Do We Still Need Non-Maximum Suppression? Accurate Confidence Estimates and Implicit Duplication Modeling with IoU-Aware Calibration

Johannes Gilg   Torben Teepe   Fabian Herzog   Philipp Wolters   Gerhard Rigoll

Technical University Munich

Johannes.Gilg@TUM.de

Abstract

Object detectors are at the heart of many semi- and fully autonomous decision systems and are poised to become even more indispensable. They are, however, still lacking in accessibility and can sometimes produce unreliable predictions. Especially concerning in this regard are the—essentially hand-crafted—non-maximum suppression algorithms that lead to an obfuscated prediction process and biased confidence estimates. We show that we can eliminate classic NMS-style post-processing by using IoU-aware calibration. IoU-aware calibration is a conditional Beta calibration; this makes it parallelizable with no hyperparameters. Instead of arbitrary cutoffs or discounts, it implicitly accounts for the likelihood of each detection being a duplicate and adjusts the confidence score accordingly, resulting in empirically based precision estimates for each detection. Our extensive experiments on diverse detection architectures show that the proposed IoU-aware calibration can successfully model duplicate detections and improve calibration. Compared to the standard sequential NMS and calibration approach, our joint modeling can deliver performance gains over the best NMS-based alternative while producing consistently better-calibrated confidence predictions with less complexity. The code for all our experiments is publicly available [1].

1. Introduction

Object detectors are indispensable in many fields today, including computer vision, robotics, and autonomous systems. They locate and identify objects in images or videos, allowing for object recognition, tracking, and scene understanding. Object detectors are also at the heart of recent advances in driver-assistance systems, surveillance systems, and augmented reality. With the increasing demand for advanced technologies in areas such as security, industrial au-
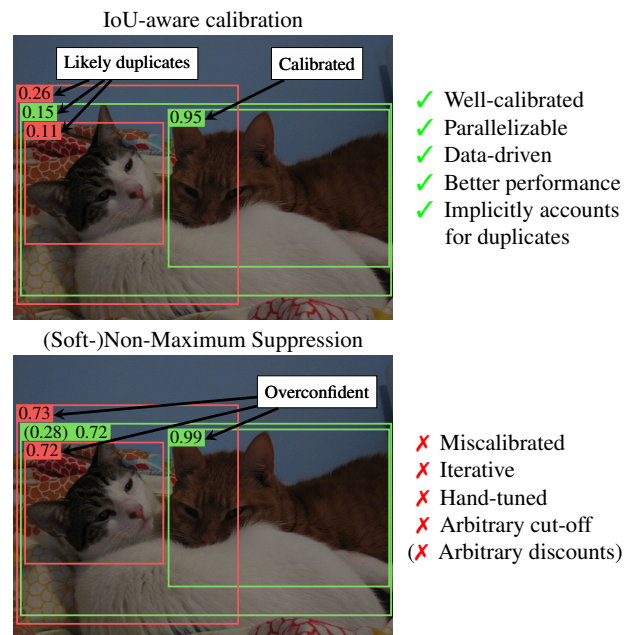

Figure 1. **Visualization of detections after IoU-aware calibration and (Soft-)NMS:** with comparison of their respective strengths (✓) and weaknesses (✗).

tomation, and even retail, object detectors are poised to play an even more important role in the future.

The proliferation of object detectors appears inevitable, but some hurdles remain. Users will likely include more non-experts, so transparency and accessibility—where today's detectors are still lacking—will be essential. Detectors will also need to be reliable and explainable, as deployments at-scale can amplify existing issues and biases. Researchers and practitioners need to examine current object detection pipelines closely and address their shortcomings in these critical areas of reliability, explainability, and biases.

A particularly troublesome part of almost all object detector's post-processing pipelines is Non-Maximum Suppression (NMS). It is a hand-crafted algorithm that is intended

---

[1]https://github.com/Blueblue4/IoU-AwareCalibration

to reduce duplicate detections of a single object. For NMS to work correctly, designers need to define a specific threshold that controls which detections are considered redundant when they overlap. Soft-NMS—the most popular improvement on the traditional NMS algorithm—introduces a different hyper-parameter that determines the degree to which overlapping detections' confidence scores are discounted via a Gaussian [1]. This rather unintuitive hyper-parameter can have a high performance and reliability impact and needs to be chosen carefully.

Using an algorithm designed for one use case in another domain or even just adapting it to frequent domain shifts can be cumbersome. It requires knowledge about the target domain, an understanding of the workings of the used NMS algorithms parameters, and how they influence the detectors' predictions. Alternatively, the parameter space can be searched for performance-maximizing settings, but this still leaves the user lacking an even superficial understanding of their implications.

Precise probability assessments are vital for automated decision-making, ensuring accurate and dependable outcomes and proper risk evaluation. The discontinuities introduced by NMS and the confidence discounting by the soft-NMS distort the distribution of detection confidences that is supposed to reflect accurate probabilities. NMS has been shown to impact the calibration of confidence predictions [36], that are often already miscalibrated to begin with [13]. When the confidence values do not reflect the empiric probability of object presence, they devolve into an inscrutable relative-likelihood ranking between different detections that are largely useless for automated systems.

We propose to map the issue of removing duplicate detections into the confidence score as one unified decision criteria with our data-driven IoU-aware calibration. We show that our proposed IoU-aware calibration can implicitly account for the likelihood of each detection being a duplicate of another detection and produce well-calibrated confidence estimates, eliminating the need for other post-processing steps like NMS and "normal" confidence calibration.

IoU-aware calibration offers several advantages:

1. it is a black-box calibration method, *i.e.* can be applied to any object detector as a post-processing,
2. it jointly tackles the problem of NMS and confidence calibration by implicitly modeling the probability of duplicate detections, making NMS obsolete and producing more reliable confidence predictions,
3. it can be vectorized as it does not require the iterative $\mathcal{O}(N^2)$ computations of soft- and greedy-NMS,
4. it has no hyper-parameters that need to be manually tuned,
5. it produces detectors with similar or better performance compared to the best NMS with better calibration on a wide range of tested detectors.

2. Related Work

Non-maximum suppression. The numerous variations of the standard NMS algorithm can be broadly categorized into NMS-improvements, task-specific adaptations, and proposed modifications to detection architectures. Improvements are *e.g.*, the soft-NMS that utilizes discounts instead of hard cut-offs [1] and Weighted Box Fusion [37] (wbf) that aims to fuse overlapping detections but also algorithmic improvements in processing speed via matrix-operations [2, 41] or spacial-priors to avoid unnecessary computations [39]. NMS-adaptations to specific domains include modifications specifically for vehicle-detections [7] and a differentiable NMS for 3D-object detection [20]. Numerous approaches require specific object detection architectures and modifications thereof [16,17,27,40,44,45]. We do not detail these approaches as we aim for a post-processing replacement for NMS that can be applied to object detectors regardless of the architecture.

