Supplementary Material for "Any-Shot Object Detection"

Shafin Rahman[1,2,3][0000−0001−7169−0318], Salman Khan[4,2][0000−0002−9502−1749], Nick Barnes[2][0000−0002−9343−9535], and Fahad Shahbaz Khan[4][0000−0002−4263−3143]

[1] North South University, Dhaka, Bangladesh
[2] Australian National University, Canberra, ACT 2601, Australia
[3] Data61, CSIRO, Canberra, ACT 2601, Australia
[4] MBZ University of Artificial Intelligence, Abu Dhabi, UAE
shafin.rahman@northsouth.edu,nick.barnes@anu.edu.au,
{salman.khan,fahad.khan}@mbzuai.ac.ae

This supplementary material provides additional qualitative results, gradient analysis for the proposed loss function and validation experiments for the hyper-parameters.

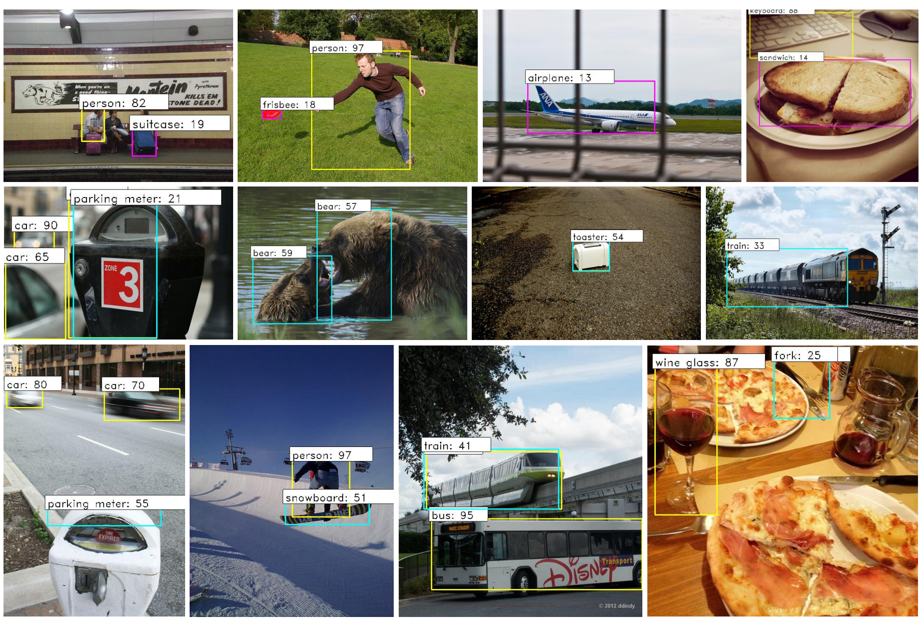

Fig. 1. More qualitative results for generalized ASD. Yellow, blue and pink boxes represent seen, few-shot and unseen objects, respectively. The proposed approach can detect classes from all the three categories.

β \ λ	0.1	0.2	0.3	0.4	0.5	0.7	0.9	1.0
0.5	30.90	30.22	28.70	26.92	24.84	22.53	17.23	0.23
1	29.70	28.82	29.25	28.06	26.96	23.73	16.83	0.24
2	30.04	29.78	29.07	29.93	27.05	23.99	18.42	0.13
5	**31.54**	31.54	30.50	29.52	27.64	25.83	17.14	0.26

Table 1. Validation studies with 5-shot detection performance

1 Qualitative Results

In Fig. 1, we show additional qualitative results of our approach on the ASD task. All visual results are generated with a single ASD model that can detect seen, unseen and few-shot classes, simultaneously.

2 Gradient analysis

In this section, we derive the gradients of our proposed loss w.r.t p. For simplicity, we only focus on the predictions corresponding to the ground-truth class. We show the gradient curve in Fig. 4(d) of the main paper.

$$\frac{\partial L}{\partial p} = \begin{cases} \frac{1}{p\,(p-p_*-1)\left(p-(1-p+p_*)^{\beta}\right)}\Big[\alpha_t\big(\,(\beta-1)\,p \\ \qquad +p_*+1\big)\left(1-\frac{p}{(1-p+p_*)^{\beta}}\right)^{\gamma}\big(\gamma p\log(p) \\ \qquad +(1-\beta\gamma\log(1-p+p_*))\,p-(1-p+p_*)^{\beta}\,\big)\Big] \\ \qquad \text{if } \Big(\alpha_t\beta\left(1-\frac{p}{(1-p+p_*)^{\beta}}\right)^{\gamma}\log(1-p+p_*) \\ \qquad -\alpha_t\left(1-\frac{p}{(1-p+p_*)^{\beta}}\right)^{\gamma}\log(p)\Big) > 0 \text{ and } y=1 \\ \frac{\alpha_t p^{\gamma}}{1-p}-\alpha_t\gamma\log(1-p)\,p^{\gamma-1} \\ \qquad \text{if } \big(\alpha_t p^{\gamma}\log(1-p)\big) < 0 \text{ and } y=0 \\ 0 \qquad \text{otherwise.} \end{cases}$$

3 Validation Set Experiments

Note that α, β, γ and λ are the hyper-parameter of our model. Among them, α and γ are focal loss hyper-parameter. Thus, we fix the recommended value of $\alpha = 0.25$ and $\gamma = 2$ as per the original RetinaNet model [1]. β and λ are the hyper-parameters proposed in our loss formulation. Therefore, we validate β and λ by further splitting 65 seen classes into 55 seen and 10 novel classes to create a validation set. Then, we perform a 5-shot detection task on this validation split. We report the validation results in Table 1. From our validation study, we select $\beta = 5$ and $\lambda = 0.1$ and use this value in all of our experiments.

References

1. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence (2018)