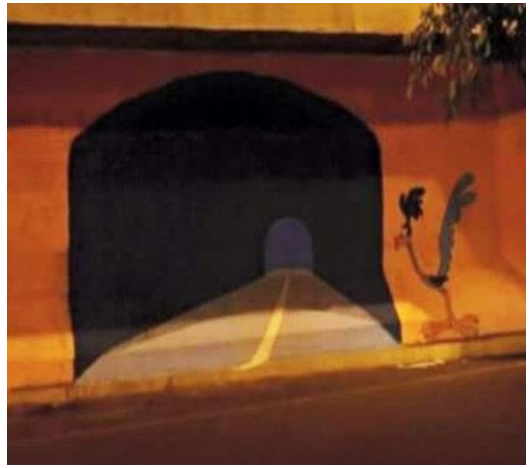
FROM BLACK-BOX TO WHITE-BOX: EXAMINING CONFIDENCE CALIBRATION UNDER DIFFERENT CONDITIONS

FRANZISKA SCHWAIGER, MAXIMILIAN HENNE, FABIAN KUEPPERS, FELIPPE SCHMOELLER ROZA ,KARSTEN ROSCHER , ANSELM HASELHOFF

THE PERCEPTION CHALLENGE

Source: Towards a Framework to Manage Perceptual Uncertainty for Safe Automated Driving, Krzysztof Czarnecki and Rick Salay



BACKGROUND: CONFIDENCE CALIBRATION FOR CLASSIFICATION

Modern neural networks are overconfident [1]

Measurement of Miscalibration: *Expected Calibration Error* (ECE):

$$ECE = \sum_{n=1}^{N} \frac{|I(n)|}{|D|} \cdot |acc(n) - conf(n)|$$

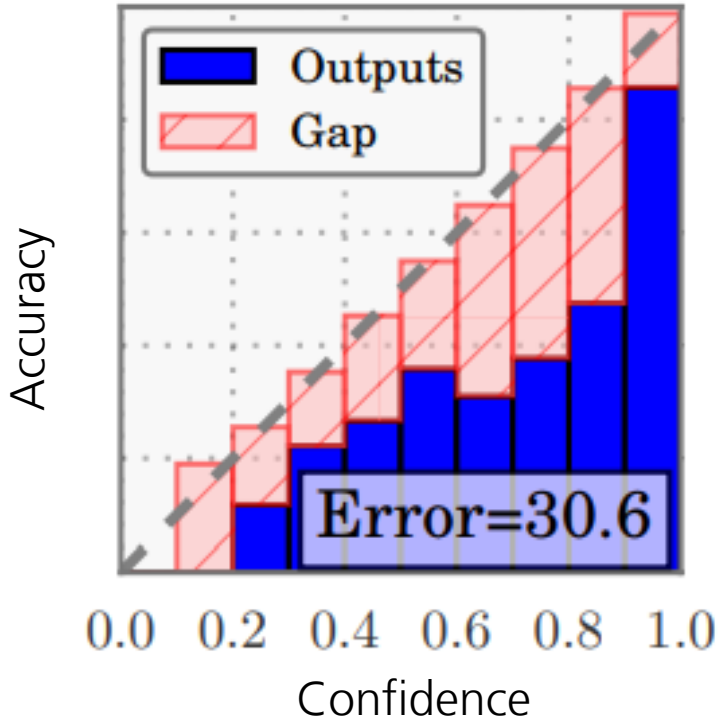
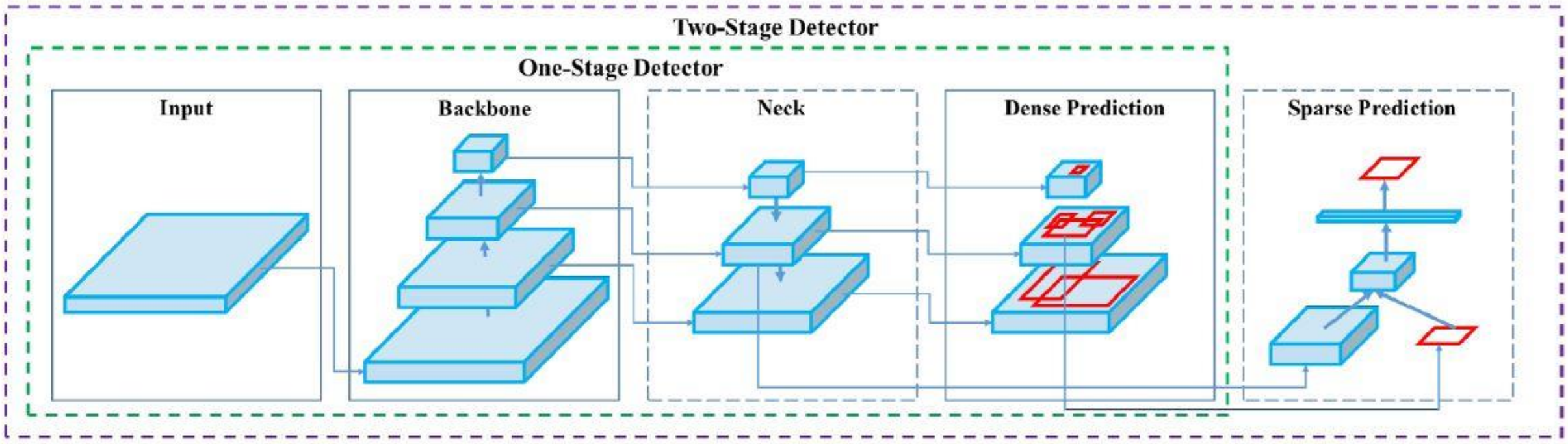


