Supplementary Materials for
Bridge Past and Future: Overcoming Information Asymmetry in Incremental Object Detection

Qijie Mo[1,3], Yipeng Gao[1,3], Shenghao Fu[1,3], Junkai Yan[1,3], Ancong Wu[1,3,⋆], and Wei-Shi Zheng[1,2,3,⋆]

[1] School of Computer Science and Engineering, Sun Yat-sen University, China
[2] Peng Cheng Laboratory, Shenzhen, China
[3] Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China
{moqj3,gaoyp23,fushh7,yanjk3}@mail2.sysu.edu.cn, wuanc@mail.sysu.edu.cn, wszheng@ieee.org

In this supplementary material, we present the detailed results for the 5-5 multi-step incremental setting (Section S1), more discussion on possible limitations (Section S2), and some ablation studies on selection metrics for potential future class objects in Bridge the Future (Section S3). We also provide a detailed process of Distillation with Future (Section S4) and more analysis on the necessity for Distillation with Future (Section S5).

S1 Detailed Results for the Multi-Step Incremental Setting

Table S1 presents the results of our experiments under 5-5 multi-step incremental setting on the PASCAL-VOC 2007 dataset [2]. We simulate this scenario by training the detector on images from the first 5 classes and then adding the $6^{th} \sim 20^{th}$ classes five by five. The table shows the class-wise average precision (AP@0.5) and the corresponding mean average precision (mAP@0.5). Our proposed BPF method outperforms the previous state-of-the-art rehearsal-based method ABR [5].

Table S1: Per-Class AP@0.5 and Overall mAP@0.5 values in different task on PASCAL-VOC 2007 5-5 setting.

Class Split	Method	aero	cycle	bird	boat	bottle	bus	car	cat	chair	cow	mAP-task1	table	dog	horse	bike	person	mAP-task2	plant	sheep	sofa	train	tv	mAP-task3	mAP-total
1-20	Joint Training	76.4	84.7	77.1	62.7	61.3	82.2	87.5	85.9	57.8	83.0	75.9	70.5	84.4	86.7	83.7	85.6	82.2	46.7	77.7	70.7	80.3	76.7	70.4	76.1
(1-5)+	ABR [5]	71.7	82.6	69.5	53.6	63.8	63.0	79.0	68.5	47.0	78.4	67.7													67.7
6-10	**BPF (Ours)**	70.1	78.3	68.3	55.9	60.1	64.9	81.3	73.3	52.3	79.8	**68.4**													**68.4**
(1-10)+	ABR [5]	68.5	79.6	67.3	51.9	56.7	60.2	75.2	62.8	38.6	62.0	62.3	54.0	66.3	76.9	74.5	77.3	69.8							64.8
11-15	**BPF (Ours)**	69.2	79.8	64.9	54.5	56.4	65.3	80.6	68.8	48.3	71.9	**66.0**	58.6	68.8	77.5	74.5	80.6	**72.0**							**68.0**
(1-15)+	ABR [5]	69.3	80.0	65.6	53.9	54.6	52.2	75.5	69.4	34.3	69.6	**62.4**	22.9	41.8	48.7	53.7	60.8	45.6	39.6	71.3	59.2	76.1	70.4	**63.3**	58.4
16-20	**BPF (Ours)**	64.8	75.9	56.7	51.8	53.8	51.1	78.8	63.6	47.1	63.5	60.7	56.2	64.1	75.9	71.2	79.4	**69.4**	40.4	61.3	61.7	61.8	70.5	59.1	**62.5**

⋆ Corresponding Author

S2 Discussion on Possible Limitations

As discussed in Section 5 in the main text, our method is based on a prior that the old classes always appear in the new training dataset so that they can be utilized to prevent catastrophic forgetting. However, when the newly introduced dataset is too small to find sufficient old class objects, especially in incremental scenarios that only add a few classes at each stage, our Bridge the Past method cannot be performed effectively, which hinders both old and new classes performance as the balance of old and new classes is a trade-off.

However, the limitation can be simply addressed by using some memory, *i.e.*, employing the same copy-paste strategy as in ABR [5] to paste some old class objects on the new training images. We follow ABR [5] and set the memory size as 2,000 for the experiments, which stores a selection of instances for all known categories. As illustrated in Table S2, with the integration of memory from both original classes and the classes subsequently added from incremental stages, our method shows significant improvement for both old and new classes. Especially in the 10-1 settings, our method achieves a 9.1% enhancement across all newly introduced classes, increasing from 48.3% to 57.4%. The results effectively highlight that the limitations of our method can be addressed by storing instances in memory.

Table S2: mAP@0.5 results on multiple incremental steps on Pascal-VOC 2007. The best performance in each is presented with **bold**, and the second best is presented with underlined. Methods with $*$ use exemplars.