A different line of research focuses on removing the need for NMS altogether by utilizing one-to-one assignment of detection proposals to ground truth objects [4, 38]. This approach can have some drawbacks, as it requires many times more training iterations and has lower performance than comparable architectures that rely on NMS [32]. This line of research is orthogonal to ours, as we show with our experiments on NMS-less architectures (see Sec. 5).

Confidence calibration. Since the findings of Guo *et al.*— deep-learning models are often highly miscalibrated—there has been a renewed interest in improving on the classic confidence calibration methods like Histogram binning [42] and Platt Scaling [13, 33]. Most closely related to our work are the improvements on Platt Scaling, like the more expressive Beta [18], the respective multi-class adaptations temperature scaling [13] and Dirichlet [19]. We claim no originality of the calibration function and rely on the previously introduced multivariate calibration methods for object detectors [21]. Our approach can be improved with further advances in multivariate calibration methods.

3. Background

Object detectors are almost always trained with a many-to-one assignment of detection proposals to ground truth objects. This approach increases the detection performance [6, 10, 32], as it can alleviate the foreground-background class imbalance problem of detectors—there are countless non-objects in every image and only a limited number of detectable objects [25]. This training strategy, by definition, incentivizes multiple detections per object. These excess detections then have to be filtered out during inference.

3.1. Non-Maximum Suppression

The combinatorial optimization problem of deciding which detections to keep and which to drop is usually still solved by using either NMS or soft-NMS [1] (see Algorithm 1). Soft-NMS and the standard NMS algorithm rely on the premise that duplicate detections are assumed to have similar scales and are highly overlapping. This "closeness" of two detections d_a and d_b is quantified via the Jaccard index, also termed Intersection over Union (IoU), of their respective bounding-boxes b_a and b_b:

$$\text{IoU}(b_a, b_b) = \frac{b_a \cap b_b}{b_a \cup b_b}. \tag{1}$$

It should not be confused with the Jaccard-distance which is $1 - \text{IoU}(b_a, b_b)$. The second shared assumption is that a higher predicted confidence $s_i > s_j$ indicates a higher likelihood for the detection d_i to be a True Positive (TP) detection than detection d_j. Thus, both algorithms greedily evaluate all N detections $\mathcal{D}$ from highest to lowest confidence s. At detection i the confidence s_i is updated to a new $\hat{s}_i$ according to:

$$\hat{s}_i = s_i \cdot f_{\text{nms}}(b_i, b_j), \qquad \forall j \mid s_j < s_i. \tag{2}$$

NMS uses an arbitrarily defined sharp cutoff threshold t_{nms} to classify detections as duplicates of a more confident detection and discounts them completely:

$$f_{\text{nms}}(b_a, b_b, t_{\text{nms}}) = \begin{cases} 1, & \text{IoU}(b_a, b_b) < t_{\text{nms}} \\ 0, & \text{IoU}(b_a, b_b) \geq t_{\text{nms}} \end{cases}, \tag{3}$$

while soft-NMS only discounts the confidence of overlapping detections according to their overlap, usually by a Gaussian:

$$f_{\text{nms}}(b_a, b_b, \sigma) = e^{-\frac{\text{IoU}(b_a, b_b)^2}{\sigma}}, \tag{4}$$

with a separate hyper-parameter σ. It was designed with the assumption that detections with an overlapping detection that is more confident can still be TP detections; they are just less likely to be so.

3.2. Object Detector Calibration

Suppose a model predicts the presence of an object with its detection d with confidence s, bounding box b, and category k. The model is *calibrated* if s_i is always equal to the probability of it being a TP detection (τ=1), *i.e.*, its precision at s_i:

$$\mathbb{P}(\tau = 1 | s = s_i) = s_i, \tag{5}$$

for any $s_i \in [0, 1]$. The concept of a TP detection in the object detection setting is more involved than in the classification setting, where the only requirement is that the predicted category k matches the ground truth label. For a detection

---

Algorithm 1 Greedy NMS pseudo code

Require:
 $\mathcal{B} = \{b_1, \dots, b_N\}$ # Bounding-boxes
 $\mathcal{S} = \{s_1, \dots, s_N\}$ # Confidences
 h # Hyperparmeter: σ or t_{nms}

1: $\mathcal{D} = \{\}; \mathcal{C} = \{\}$ # output Boxes and Confidences
2: **while** $\mathcal{S} \neq \varnothing$ **do**
3:   $i = \text{argmax}(\mathcal{S})$
4:   $\mathcal{S} = \mathcal{S} \setminus s_i; \mathcal{B} = \mathcal{B} \setminus b_i$
5:   $\mathcal{C} = \mathcal{C} \cup s_i; \mathcal{D} = \mathcal{D} \cup b_i$
6:   **for** $b_j \in \mathcal{B}$ **do**
7:    $s_j = s_j \cdot f_{\text{nms}}(s_j, s_i, h)$ # Eq. (3) or Eq. (4)
8:    **if** $s_j == 0$ **then**
9:     $\mathcal{S} = \mathcal{S} \setminus s_j; \mathcal{B} = \mathcal{B} \setminus b_j$ # optional
10:    **end if**
11:   **end for**
12: **end while**
13: **return** $\mathcal{D}, \mathcal{C}$

---

to be evaluated as a TP ($\tau = 1$) it needs to have the highest confidence s among the set of all detections $\mathcal{D}$ that have the correct category k and a bounding-box b that sufficiently overlaps a ground truth object. The sufficient overlap is determined according to a threshold t_{IoU} on the IoU (Eq. (1)) of the ground truth and detected object. This introduces some ambiguity into the concept of a TP detection as to what a "good" value for t_{IoU} is. In the performance evaluation metric mAP this is resolved by evaluating over a range of thresholds $t_{\text{IoU}} \in [0.5, 0.55, ..., 0.95]$ [26]. A binary label is usually required for the confidence calibration, so we use $t_{\text{IoU}} = 0.5$, same as mAP$_{50}$, unless specified otherwise [21].

