SUPPLEMENTARY MATERIAL
for
ARUBA: An Architecture-Agnostic Balanced Loss for Aerial Object Detection

In this supplementary material, we include the following details, which we could not include in the main paper owing to space constraints:

- Additional qualitative results
- Illustrative example of the Gaussian amplification step in our pipeline
- Justification for class-wise segregation
- Discussion of failure cases

Additional Qualitative Results. Figure S1 shows additional qualitative results on the HRSC2016 dataset [4]. These results continue to support our claim that our approach is able to result in prediction of more small objects compared to the baseline.

Figure S2 shows additional qualitative results which shows how our method reduces false positives when compared to baseline methods. The figure shows bounding box predictions on images of the DOTA_v1.5 dataset [1] using the baseline model ReDet [3] vs Ours. As observed, the baseline method *(top)* wrongly predicts few small objects as planes (indicated in red boxes) whereas our method succeeds in reducing false positives because of the effectiveness of the proposed **AR**chitect**U**re-agnostic **BA**lanced (**ARUBA**) Loss.

Illustrative Example of Gaussian Amplification (GA). To demonstrate how Gaussian amplification adds the context of size neighborhood, we consider an example. Let $B = (1, 100, \mathbf{20}, 10, \mathbf{20}, 2)$ be the size distribution, i.e. frequency counts of object instances falling into different size bins (Figure S3a). Note that B has two bins with number of object instances **20** ($b3$ and $b5$). Let the window size w and variance σ be 5 and 1 respectively. Based on the properties that we defined (in the main paper, Sec 3), the value of our discrete Gaussian kernel is $K = (0.14, 0.60, 1, 0.60, 0.14)$. When we convolve this kernel with B, we get $B' = (64, 114, \mathbf{89}, 49, \mathbf{30}, 15)$ (Figure S3b). After convolution with K, the bin (**20**) that has a high neighborhood of 100 gets more amplified compared to the other (**20**). Also, the instances in bin 100 increase to 114. Although two size bins ($b3$ and $b5$ in Figure S3a) have same number of object instances, they experience different levels of imbalance because of the difference in the neighborhood. We leverage this observation by applying Gaussian amplification.

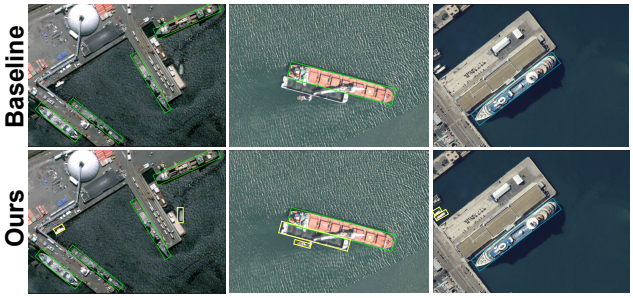


Figure S1: Predictions on images of HRSC2016 dataset [4] using ReDet [3] vs Ours. **Top**: The baseline method fails to detect small sized objects. **Bottom**: Ours is able to recognize additional small objects because of our **AR**chitect**U**re-agnostic **BA**lanced (ARUBA) loss. Yellow boxes indicate objects additionally detected.

Justification for Class-wise Segregation. As stated in Sec 3 of the main paper, the first step in our pipeline is the class-wise segregation of training object instances. To explain the need for this step, we extracted the Car and $Pedestrian$ classes from the VisDrone dataset [2]. We divided the objects of train and validation sets into three kinds - $small$, $medium$ and $large$, based on their sizes. We trained the baseline model [3] on data from both these classes on the $small$ category of objects and tested its performance on the three kinds of objects from the Car class. Table S1 summarizes these results. We observed that the model's performance on medium-sized Car instances is because of the small instances from the Car class and not the $Pedestrian$ class. This suggests that the effect of neighborhood should be considered within a class rather than across classes, supporting the need for this step in our pipeline.