Method	15-1 (6 tasks)			10-1 (10 tasks)		
	1-15	16-20	1-20	1-10	11-20	1-20
Joint Training	78.0	70.4	76.1	75.9	76.3	76.1
Fine-tuning [5]	0.0	10.5	5.3	0.0	5.1	2.6
Faster ILOD [6]	66.9	44.5	61.3	52.9	41.5	47.2
MMA [1]	68.3	54.3	64.1	59.2	48.3	53.8
BPF (Ours)	**71.5**	53.1	<u>66.9</u>	<u>62.2</u>	48.3	55.2
ABR* [5]	68.7	<u>56.7</u>	65.7	62.0	<u>55.7</u>	<u>58.9</u>
BPF (Ours)*	<u>71.3</u>	**57.3**	**67.8**	**62.7**	**57.4**	**60.1**

S3 Additional Analysis for Bridge the Future

In Bridge the Future, regions with both high attention scores from feature maps and objectness scores from class-agnostic RPN are excluded from negative samples when training the RoI head, thus maintaining the model's consistency with future stage background definitions. In Table S3, we show that both the RPN and spatial attention can independently identify potential foreground objects. And combining them can more accurately identify them, thus performing better.

Table S3: Ablation Study on selection metrics for potential future class objects in Bridge the Future.

RPN objectness	Attention scores	VOC(10-10) 1-10	11-20	1-20
		71.2	73.3	72.3
	✓	71.4	73.7	72.6
✓		71.3	73.7	72.5
✓	✓	**71.7**	**74.0**	**72.9**

S4 Detailed Process of Distillation with Future

Algorithm 1 shows the detailed process of Distillation with Future strategy. Firstly, we obtain 64 proposals $\mathcal{R}$ for distillation, which are randomly selected out of the top 128 proposals with the highest objectness scores from the RPN network of the old model $\mathcal{M}_{t-1}$. Then we divide the $\mathcal{R}$ into $\mathcal{R}_1, \mathcal{R}_2$ based on the IoU with ground truth labels $\mathcal{Y}_t$. For region r_i in $\mathcal{R}_1$, which is likely to be the region for old classes, we take $\mathcal{M}_{t-1}$ as the primary model and reconstruct its background representation with the model $\mathcal{M}_t^{im}$. Then we get the final distillation probabilities: $p_i^{dist} = [p_i^{\mathcal{C}_{1:t-1},t-1}, \hat{p}_i^{\mathcal{C}_t \cup \mathcal{B},im}] \in \mathbb{R}^{|\mathcal{C}_{1:t}|+1}$. On the contrary, for region r_i in $\mathcal{R}_2$, the final distillation probabilities are: $p_i^{dist} = [\hat{p}_i^{\mathcal{C}_{1:t-1},t-1}, p_i^{\mathcal{C}_t,im}, \hat{p}_i^{\mathcal{B},t-1}] \in \mathbb{R}^{|\mathcal{C}_{1:t}|+1}$. After getting the probabilities p_i from current model $\mathcal{M}_t$ such that $p_i^t \in \mathbb{R}^{|\mathcal{C}_{1:t}|+1}$, the probability distillation loss is:

$$\mathcal{L}_{dist,cls}^{roi} = \mathcal{L}_{KL}(p_i^{dist}, p_i^t), \tag{1}$$

where $\mathcal{L}_{KL}$ represents Kullback-Leibler divergence. Regarding the box distillation, we take the output boxes b_i^{dist} from the old model $\mathcal{M}_{t-1}$ for regions in $\mathcal{R}_1$ and the intermediate model $\mathcal{M}_t^{im}$ for regions in $\mathcal{R}_2$. The box distillation loss is:

$$\mathcal{L}_{dist,bbox}^{roi} = \mathcal{L}_2(b_i^{dist}, b_i^t), \tag{2}$$

where $\mathcal{L}_2$ represents the L2 loss.

The final distillation loss in RoI Head is:

$$\mathcal{L}_{dist}^{roi} = \mathcal{L}_{dist,cls}^{roi} + \mathcal{L}_{dist,bbox}^{roi}. \tag{3}$$

Table S4: The recall@50 of proposals for distillation generated by the old model's RPN for old classes 1-10 and new classes 11-20. The results are obtained from the second incremental task under the PASCAL VOC 10-10 setting.

Classes	old classes (1-10)	new classes (11-20)
Recall@50	70.5	58.6

Algorithm 1 Distillation with Future

Input: Previous model: $\mathcal{M}_{t-1}$; Current model: $\mathcal{M}_t$; Current dataset: $\mathcal{D}_t$; Current ground truth $\mathcal{Y}_t$;.

