Table I: Ablation study about the quality branch and loss type.

From	Addition Branch	Classification Loss	FCOS [26]			ATSS [31]		
			AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}
paper	centerness	FL	38.5	56.8	41.6	39.2	57.4	42.2
paper	IoU	FL	38.7	56.7	**42.0**	39.6	57.6	**43.0**
paper	no	FL	37.8	56.2	40.8	38.0	56.5	40.7
rebuttal	no	L2	19.2	27.4	21.0	20.4	28.6	22.1
paper	no	QFL **(ours)**	**39.0**	**57.8**	41.9	**39.9**	**58.5**	**43.0**

Table II: Performance comparisons. ✓ of MS_{test} denotes multi-scale testing.

From	Method	Backbone	MS_{test}	FPS	AP
paper	ATSS [31]	X101-32x**8d**-DCN		6.9	47.7
rebuttal	ATSS [31]	X101-32x**4d**-DCN		10.0	47.4
paper	GFL **(ours)**	X101-32x**4d**-DCN		10.0	**48.2**
rebuttal	ATSS [31]	X101-32x8d-DCN	✓	-	50.7
rebuttal	ATSS [31]	X101-32x4d-DCN	✓	-	50.3
rebuttal	GFL **(ours)**	X101-32x4d-DCN	✓	-	**51.1**

Thanks a lot for the informative and constructive reviews. In general, the reviewers appreciate the novelty and motivation of GFL, but also raise several concerns about its contribution. We argue that the contributions of GFL are *significant*: We revisit the common problems in the classification/regression representations of recent dense detectors, and then provide an effective solution (GFL) for them. The proposed GFL is *simple* (without the need of extra quality estimation branch), *fast* (cost-free in speed) and *effective* (consistent 0.6∼1.4 AP improvement). It provides a practical and general format for the detection head, which is compatible for most dense detectors. Therefore, we think that GFL has the potential to be widely applied in the field of dense object detection.

To Reviewer #1 **Q1**: Ablation study about the separate branch and different forms of loss function. **A1**: We conduct additional experiments with L2 regression loss to compare against QFL (Table I). Some other related results from Table 1(a) in the original paper are also included in Table I. We observe that (1) ignoring the separate branch considerably hurts the performance (FCOS: 38.5→37.8, ATSS: 39.2→38.0); (2) the proposed new loss function (QFL) is essential since the performance of simple L2 regression loss drops dramatically (FCOS: 39.0→19.2, ATSS: 39.9→20.4). The major reason is that the simple regression loss lacks the good property of Focal Loss and GFL which handle the class imbalance problem well. We will update these analyses to the paper in the later version.

Q2: Quality of the predicted IoUs. **A2**: We calculate the Pearson correlation coefficient between the predicted IoUs and the actual IoUs (R-50) over all the validation images. The statistic of GFL is 0.78, which is larger than ATSS (0.72). It shows that GFL indeed improves the quality of the predicted IoUs. We will update these results in the revised version.

Q3: Meanings of Δ and regression target. **A3**: $\Delta = y_{i+1} - y_i, \forall i \in [0, n-1]$. "regression target" is defined in "we adopt the relative offsets from the location to the four sides of a bounding box as the regression targets", which is in line 156-157 of the original paper. The regression targets are obtained from each positive location and its corresponding gt bbox, divided by the stride value of its corresponding FPN level. We will make it clearer for readers.

To Reviewer #2 **Q1**: DFL in regression tasks. **A1**: Here the "Focal" in DFL has a completely different meaning: it forces the network to rapidly *focus* on the probabilities near the target label (line 174-182). Instead, the "Focal" in QFL means *focusing* the model on hard examples. They have different meanings but share a generalized formulation (GFL).

Q2: Could bbox confidence benefit the NMS? **A2**: Yes. We use variance voting method in "Softer-nms" and improve GFL (R-50) by 0.2 AP. In "Softer-nms", variance voting achieves similar gains for Faster RCNN, i.e., ∼0.3 AP.

Q3: Relation between DFL and the keypoint detection paper. **A3**: Although the integral form seems similar in keypoint detection community, our work is the first to introduce the *integral form* of a *General distribution* into the object detection field. Meanwhile, we also provide a derivation by extending the concept of Dirac delta distribution, from a theoretical perspective. Further, we design a novel DFL that quickly focuses on learning probabilities near gt labels. We will cite and discuss these related works (including LCRNet, etc. mentioned by Reviewer #3) in the revised manuscript.

To Reviewer #3 **Q1**: Unclear speed of methods (ATSS: 6.9 FPS, GFL: 10.0 FPS) with X-101-DCN & multi-scale testing. **A1**: There are some misunderstanding here. The gap of the speeds only comes from the backbone part of ATSS (X-101-32x**8d**-DCN) and GFL (X-101-32x**4d**-DCN), where the bottleneck feature dimension of ATSS (32x**8d**) is **twice** that of GFL (32x**4d**). GFL is cost-free and has the same speed as ATSS under the same backbone (Table 3, II). Due to GPU memory constraint (11G), we cannot train models with very large backbone (X-101-32x**8d**-DCN). So we make more comparisons between GFL and ATSS with backbone X-101-32x4d-DCN in Table II. The multi-scale testing results are also included. We observe that GFL improves ATSS by 0.8 AP under the same X-101-32x4d-DCN backbone, and the multi-scale testing result of GFL really pushes the state-of-the-art number of ATSS (50.7→51.1).

Q2: About Fig. 3. **A2**: Fig. 3 is a real output of the model (R-50). The config and pretrained model are in Supplementary Material (SM), which can reproduce these figures exactly. More results (Fig. 13, 14) are also provided in SM.

Q3: Details of changing from centerness to IoU & limited "valuable improvement". **A3**: The standard version of FCOS/ATSS defaults to use the centerness branch. Our proposed QFL is designed to optimize a novel "classification-*quality*" joint representation, whilst identifying the "*quality*" here to be "IoU" instead of "centerness" deserves a "valuable improvement" because there is no previous work pointing out that issue. Further, we give a deep and statistical analysis to prove that IoU is better than centerness as a quality measurement (see **Section E** of SM), exposing an informative fact that centerness has a *fatal flaw* in its definition: centerness can make some quality gt labels too small (close to 0) to recall a set of objects. For the first time, we provide a convincing reason and suggest that the community should use IoU instead of centerness although centerness is very successful in FCOS and ATSS. We will add these analyses into the main paper to make our unique contributions (including the DFL part, see **#2 A3**) clearer.

To Reviewer #4 **Q1**: The presentation. **A1**: Thanks for the kind suggestions about the presentation regarding the title, abstract and formulations. We will try our best to streamline the abstract and title, and clarify possible formulations.