3.2.1 Measuring Calibration Error

The Expected Calibration Error (ECE) tries to capture the expected difference of the left- and right-hand sides of Eq. (5) over some data distribution and the whole confidence interval. To make this evaluation practical, certain simplifications are used. The confidences s can take any value on the interval [0,1] and have to be discretized by splitting the interval into a fixed number (B), usually B=10, of bins b, each containing n_b detections. The precision for each bin $prec(b) = \mathbb{P}(\tau = 1 | s \in b)$ is evaluated over N detections on a labeled hold-out set. The result can be visualized in a reliability diagram [8] or averaged to produce the detection ECE-metric [13, 21]:

$$\text{ECE} = \sum_{b=1}^{B} \frac{n_b}{N} |prec(b) - conf(b)|. \tag{6}$$

The standard ECE was shown to have numerous pathologies. As a remedy, Nixon *et al.* introduced two variations on the ECE. The Static Calibration Error (SCE) aims to resolve

the problem of class dependency by calculating the ECE for each class separately and then averaging the results. The second metric, Adaptive Calibration Error (ACE), uses an adaptive binning scheme. The confidences of object detectors are usually not uniformly distributed but are clustered more densely at the extreme values close to 0 and 1. To account for this imbalance, ACE uses spaced bins, so each bin contains approximately the same number of detections [30]. In this work, all of the calibration-metrics, *i.e.* ECE, ACE, and SCE, are stated in percent unless otherwise indicated.

Proper scoring rules like negative log likelihood (NLL) are often used to fit and validate calibration methods [18]. NLL is minimized only if there is no miscalibration but does not directly capture the calibration error [12, 14].

3.2.2 Confidence Calibration

Modern neural networks are usually not well calibrated [13], and neither are modern deep-learning based object detectors [29]. The goal of model-agnostic (black-box) confidence calibration is to adjust the confidence estimates produced by a statistical model to better reflect the actual probability of the predicted outcomes using some mapping $s_i \mapsto f_{cal}(s_i)$, that is inspired by probability distributions. In the context of object detection:

$$\mathbb{P}(\tau = 1 | s = s_i) \stackrel{!}{=} f_{cal}(s_i, \boldsymbol{\theta}). \qquad (7)$$

According to this notation, the calibration functions are parametrized with $\boldsymbol{\theta}$. There are also non-parametric calibration functions like Histogram Binning [42], Isotonic Regression [5], and Bayesian Binning [28], which we omit here because they pose some unique problems in the multivariate calibration setting [11]. Common parametrized calibration methods include Platt scaling [33], Beta calibration [18], and Temperature scaling [13]. Note that most parametrized calibration functions are defined for logit outputs not the final confidence values s_i we use for simplicity, but a conversion is possible.

The parameters of the calibration function can not be directly optimized on their defined objective shown in Eq. (7). They are instead optimized according to some scoring rule, like the NLL-loss or the Brier Score [3] resulting in an optimization criteria like:

$$\arg\min_{\boldsymbol{\theta}} \sum_i (\tau_i - f_{cal}(s_i, \boldsymbol{\theta}))^2. \qquad (8)$$

In this form, the calibration can be posed as an optimization of a logistic regression problem and can be solved using gradient-based optimization procedures.

3.2.3 Multivariate Confidence Calibration

Confidence calibration can be further extended to model bi- or multivariate probability distributions to apply them to multi-output models. Object detectors produce additional regression outputs, *i.e.*, their bounding-box predictions, that can be used for multivariate calibration. The objective of a multivariate calibration of an object detector *e.g.*:

$$\mathbb{P}(\tau = 1 | s = s_i, b = b_i) \stackrel{!}{=} f_{cal}(s_i, b_i, \boldsymbol{\theta}), \qquad (9)$$

the motivation is to remove any conditional confidence biases with regard to object positions or size. Multivariate calibration functions can be derived from their univariate counterparts [23, 31]. If we model the conditioned-on variables—in this case s and b—as independent of each other, the modification of the calibration function is a relatively straightforward vectorized version of its univariate counterpart. When we reasonably assume that the variates are dependent on each other, we also need to model the covariance matrices. We use the implementation by Küppers *et al.*, which introduced conditionally dependent and independent multivariate adaptations for the logistic and Beta calibration in their multivariate calibration framework for object detectors [21].

Unlike the uni-variate calibration, the multivariate confidence calibration can impact a model's performance if there is any conditional bias regarding the additional model outputs [11].

4. IoU-aware calibration

Object detectors are an integral component of the vision stack in automated decision-making systems. To ensure trust in these systems, it is crucial that object detectors reliably produce predictions with precise probability estimates. The reliability of these predictions is essential in making informed decisions and achieving dependable outcomes in automated systems. In stark contrast to this goal, the hand-crafted discounting and suppression of the ubiquitous NMS-methods obscure the meaning of detector outputs.

The prevalent training paradigm of many-to-one detection-to-object assignment—and, to a lesser extent, the one-to-one assignment—makes the detectors produce overlapping duplicate detections for individual objects. These confident duplicate detections can severely impact the detector's performance and are, therefore, usually discarded or heavily discounted, depending on their overlap with other detections by heuristic-based algorithms.

We argue that if the confidence of predictions properly accounted for the likelihood of the detections being duplicates, we could eliminate the NMS post-processing altogether. As likely-duplicated detections would have very low

confidence—reflecting their low overall probability of being the true detection. Therefore, we aim to develop a data-driven approach that produces transparent decision criteria to improve the accuracy and reliability of object detection. This approach must ensure that the empirical probability of overlapping detections being duplicates is properly estimated, leading to well-calibrated reliable confidence predictions.

If we allow for the implicit assumption of most NMS approaches, then the likelihood of a detection being a duplicate δ_i is dependent on its confidence s_i, and the IoU with all other detections of the same class, which we represent via the Jaccard-distance vector $\mathbf{j}_i$ which is $[1 - IoU(b_i, b_1), 1 - IoU(b_i, b_2), ... 1 - IoU(b_i, b_N)]^T$. If we want to include the explicitly calculated probabilities of a detection being a duplicate in the calibration definition of Eq. (5), it results in:

$$\hat{s}_i = \underbrace{\mathbb{P}(\tau = 1|\delta_i)\mathbb{P}(\delta_i|s_i, \mathbf{j}_i)}_{=0} + \underbrace{\mathbb{P}(\tau = 1|\bar{\delta}_i)\mathbb{P}(\bar{\delta}_i|s_i, \mathbf{j}_i)}_{\mathbb{P}(\tau = 1|s_i, \mathbf{j}_i)}.$$
(10)

This requires us to explicitly determine the conditional probability of a detection not being a duplicate ($\mathbb{P}(\bar{\delta}_i|s_i, \mathbf{j}_i)$) and the conditional probability of a non-duplicate detection being a correct, i.e. TP, detection ($\mathbb{P}(\tau = 1|\bar{\delta}_i)$). We can simplify this by empirically determining $\mathbb{P}(\tau = 1|s_i, \mathbf{j}_i)$, thereby skipping the explicit calculation of the non-duplicate probability. This approach transforms the problem into a bi-variate confidence calibration (as defined in Eq. (9)).