Initialization: $\mathcal{L}_{dist}^{roi} = 0$

Get intermediate model $\mathcal{M}_t^{im}$, trained using $\mathcal{D}_t$ in a fully supervised way

for all $\mathcal{I} \in \mathcal{D}_t$ **do**

  Get proposals $\mathcal{R}$ for distillation, generated by $\mathcal{M}_{t-1}$

  Divide $\mathcal{R}$ into $\mathcal{R}_1, \mathcal{R}_2 \subset \mathcal{R}$, based on their intersection over union with $\mathcal{Y}_t$

  for all $r_i \in \mathcal{R}_1$ **do**

    Get probabilities $p_i^{t-1}, p_i^{im}, p_i^t$ from $\mathcal{M}_{t-1}, \mathcal{M}_t^{im}, \mathcal{M}_t$

    Reconstruct M_{t-1} background representation: $\hat{p}_i^{c,im} = p_i^{c,im} \times p_i^{b,t-1}$

    Get distillation probabilities: $p_i^{dist} = [p_i^{\mathcal{C}_{1:t-1},t-1}, \hat{p}_i^{\mathcal{C}_t \cup \mathcal{B},im}] \in \mathbb{R}^{|\mathcal{C}_{1:t}|+1}$

    Get probability distillation loss: $\mathcal{L}_{dist,cls}^{roi} = \mathcal{L}_{KL}(p_i^{dist}, p_i^t)$

    Get boxes b_i^{t-1}, b_i^t from $\mathcal{M}_{t-1}, \mathcal{M}_t$

    Get box distillation loss: $\mathcal{L}_{dist,bbox}^{roi} = \mathcal{L}_2(b_i^{t-1}, b_i^t)$

    Get distillation loss: $\mathcal{L}_{dist}^{roi} = \mathcal{L}_{dist}^{roi} + \mathcal{L}_{dist,cls}^{roi} + \mathcal{L}_{dist,bbox}^{roi}$

  end for

  for all $r_i \in \mathcal{R}_2$ **do**

    Get probabilities $p_i^{t-1}, p_i^{im}, p_i^t$ from $\mathcal{M}_{t-1}, \mathcal{M}_t^{im}, \mathcal{M}_t$

    Reconstruct M_t^{im} background representation: $\hat{p}_i^{c,t-1} = p_i^{c,t-1} \times p_i^{b,im}$

    Get distillation probabilities: $p_i^{dist} = [\hat{p}_i^{\mathcal{C}_{1:t-1},t-1}, p_i^{\mathcal{C}_t,im}, \hat{p}_i^{\mathcal{B},t-1}] \in \mathbb{R}^{|\mathcal{C}_{1:t}|+1}$

    Get probability distillation loss: $\mathcal{L}_{dist,cls}^{roi} = \mathcal{L}_{KL}(p_i^{dist}, p_i^t)$

    Get boxes b_i^{im}, b_i^t from $\mathcal{M}_t^{im}, \mathcal{M}_t$

    Get box distillation loss: $\mathcal{L}_{dist,bbox}^{roi} = \mathcal{L}_2(b_i^{im}, b_i^t)$

    Get distillation loss: $\mathcal{L}_{dist}^{roi} = \mathcal{L}_{dist}^{roi} + \mathcal{L}_{dist,cls}^{roi} + \mathcal{L}_{dist,bbox}^{roi}$

  end for

end for

Output: Distillation with Future loss: $\mathcal{L}_{dist}^{roi}$

S5 Additional Analysis for Distillation with Future

Knowledge distillation [3] serves as a potent strategy in incremental object detection. Previous methods [1,4,5] transfer the knowledge of the old model to the current model using the current dataset. The old model is usually biased toward old classes since it has not trained with new classes, making it hard to extract good representations for new classes. However, we find that a number of proposals used in the distillation process do contain the newly introduced classes, and directly distilling the old model's knowledge of these proposals is not wise since the old model has not learned the new classes.

Specifically, Table S4 shows that during the second stage of the PASCAL VOC 10-10 setting, the distilled proposals achieve a recall rate of 70.5% for old classes and **58.6% for objects of newly introduced classes**. This highlights the importance of our proposed intermediate model, which excels in detecting objects of newly introduced classes. During the distillation process, we select the old model as the primary distillation model in regions containing old class objects, whereas the intermediate model is chosen as the primary in areas with new class objects. The other model serves as an auxiliary, reconstructing the background probability information of the primary distillation model. This approach provides complementary and comprehensive guidance to the current model.

References

1. Cermelli, F., Geraci, A., Fontanel, D., Caputo, B.: Modeling missing annotations for incremental learning in object detection. In: CVPR (2022)
2. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. Springer IJCV (2010)
3. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
4. Joseph, K., Rajasegaran, J., Khan, S., Khan, F.S., Balasubramanian, V.N.: Incremental object detection via meta-learning. IEEE TPAMI (2021)
5. Liu, Y., Cong, Y., Goswami, D., Liu, X., van de Weijer, J.: Augmented box replay: Overcoming foreground shift for incremental object detection. In: ICCV (2023)
6. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: NeurIPS (2015)