Failure Cases. Figure S4 shows predictions of both the baseline [3] and our model on an image from the

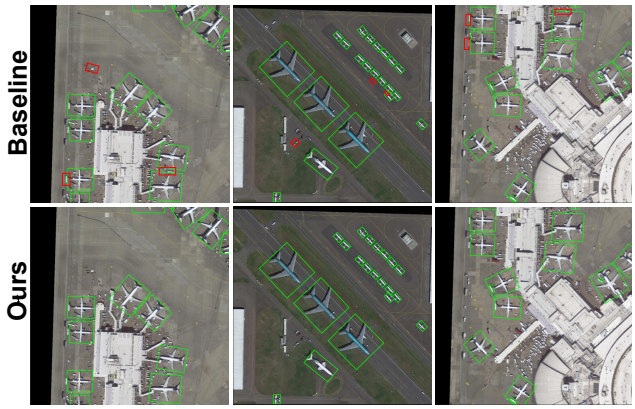
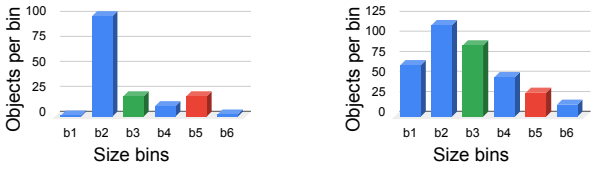


Figure S2: Predictions on images from DOTA_v1.5 dataset [1] using ReDet [3] vs Ours. The results show how our method reduces false positives (shown in red boxes in the figure), **Top:** The baseline method struggles while predicting small-sized objects and results in false positives. **Bottom:** Ours is able to detect small objects accurately.



(a) Before applying GA (b) After applying GA

Figure S3: Working of Gaussian Amplification(GA): **a)** Size distribution before applying GA. **b)** Size distribution after applying GA.

Trained on	Tested on		
	Small	Medium	Large
Small	12.28	6.82	1.21

Table S1: Performance of baseline [3] on two object categories of VisDrone dataset [2]. The train and val sets are divided into three bins - $small$, $medium$, and $large$.

HRSC2016 dataset. We observe that both the baseline and our method incorrectly predict a large object as 'Ship' (indicated in red box). Although our method improves performance on small objects in numerous images (as evident from Figure S1), there are a few cases where our model performs the same as the baseline on large objects. Handling this challenge will be an interesting direction of future work.

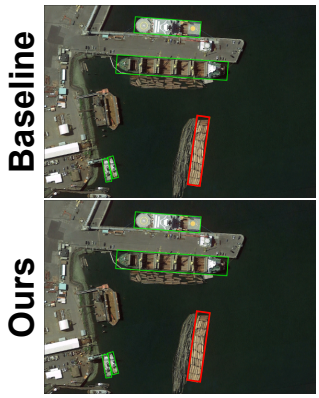


Figure S4: *Failure case.* Predictions on an image from HRSC2016 [4] dataset using ReDet [3] vs Ours. This figure shows a case where both baseline and our method results in a false positive while predicting a large-sized object (indicated with red boxes).

References

[1] Dota1.5 dataset: Object detection in aerial images. https://captain-whu.github.io/DOAI2019/dataset.html. 1, 2

[2] Dawei Du, Pengfei Zhu, Longyin Wen, Xiao Bian, Haibin Lin, Qinghua Hu, Tao Peng, Jiayu Zheng, Xinyao Wang, Yue Zhang, Liefeng Bo, Hailin Shi, Rui Zhu, Aashish Kumar, Aijin Li, Almaz Zinollayev, Anuar Askergaliyev, Arne Schumann, Binjie Mao, Byeongwon Lee, Chang Liu, Changrui Chen, Chunhong Pan, Chunlei Huo, Da Yu, DeChun Cong, Dening Zeng, Dheeraj Reddy Pailla, Di Li, Dong Wang, Donghyeon Cho, Dongyu Zhang, Furui Bai, George Jose, Guangyu Gao, Guizhong Liu, Haitao Xiong, Hao Qi, Haoran Wang, Heqian Qiu, HongLiang Li, Huchuan Lu, Ildoo Kim, Jaekyum Kim, Jane Shen, Jihoon Lee, Jing Ge, Jingjing Xu, Jingkai Zhou, Jonas Meier, Jun Won Choi, Junhao Hu, Junyi Zhang, Junying Huang, Kaiqi Huang, Keyang Wang, Lars Sommer, Lei Jin, and Lei. Zhang. Visdrone-det2019: The vision meets drone object detection in image challenge results. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, Oct 2019. 1, 2

[3] Jiaming Han, Jian Ding, Nan Xue, and Gui-Song Xia. Redet: A rotation-equivariant detector for aerial object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2786–2795, 2021. 1, 2

[4] Zikun Liu, Liu Yuan, Lubin Weng, and Yiping Yang. A high resolution optical satellite image dataset for ship recognition and some new baselines. In *International conference on pattern recognition applications and methods*, volume 2, pages 324–331. SCITEPRESS, 2017. 1, 2