The conditioning on $\mathbf{j}_i$ poses a practical problem as it is not a single parameter but a vector with a length that varies with the number of detections. The NMS methods solve this problem by iterating over all the box-parings of boxes with higher confidence (see Eq. (2)), but we can also use a basic summary statistic, like the minimum, of the Jaccard-distances similar to the approach proposed for the Matrix Non-Maximum Suppression [41]. These summary statistics can easily be calculated in parallel over the whole Jaccard-Distance-Matrix of (1 - IoU) values between all detections:

$$j_{i,min} = \min_{\forall s_k > s_i} \{1 - \text{IoU}(b_i, b_k)\},$$
(11)

removing the need for the serial computation with $\mathcal{O}(N^2)$ complexity of the NMS algorithm for N detections (compare Algorithm 1 to Algorithm 2). We will use $j_{i,min}$ as a proxy for $\mathbf{j}_i$ unless specified otherwise and ablate this choice later (see Sec. 5.1). From these design choices, our optimization objective becomes:

$$\arg\min_{\boldsymbol{\theta}} \sum_i (\tau_i - f_{cal}(s_i, j_{i,min}, \boldsymbol{\theta}))^2.$$
(12)

For the calibration function, we choose the Bi-variate conditional Beta calibration, as it is more expressive than the Lo-

---

Algorithm 2 Vectorized implementation of IoU-aware calibration in pseudo code

Require:
    $\boldsymbol{b} = [b_1, \ldots, b_N]$ # Bounding-boxes, shape: [N, 4]
    $\boldsymbol{s} = [s_1, \ldots, s_N]$ # Confidences, shape: [N]

1: $\boldsymbol{b}, \boldsymbol{s} = \text{sort\_descending}((\boldsymbol{b}, \boldsymbol{s}), \text{by}=\boldsymbol{s})$ # [N, 4], [N]
2: $\mathbf{J} = \text{ones}(N, N) - \text{IoU}(\boldsymbol{b}, \boldsymbol{b}^T)$ # Jaccard-dists., [N, N]
3: $\mathbf{J} = \mathbf{J} \cdot \text{lower\_triangular}(N, -1)$ # Mask sup., [N, N]
4: $\mathbf{j}_{min} = \min(\mathbf{J}, \text{axis}=1)$ # *cf*. Eq. (11), [N]
5: $\boldsymbol{s} = \text{Conditional\_Beta\_Calibration}(\boldsymbol{s}, \mathbf{j}_{min})$ # [N]
6: **return** $\boldsymbol{b}, \boldsymbol{s}$ # [N, 4], [N]

---

gistic calibration; it produces the identity function if the predictions of a model are already well calibrated, at the cost of a few more parameters [18]. We assume conditional dependence between our overlap-proxy (j_i) and the confidence s_i because we calculate the minimum Jaccard-Distance only from the more confident detections, *i.e.*, the detections that would "suppress" detection d_i in the classical NMS setting. This already introduces an indirect dependence between the two variables, which we want to account for in our modeling, but we also ablate this choice.

With this approach, we can now calibrate detections conditioned on a summary statistic of the Jaccard-Distances to the other detections, thereby implicitly accounting for the likelihood of each detection being a duplicate. This should help object detectors reliably produce predictions with precise probability estimates and remove the need for NMS post-processing.

5. Experiments

We now verify the theoretical justifications for our IoU-aware calibration by conducting detailed experiments and analyze individual design choices through ablations.

The setup. For our initial experiments we use a two-stage architecture, as the region-proposal mechanism makes them produce more duplicate detections, making the effects more measurable. We use a modern pre-trained version of the highly-popular Faster-RCNN [35] with a Resnet50 [15] backbone and a FPN-Neck [24], trained with multi-scale data augmentation for 36 epochs on Common Objects in Context dataset [26] (COCO). We remove the NMS post-processing and limit the maximum number of detections per image to 400 to allow for the additional duplicate detections. The number of detections is later limited to the standard 100 per image for evaluation. We use the multivariate calibration framework for object detectors [21] for the implementation of the Logistic and Beta calibration and their optimization. As usual for calibration methods, we split the COCO validation set, val2017, into a fitting and an evaluation sub-set. Specifically, we evaluate across 10 random

60:40 train:test image-wise splits and report the mean performance with the maximum difference from it. To evaluate if a detection is a TP ($\tau = 1$) we use the official COCO-evaluation script and a IoU-threshold of t_{IoU}=0.5 [26].

For NMS, the reported metrics are evaluated on the *same splits*, and the hyper-parameters for the different methods are determined by a parameter grid-search on the whole `val2017`, reflecting the *best possible performance* achievable by NMS. When only one NMS method is shown, we report the best performing out of Gaussian soft-NMS, standard NMS, and wbf.

We apply our proposed IoU-aware calibration in the described cross-validation scheme. We want to test if it can implicitly account for the likelihood of each detection being a duplicate. If our method can capture it, we should see a similar performance increase to a NMS-based post-processing. We verify this via the performance-metrics mAP and mAP$_{50}$. Additionally, we aim for the detector to reliably produce predictions with precise probability estimates, which we capture with the calibration-metrics: ECE, ACE, SCE and to a lesser extent with the NLL. Our method tightly couples the two issues of performance and calibration. Since a performance increase means our method is able to implicitly capture at least the relative likelihood of duplication, the explicitly optimized-for empiric probability assessment ($\mathbb{P}(\tau = 1|s_i, \mathbf{j}_i)$) is likely also a well-calibrated prediction (see Eq. (10)).

Since NMS has been shown to cause severe miscalibration [36], we add a uni-variate confidence calibration with the Beta calibration for a better calibration-metrics baseline.

Initial results. The results are shown in Tab. 1. As expected, our IoU-aware calibration produces well-calibrated detections. The significant improvement over the best calibrated NMS results provides evidence that the implicit modeling of the probability of duplicate detections is indeed a critical necessity to produce precise probability estimates. The IoU-aware calibration also performs very well on the performance-metrics. It even outperforms the best NMS, in this case soft-NMS by a very slight margin and NMS by 0.6 mAP. This is remarkable since our method relies only on the Jaccard distance to the most overlapping other detection. In contrast, the iterative NMS-methods consider all overlapping detections with their iterative approach.