Fig. 1: Reliability Diagram of a 110-layer ResNet on CIFAR-100 [1]

OBJECT DETECTION PIPELINE



A Bochkovskiy et. al, Yolov4: Optimal speed and accuracy of object detection

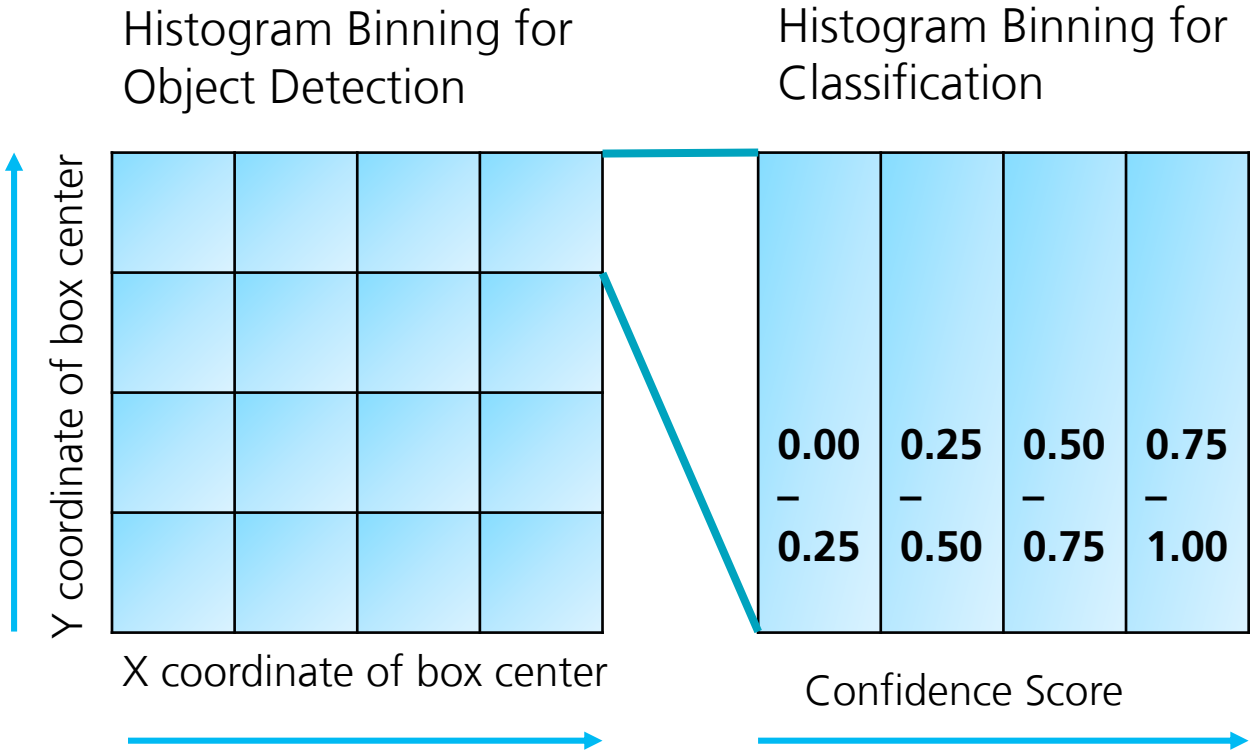
BACKGROUND: CONFIDENCE CALIBRATION FOR OBJECT DETECTION

Modern object detectors are also not well calibrated [2]

Measurement of Miscalibration:

Detection Expected Calibration Error (D-ECE):

$$D - ECE_K = \sum_{n=1}^{N_{total}} \frac{|I(n)|}{|D|} \cdot |prec(n) - conf(n)|$$



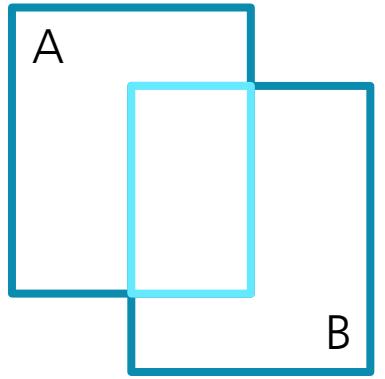
Going from 1D Histogram Binning to multi-dimensional Histogram Binning

