0.3% TP ↓

Confidence Threshold	NMS			NMS + NRS		
	# of TP	*# of FP*	*mAP*	*# of TP*	*# of FP*	*mAP*
0.5	8849	1356	70.8	8849	1305	70.8
0.1	9396	2894	74.8	9390	2593	74.8
0.05	9574	3869	75.8	9561	3334	75.8
0.01	9908	7652	77.5	9886	6114	**77.6**
0.005	10020	10490	77.9	9992	7512	**78.1**

- Elimination of unreliable detections.
  - Classification: 39.0% ↓ ; Non-classification: 13.2% ↓

Method	# of FP (classification error)	# of FP (non-classification error)
NMS	6164	4326
NMS + NRS	3757	3755



sheep: 11%; -> classification error
person: 1%; -> non-classification error

- Inference latency.
  - Image resolution: 416 x 416
  - NRS: 0.01 ~ 0.02ms per image on Intel(R) Xeon(R) Gold 5118 CPU @2.30GHz.
  - Total inference latency: 29ms per image on one NVIDIA Titan Xp GPU.

Conclusion

- NRS efficiently eliminates the false positives while the loss of true positives can be ignored.
- NRS is simple to be implemented and the computational complexity is minimal.
- Comparison with recent NMS algorithms.

NMS		Extra Training	Extra Neural Network
Greedy-based NMS	Traditional-NMS	No	No
	Soft-NMS	No	No
	Softer-NMS	Yes	No
	Fitness-NMS	Yes	No
	IoU-guided NMS	Yes	Yes
	Adaptive-NMS	Yes	Yes
Learning-based NMS	Gossip Net	Yes	Yes
	Object Relation Module	Yes	Yes
NRS		No	No

- Lack of NRS
  - Empirically predefined R thresholds. Need some strategies to find optimal parameters.

Thank You