What does IoU-aware calibration do? We visualize the calibration function to examine what the learned f_{cal} does (see Fig. 2 left) and compare it to the Gaussian soft-NMS (see Fig. 2 right). Fig. 2 reveals that the IoU-aware calibration adjusts the shape of the confidence mapping not only by the amount of overlap (in IoU) but also by the initial confidence value. While the curve approximates a Gaussian for s=0.9, it resembles to a linear decay for s=0.3. The confidence discounting for high IoUs is also much reduced, especially for initially high confidence values, com-

Post-Processing		Performance-Metrics		Calibration-Metrics			
NMS	Calibration	mAP↑	mAP$_{50}$↑	ECE↓	ACE↓	SCE↓	NLL↓
none	Beta	13.55±0.58	17.68±0.81	0.11±0.03	0.07±0.03	0.59±0.04	0.07±0.00
NMS	none	40.76±0.99	61.10±1.32	7.82±0.11	7.82±0.11	7.69±0.11	0.21±0.01
NMS	Beta	40.76±0.99	61.10±1.32	0.22±0.14	0.21±0.08	1.97±0.16	0.17±0.01
soft	none	41.34±1.01	61.18±1.28	6.12±0.07	6.12±0.07	6.33±0.09	0.17±0.00
soft	Beta	41.34±1.01	61.18±1.28	0.23±0.07	0.22±0.06	1.61±0.11	0.13±0.00
wbf	none	37.50±0.76	58.00±0.88	4.47±0.12	4.41±0.12	4.64±0.15	0.19±0.01
wbf	Beta	37.50±0.76	58.00±0.88	0.47±0.09	0.43±0.15	2.31±0.20	0.17±0.00
none	Cond.-Beta	**41.36**±1.01	**61.28**±1.32	**0.03**±0.01	**0.04**±0.02	**0.31**±0.02	**0.03**±0.00

Table 1. **Comparison of IoU-aware calibration to traditional NMS methods with calibration.** Averaged results with maximum positive and negative difference shown in brackets, lifted and lowered respectively.
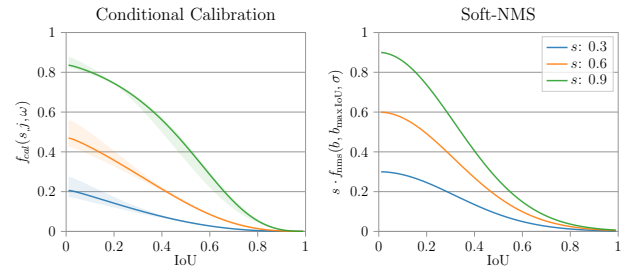

Figure 2. **Comparison of proposed IoU-aware calibration and soft-NMS.** Shows how confidence of detections is adjusted, depending on the IoU with a more confident detection with three different initial confidences s. Confidence intervals in lighter colours. Note how the conditional calibration can account for the different confidences with slightly varying curve shape and discounting factor. We show the best-performing soft-NMS with σ=0.2.

pared to the soft-NMS. The soft-NMS is applied iteratively (see Eq. (2)) for each more-confident detection, so detections with multiple overlapping higher-confident detections would have their confidence discounted even stronger.

5.1. Ablations

Our IoU-aware calibration delivers good results, which is a indication that our design- and modeling choices are well-founded and accurately reflect the underlying problem. Nonetheless, we evaluate the impact of each of the modeling choices and confirm their validity via ablations.

Choice of calibration function. We chose the Beta calibration function over the Logistic function because of its higher expressiveness. We assumed a conditional dependence between the confidence s_i of the detection and our proxy for the overlap with more confident detections, $j_{i,\min}$. Visual inspection of Fig. 2 confirms that the learned f_{cal} models some interaction between the two variables. We compare the conditional dependent bi-variate Beta calibration to the independent version as well as the dependent and independent bi-variate Logistic calibration function.

As shown in Tab. 2, there are no significant differences between the calibration functions, except for the independent Logistic calibration, which performs significantly worse for

Calibration-Method	Performance-Metrics		Calibration-Metrics			
	mAP↑	mAP$_{50}$↑	ECE↓	ACE↓	SCE↓	NLL↓
Ind.-Logistic	40.25±0.94	59.59±1.22	0.13±0.03	0.14±0.04	0.46±0.02	**0.03**±0.00
Cond.-Logistic	41.33±1.00	**61.42**±1.32	0.04±0.01	0.05±0.02	**0.31**±0.03	**0.03**±0.00
Ind.-Beta	41.32±0.98	61.20±1.31	0.08±0.02	**0.03**±0.02	0.33±0.02	**0.03**±0.00
Cond.-Beta	**41.36**±1.01	61.28±1.32	**0.03**±0.01	0.04±0.02	**0.31**±0.02	**0.03**±0.00

Table 2. **Ablation of different calibration methods.** Conditionally dependent and independent versions of the Bi-variate Logistic- and Beta calibration are evaluated.

both performance- and calibration-metrics. The independent Beta calibration has only a slightly higher ECE than its conditional counterpart. The performance differences between conditional Logistic and Beta calibrations are only minor. We use the Beta calibration for our experiments but note that a conditional Logistic calibration would likely produce similar results.

Conditioning on other summary-statistics. We chose our summary statistic of $\mathbf{j}_i$ by taking the minimum Jaccard-distance *i.e.*, the maximum IoU. The greedy NMS methods apply the confidence adjustments iteratively (see Eq. (2)), thereby accounting for individual overlapping detections, not only the extrema. This approach is not transferable to our calibration method, but we can look for a summary statistic that accounts for the additional overlapping detections. Ideally, the value should be bounded in [0, 1], making the product of the Jaccard-distances an ideal choice. We denote the resulting variable as $j_{i,\Pi}$. We compute the statistics $j_{i,\Pi}$ and $j_{i,min}$ from $\mathbf{j}_i$, over all the detections with *higher* confidences than the detection d_i that has its confidence s_i adjusted. In the context of NMS, these would be the overlaps with *suppressing* bounding boxes; this leaves the potential influence of boxes *suppressed* by detection d_i unaccounted for. To address this, we can also compute the proposed summary statistics for this, left out, suppressing part of the $\mathbf{j}_i$ and denote them as $j_{i,min}^{sup}$ and $j_{i,\Pi}^{sup}$ even though we have no prior reason to assume that these statistics are relevant for our calibration objective.