BACKGROUND: NON-MAXIMUM-SUPPRESSION (NMS)

Method: selects a single bounding box with the highest confidence score out of many overlapping bounding boxes

Intersection over Union (IoU): controls how aggressively overlapping boxes are discarded

$$IoU(A,B) = \frac{A \cap B}{A \cup B}$$



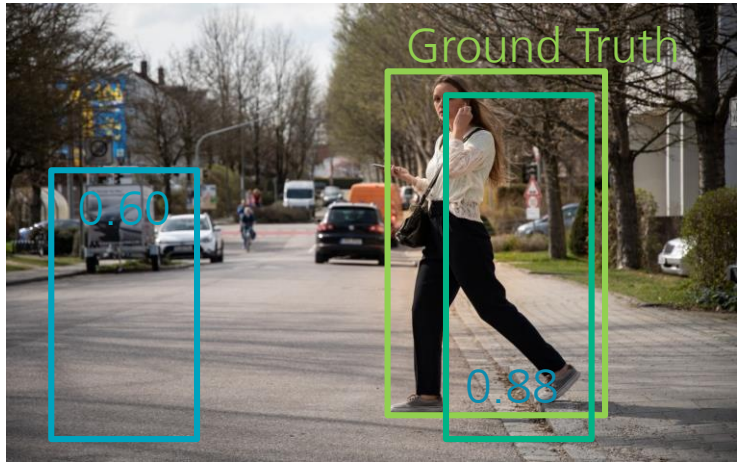
WHITE-BOX VS. BLACK-BOX

$$\text{TP: } IoU(\square, \square) \geq threshold$$

$$\text{FP: } IoU(\square, \square) < threshold$$

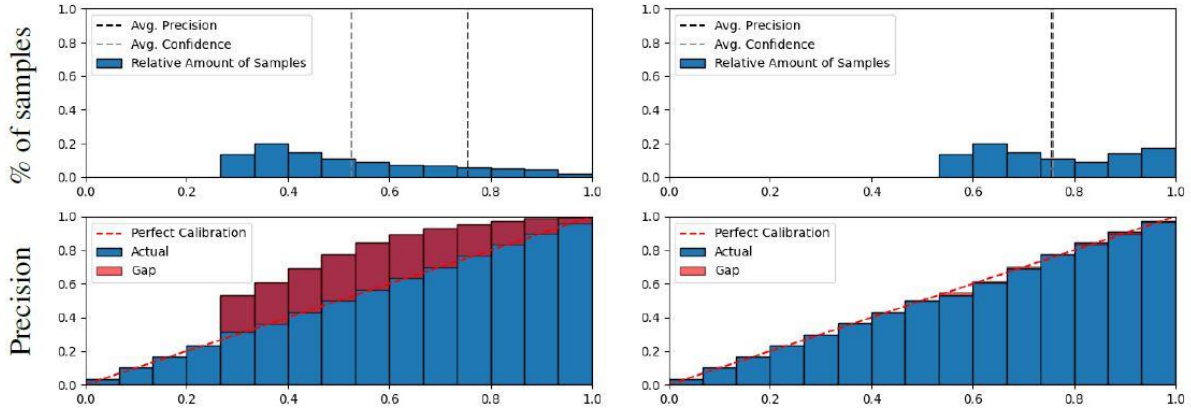
Miscalibration of object detectors is measured before (white-box) and after (black-box) NMS