Comparing the performance changes by conditioning on the different variables (see Tab. 3), we find that the conditioning on the minimum Jaccard-distance of the suppressing detections j_{min} leads to significantly better performance and calibration scores than any of the alternatives. The introduced product j_Π is not as effective as j_{min} at capturing the likelihood of duplicate detections and the statistics over the suppressed detections, j_{min}^{sup} and j_Π^{sup}, are only slightly barely better at it than the baseline, which can be seen on the performance metrics.

Conditioning on additional variables. Up to this point, we only performed Bi-variate calibration using the confidence s and a single summary statistic over all Jaccard-distances to the other detections ($\mathbf{j}_i$). While the other proposed metrics did not perform well on their own, we could combine them with the best statistic, j_{min}, to a multivariate

Conditioned-Variable	Performance-Metrics		Calibration-Metrics			
	mAP↑	mAP$_{50}$↑	ECE↓	ACE↓	SCE↓	NLL↓
j_{min}	**41.34**±1.02	**61.26**±1.34	**0.04**±0.01	**0.04**±0.02	**0.31**±0.03	**0.03**±0.00
j_Π	40.50±0.89	59.35±1.17	0.06±0.03	0.05±0.02	0.37±0.04	**0.03**±0.00
j_{min}^{sup}	14.48±0.60	19.26±0.77	0.10±0.03	0.08±0.02	0.60±0.04	0.07±0.00
j_Π^{sup}	13.72±0.77	17.95±0.98	0.08±0.18	0.09±0.20	0.56±0.09	0.07±0.00
none	13.55±0.58	17.68±0.81	0.11±0.03	0.07±0.03	0.59±0.04	0.07±0.00

Table 3. **Ablation conditional calibration variables.** Comparing the performance by taking summary statistics over all the *suppressing* detections (j_Π and j_{min}) and over the detections that are being *suppressed* (j_{min}^{sup} and j_Π^{sup}).

Variables				Performance-Metrics		Calibration-Metrics		
j_{min}	j_Π	j_{min}^{sup}	j_Π^{sup}	mAP↑	mAP$_{50}$↑	ECE↓	ACE↓	SCE↓
✓				41.21±1.39	61.07±1.85	0.05±0.16	0.04±0.12	0.32±0.09
✓	✓			41.33±1.07	61.21±1.57	0.04±0.03	0.03±0.04	0.31±0.03
✓		✓		41.14±1.23	60.97±1.61	0.06±0.15	0.05±0.11	0.32±0.08
✓	✓	✓		41.22±1.15	61.06±1.52	0.05±0.12	0.04±0.05	0.32±0.05
✓			✓	41.14±1.31	60.97±1.75	0.07±0.10	0.05±0.12	0.32±0.04
✓		✓	✓	41.09±1.21	60.93±1.60	0.08±0.13	0.05±0.11	0.33±0.08
✓	✓		✓	41.28±1.08	61.13±1.49	0.04±0.03	0.03±0.04	0.31±0.02
✓	✓	✓	✓	41.27±1.07	61.11±1.45	0.04±0.06	0.03±0.01	0.31±0.05

Table 4. **Ablation of combination of additional calibration variables.** We apply conditional multivariate calibration with j_{min} and different combination of the additional summary statistics (j_Π, j_{min}^{sup} and j_Π^{sup}).

Object Detector	Post-Processing	mAP↑	mAP$_{50}$↑
Sparse-RCNN RN50 [38]	none (default)	45.49±1.33	64.38±1.51
	NMS	**45.67**±1.33	**64.77**±1.50
	cond. Beta	45.48±1.33	64.37±1.51
CenterNet HG [46]	none (default)	40.66±0.84	59.28±1.28
	NMS	40.73±0.83	59.45±1.28
	cond. Beta	**40.76**±0.82	**59.61**±1.29
Detr RN50 [4]	none (default)	40.49±1.10	60.89±1.68
	NMS	40.56±1.08	61.11±1.65
	cond. Beta	**40.64**±1.07	**61.62**±1.70

Table 5. **We compare post-processing methods on Detection architectures designed for No-NMS.** As a sanity check we evaluate the IoU-aware calibration on architectures that do not require NMS.

confidence calibration to see if jointly they are better able to model the likelihood of duplication.

We explore all possible combinations of the other available summary statistics, j_Π, j_{min}^{sup} and j_Π^{sup} with j_{min} and s (see Tab. 4). There is no significant improvement, as seen in the performance and calibration metrics. Combinations that include j_Π seem to have a slight edge over the ones that do not include it. However, unlike the contribution of conditioning on j_{min}, these gains are far from significant enough to justify modeling the conditioning on an additional variable.

5.2. Results

Finally, we verify the effectiveness of our IoU-aware calibration on a wide range of object detection architectures. We aim to get representatives of the different design

Object-	Post-Processing		`test-dev2017`		`val2017` (Cross-Val)					
Detector	NMS	Calibration	mAP↑	mAP$_{50}$↑	mAP↑	mAP$_{50}$↑	ECE↓	ACE↓	SCE↓	NLL↓
YoloV3-608 [34]	✓	none	34.2	56.4	34.78±0.87	57.16±1.17	1.55±0.12	1.31±0.13	2.95±0.15	0.19±0.00
	✓	Beta	34.2	56.4	34.78±0.87	57.16±1.17	0.28±0.07	0.25±0.10	2.71±0.18	0.19±0.00
	✗	cond. Beta	**34.6**	**56.7**	**35.28**±0.87	**57.62**±1.17	**0.22**±0.12	**0.13**±0.07	**1.22**±0.09	**0.09**±0.00
RetinaNet RN101 [25]	✓	none	**38.8**	**56.3**	**38.71**±1.18	**55.50**±1.28	6.85±0.06	6.86±0.09	7.49±0.10	0.17±0.00
	✓	Beta	**38.8**	**56.3**	**38.71**±1.18	**55.50**±1.28	0.19±0.12	0.20±0.12	2.00±0.14	0.12±0.00
	✗	cond. Beta	38.6	56.0	38.64±1.14	55.31±1.28	**0.06**±0.02	**0.04**±0.03	**0.28**±0.02	**0.02**±0.00
Faster-RCNN RN50 [35]	✓	none	**41.4**	61.6	**41.34**±1.01	61.18±1.28	6.12±0.07	6.12±0.07	6.33±0.09	0.17±0.00
	✓	Beta	**41.4**	61.6	**41.34**±1.01	61.18±1.28	0.23±0.07	0.22±0.06	1.61±0.11	0.13±0.00
	✗	cond. Beta	41.3	**61.8**	**41.35**±1.01	**61.28**±1.32	**0.04**±0.01	**0.04**±0.02	**0.31**±0.03	**0.03**±0.00
Varifocalnet RN50 [43]	✓	none	47.7	**65.8**	47.85±1.00	**65.65**±1.39	6.47±0.06	5.77±0.08	6.92±0.10	0.14±0.00
	✓	Beta	47.7	**65.8**	47.85±1.00	**65.65**±1.39	0.15±0.04	0.10±0.04	1.31±0.07	0.10±0.00
	✗	cond. Beta	**47.9**	65.7	**48.16**±1.02	65.59±1.39	**0.09**±0.03	**0.05**±0.01	**0.35**±0.03	**0.03**±0.00
YOLOX-L [10]	✓	none	49.3	67.5	49.44±1.72	67.22±1.95	1.83±0.14	1.80±0.10	4.55±0.26	0.23±0.00
	✓	Beta	49.3	67.5	49.44±1.72	67.22±1.95	0.26±0.12	0.27±0.13	3.99±0.23	0.22±0.00
	✗	cond. Beta	**49.5**	**67.8**	**49.69**±1.65	**67.50**±1.91	**0.19**±0.03	**0.08**±0.07	**0.94**±0.07	**0.06**±0.00
HTC CBNetv2 Swin-L † [22]	✓	none	58.4	76.0	58.56±1.55	75.75±1.70	8.63±0.33	8.63±0.33	9.13±0.51	0.26±0.01
	✓	Beta	58.4	76.0	58.56±1.55	75.75±1.70	0.38±0.12	0.35±0.19	4.10±0.34	0.21±0.01
	✗	cond. Beta	**58.8**	**76.8**	**58.99**±1.44	**76.45**±1.59	**0.06**±0.03	**0.07**±0.05	**0.62**±0.05	**0.03**±0.00
EVA Cascade Mask-RCNN † [9]	✓	none	**63.0**	81.6	**63.07**±1.11	**81.56**±1.09	1.21±0.07	1.21±0.07	1.47±0.08	0.08±0.00
	✓	Beta	**63.0**	81.6	**63.07**±1.11	**81.56**±1.09	**0.09**±0.03	**0.06**±0.05	**0.98**±0.05	**0.07**±0.00
	✗	cond. Beta	**63.0**	**81.9**	**63.07**±1.11	**81.56**±1.10	**0.09**±0.03	0.07±0.02	**0.98**±0.06	**0.07**±0.00

Table 6. **Proposed IoU-aware calibration is applied on a variety of Object detectors.** We apply our conditional calibration to a wide range of object detectors, including state-of-the-art detectors, and compare it to the best (oracle) NMS alternative. We evaluate the detections for performance- and calibration-metrics on 10 random splits on the validation set and report the mean and maximum deviation. On the test-dev set, we show only performance-metrics produced by the evaluation server, as ground truth labels are not publicly available. Detectors marked with † are also instance-segmentation models.

philosophies: from single-stage to two-stage, from small, real-time to huge billion-parameter models, anchor-based and anchor-free, designed for NMS or without, transformer- and CNN-based models, and different backbone-networks. The results on all object detectors are shown in Tab. 6. As a sanity check we also verify that there is little performance variation on architectures that are designed without NMS post-processing (see Tab. 5). The method consistently outperforms standard NMS—for some models by up to 0.7 mAP—and can often improve on the oracle-tuned soft-NMS by up to 0.4 mAP.

6. Discussion

The performance of IoU-aware calibration is on-par with with the best fine-tuned NMS-method across models which proves that it is able to model the underlying problem of duplicate detections regardless of architecture and better than any individual type of NMS. The better-calibrated confidence predictions further show that the calibrations is not only good at implicitly capturing the likelihood of duplication, but the duplication likelihood is also a crucial intermediary for confidence predictions that accurately reflect the empirical precision. This shows that we can fuse the de-duplication into the calibration into a single step, simplifying the post-processing and improving calibration and performance with reduced complexity.

Limitations. The IoU-aware calibration is a data-driven approach, so a distribution shift between the data used for calibration and data seen at deployment can lead to over- or under-confident predictions. It can also lead to performance degradation in significantly more crowded scenes than the calibration subset.

7. Conclusion

We proposed IoU-aware calibration as a data-driven alternative to the classic iterative NMS and calibration post-processing. We showed how our method implicitly accounts for the likelihood of each detection being a duplicate of another and adjusts the confidence score accordingly. This results in well-calibrated probability estimates for the precision of each detection. With comprehensive ablation studies, we demonstrated the validity of our design choices. Furthermore, we conducted extensive experiments across various detection architectures, we found that the proposed IoU-aware Calibration method results in performance gains comparable to those obtained by an oracle chosen fine-tuned non-maximum suppression algorithm. In some cases, it even achieves performance gains over the best NMS-based alternative while also producing consistently better calibrated confidence predictions than comparable calibrations. We see this as a significant step away from hand-crafted algorithms toward data-driven detectors with interpretable outputs.

References

[1] Navaneeth Bodla, Bharat Singh, Rama Chellappa, and Larry S Davis. Soft-nms–improving object detection with one line of code. In *ICCV*, pages 5561–5569, 2017. 2, 3

[2] Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. Yolact: Real-time instance segmentation. In *ICCV*, October 2019. 2

[3] Glenn W Brier et al. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950. 4

[4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *ECCV*, pages 213–229, 2020. 2, 7

[5] Nilotpal Chakravarti. Isotonic median regression: a linear programming approach. *Mathematics of operations research*, 14(2):303–308, 1989. 4

[6] Qiang Chen, Xiaokang Chen, Jian Wang, Shan Zhang, Kun Yao, Haocheng Feng, Junyu Han, Errui Ding, Gang Zeng, and Jingdong Wang. Group detr: Fast detr training with group-wise one-to-many assignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6633–6642, 2023. 2

[7] Guochen Cui, Shufeng Wang, Yongqing Wang, Zhe Liu, Yadong Yuan, and Qiaoqiao Wang. Preceding vehicle detection using faster r-cnn based on speed classificatiofrom black-box to white-box: examining confidence calibration under different conditionsn random anchor and q-square penalty coefficient. *Electronics*, 8(9):1024, 2019. 2

[8] Morris H DeGroot and Stephen E Fienberg. The comparison and evaluation of forecasters. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 32(1-2):12–22, 1983. 3