Training and evaluation is done on the COCO2017 validation dataset



WHITE BOX CALIBRATION RESULTS

RetinaNet



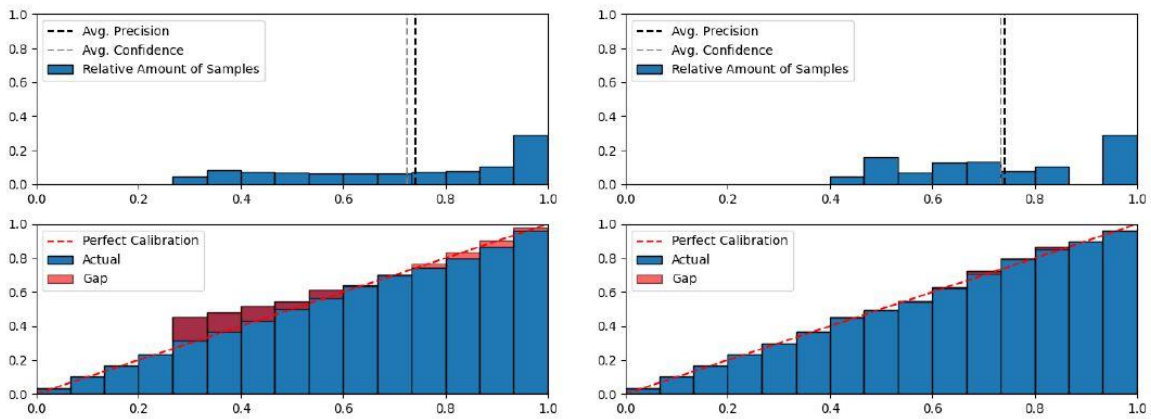
Before

Uncalibrated D-ECE:
22.9%

After

Calibrated D-ECE:
0.981%

Faster R-CNN



Before

Uncalibrated D-ECE:
4.20%

After

Calibrated D-ECE:
0.861%

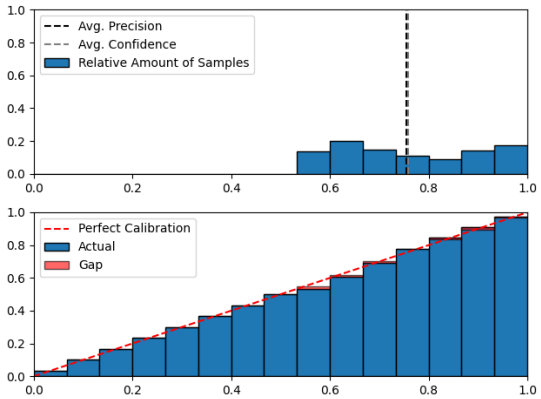
Fraunhofer IKS

QUALITATIVE RESULTS

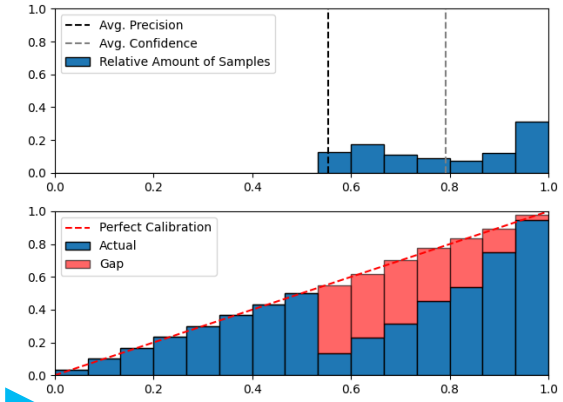
NMS potentially degrades initially well-calibrated predictors

After white-box calibration

Confidence Histogram (top) and Reliability Diagram (bottom) of RetinaNet

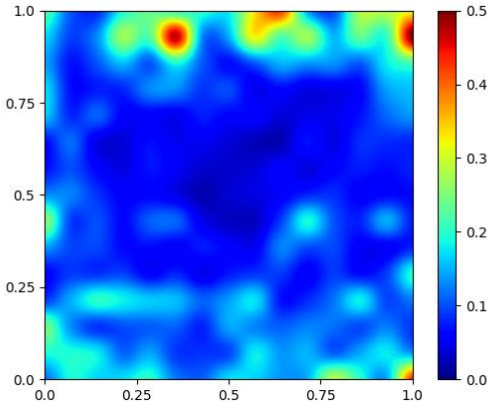


NMS

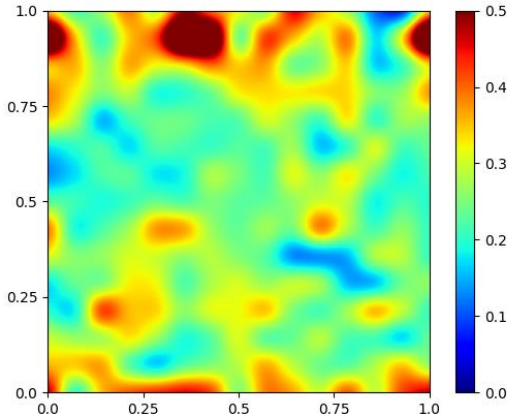


Position-dependent Heatmap of RetinaNet

D-ECE



D-ECE



Fraunhofer IKS

CONCLUSION

- Focal loss decreases average confidence
- White Box Calibration works
- NMS making predictions overconfident
- Detections on border of image tend to be worse calibrated than near the center
- Future work should focus on other box aggregation methods than NMS (e.g. average boxes)

REFERENCES

All quantitave calibration results are given in our paper in more detail.

References

[1] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger.

On Calibration of Modern Neural Networks.

In Proceedings of the 34th International Conference on Machine Learning, volume 70 of Proceedings of Machine Learning Research, pages 1321–1330. PMLR, August 2017.

[2] Küppers, F.; Kronenberger, J.; Shantia, A.; and Haselhoff, A. ¨ 2020.

Multivariate Confidence Calibration for Object Detection.

In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 326–327.