[9] Yuxin Fang, Wen Wang, Binhui Xie, Quan Sun, Ledell Wu, Xinggang Wang, Tiejun Huang, Xinlong Wang, and Yue Cao. Eva: Exploring the limits of masked visual representation learning at scale. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19358–19369, 2023. 8

[10] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021. 2, 8

[11] Johannes Gilg, Torben Teepe, Fabian Herzog, and Gerhard Rigoll. The box size confidence bias harms your object detector. In *WACV*, pages 1471–1480, 2023. 4

[12] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007. 4

[13] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In Doina Precup and Yee Whye Teh, editors, *ICML*, volume 70, pages 1321–1330, 2017. 2, 3, 4

[14] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009. 4

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 5

[16] Yihui He, Xiangyu Zhang, Marios Savvides, and Kris Kitani. Softer-nms: Rethinking bounding box regression for accurate object detection. *arXiv preprint arXiv:1809.08545*, 2(3):69–80, 2018. 2

[17] Borui Jiang, Ruixuan Luo, Jiayuan Mao, Tete Xiao, and Yuning Jiang. Acquisition of localization confidence for accurate object detection. In *ECCV*, pages 784–799, 2018. 2

[18] Meelis Kull, Telmo Silva Filho, and Peter Flach. Beta calibration: a well-founded and easily implemented improvement on logistic calibration for binary classifiers. In *AISTATS*, volume 54, pages 623–631, 2017. 2, 4, 5

[19] Meelis Kull, Miquel Perello Nieto, Markus Kängsepp, Telmo Silva Filho, Hao Song, and Peter Flach. Beyond temperature scaling: Obtaining well-calibrated multi-class probabilities with dirichlet calibration. *NeurIPS*, 32, 2019. 2

[20] Abhinav Kumar, Garrick Brazil, and Xiaoming Liu. Groomed-nms: Grouped mathematically differentiable nms for monocular 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8973–8983, June 2021. 2

[21] Fabian Kuppers, Jan Kronenberger, Amirhossein Shantia, and Anselm Haselhoff. Multivariate confidence calibration for object detection. In *CVPRW*, pages 326–327, 2020. 2, 3, 4, 5

[22] Tingting Liang, Xiaojie Chu, Yudong Liu, Yongtao Wang, Zhi Tang, Wei Chu, Jingdong Chen, and Haibin Ling. Cbnet: A composite backbone network architecture for object detection. *TIP*, 31:6893–6906, 2022. 8

[23] David L Libby and Melvin R Novick. Multivariate generalized beta distributions with applications to utility assessment. *Journal of Educational Statistics*, 7(4):271–294, 1982. 4

[24] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *CVPR*, pages 2117–2125, 2017. 5

[25] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, pages 2980–2988, 2017. 2, 8

[26] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pages 740–755, 2014. (CC-BY 4.0): https://cocodataset.org. 3, 5, 6

[27] Songtao Liu, Di Huang, and Yunhong Wang. Adaptive nms: Refining pedestrian detection in a crowd. In *CVPR*, June 2019. 2

[28] Mahdi Pakdaman Naeini, Gregory Cooper, and Milos Hauskrecht. Obtaining well calibrated probabilities using bayesian binning. In *AAAI*, volume 29, 2015. 4

[29] Lukas Neumann, Andrew Zisserman, and Andrea Vedaldi. Relaxed softmax: Efficient confidence auto-calibration for safe pedestrian detection. In *NeurIPSW*, 2018. 4

[30] Jeremy Nixon, Michael W Dusenberry, Linchuan Zhang, Ghassen Jerfel, and Dustin Tran. Measuring calibration in deep learning. In *CVPRW*, volume 2, 2019. 4

[31] Ingram Olkin and Thomas A Trikalinos. Constructions for a bivariate beta distribution. *Statistics & Probability Letters*, 96:54–60, 2015. 4

[32] Jeffrey Ouyang-Zhang, Jang Hyun Cho, Xingyi Zhou, and Philipp Krähenbühl. Nms strikes back. *arXiv preprint arXiv:2212.06137*, 2022. 2

[33] John Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999. 2, 4

[34] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 8

[35] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *PAMI*, Jun 2017. 5, 8

[36] Franziska Schwaiger, Maximilian Henne, Fabian Küppers, Felippe Schmoeller Roza, Karsten Roscher, and Anselm Haselhoff. From black-box to white-box: examining confidence calibration under different conditions. *arXiv preprint arXiv:2101.02971*, 2021. 2, 6

[37] Roman Solovyev, Weimin Wang, and Tatiana Gabruseva. Weighted boxes fusion: Ensembling boxes from different object detection models. *Image and Vision Computing*, 107:104117, 2021. 2

[38] Peize Sun, Rufeng Zhang, Yi Jiang, Tao Kong, Chenfeng Xu, Wei Zhan, Masayoshi Tomizuka, Lei Li, Zehuan Yuan, Changhu Wang, et al. Sparse r-cnn: End-to-end object detection with learnable proposals. In *CVPR*, pages 14454–14463, 2021. 2, 7

[39] Rohun Tripathi, Vasu Singla, Mahyar Najibi, Bharat Singh, Abhishek Sharma, and Larry Davis. Asap-nms: Accelerating non-maximum suppression using spatially aware priors. *arXiv preprint arXiv:2007.09785*, 2020. 2

[40] Lachlan Tychsen-Smith and Lars Petersson. Improving object localization with fitness nms and bounded iou loss. In *CVPR*, June 2018. 2

[41] Xinlong Wang, Rufeng Zhang, Tao Kong, Lei Li, and Chunhua Shen. Solov2: Dynamic and fast instance segmentation. *NeurIPS*, 33:17721–17732, 2020. 2, 5

[42] Bianca Zadrozny and Charles Elkan. Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. In *ICML*, volume 1, pages 609–616, 2001. 2, 4

[43] Haoyang Zhang, Ying Wang, Feras Dayoub, and Niko Sunderhauf. Varifocalnet: An iou-aware dense object detector. In *CVPR*, pages 8514–8523, 2021. 8

[44] Hao Zhao, Jikai Wang, Deyun Dai, Shiqi Lin, and Zonghai Chen. D-nms: A dynamic nms network for general object detection. *Neurocomputing*, 512:225–234, 2022. 2

[45] Penghao Zhou, Chong Zhou, Pai Peng, Junlong Du, Xing Sun, Xiaowei Guo, and Feiyue Huang. Noh-nms: Improving pedestrian detection by nearby objects hallucination. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 1967–1975, 2020. 2

[46] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as points. *arXiv preprint arXiv:1904.07850*, 2019. 7