

Incremental Object Detection via Future-Aware Decoupled Cross-Head Distillation

Supplementary Material

Table 6. Ablation study of RPCS on VOC 10+10: comparison with prototype-based methods without semantic calibration. AP is reported for Old, New, All, and task averaged (Avg).

Methods	Old	New	All
Baseline	75.7	75.0	75.4
Without Semantic Calibration	74.9	75.4	75.1
RPRC	76.0	75.9	75.9

A. Training pipeline for FaCHD

Algorithm 1 illustrates the training details of our method

B. Effect of Semantic Calibration

Table 6 compares a prototype method without semantic drift calibration against its counterpart with semantic drift compensation under the VOC 10-10 setting. The baseline uses only the FaCHD method, Without Semantic Calibration denotes the baseline augmented with prototype replay but without semantic drift calibration, and RPRC denotes the baseline combined with semantic drift compensation. With semantic drift compensation, all metrics improve, with the performance on old classes increasing from 74.9% to 76.0%, on new classes from 75.4% to 75.9%, and on all classes from 75.1% to 75.9%. It is worth noting that directly applying prototype replay without semantic drift compensation leads to a drop in performance, which further highlights the necessity and effectiveness of semantic drift compensation.

C. Analysis of detection predictions

As shown in Figure 3, we performed a qualitative comparison of the detection results between the baseline method and the proposed method on the PASCAL VOC 2007 dataset with the 10+10 setting. In Figure 3(a), our method alleviated the forgetting of the old class *boat*, while the baseline method missed detecting the boat object. In Figure 3(b), our method reduced the repeated detection of the old class *person*. In Figure 3(c), our method prevented the misclassification of the new class *person* as the old class *chair*. These results demonstrate that our method effectively mitigates both the missed detections and misclassifications between new and old classes while maintaining strong detection performance for the old class objects.

Algorithm 1 Future-Aware Decoupled Cross-Head Distillation

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

- 1: Initialization: $\mathcal{L}_{\text{FaCHD}} = 0$.
- 2: Get intermediate model $\mathcal{M}_t^{\text{im}}$, trained using \mathcal{D}_t in a fully supervised way.
- 3: **for all** $\mathcal{I} \in \mathcal{D}_t$ **do**
- 4: Get proposals \mathcal{R} for distillation, generated by \mathcal{M}_{t-1} .
- 5: Divide \mathcal{R} into $\mathcal{R}_1, \mathcal{R}_2 \subset \mathcal{R}$, based on their intersection over union with \mathcal{Y}_t .
- 6: **for all** $\mathbf{r} \in \mathcal{R}_1$ **do**
- 7: Get prediction $\mathbf{p}_r^{t-1}, \mathbf{p}_r^{\text{im}}$ from $\mathcal{M}_{t-1}, \mathcal{M}_t^{\text{im}}$.
- 8: Reconstruct \mathcal{M}_{t-1} background representation: $\hat{\mathbf{p}}_r^{c,\text{im}} = \mathbf{p}_r^{c,\text{im}} \cdot \mathbf{p}_r^{b,t-1}$.
- 9: Get distillation prediction by concatenating: $\bar{\mathbf{p}}_r = \text{concat}(\hat{\mathbf{p}}_r^{c,\text{im}}, \mathbf{p}_r^{c,t-1})$.
- 10: Get ROI features \mathbf{z} from \mathcal{M}_t .
- 11: Get cross-head prediction from \mathcal{M}_{t-1} and $\mathcal{M}_t^{\text{im}}$: $\mathbf{p}^{ch,t-1} = \text{softmax}(\mathcal{H}_{t-1}(\mathbf{z}))$, $\mathbf{p}^{ch,\text{im}} = \text{softmax}(\mathcal{H}_t^{\text{im}}(\mathbf{z}))$.
- 12: Reconstruct cross-head prediction: $\hat{\mathbf{p}}_r^{ch,c,\text{im}} = \mathbf{p}_r^{ch,c,\text{im}} \cdot \mathbf{p}_r^{ch,b,t-1}$.
- 13: Get joint cross-head prediction: $\bar{\mathbf{p}}_r^{ch} = \text{concat}(\hat{\mathbf{p}}_r^{ch,c,\text{im}}, \mathbf{p}_r^{ch,c,t-1})$.
- 14: $\mathcal{L}_{\text{FaCHD}} = L_{\text{KL}}(\bar{\mathbf{p}}_r, \bar{\mathbf{p}}_r^{ch})$.
- 15: **end for**
- 16: **for all** $\mathbf{r} \in \mathcal{R}_2$ **do**
- 17: Get prediction $\mathbf{p}_r^{t-1}, \mathbf{p}_r^{\text{im}}$ from $\mathcal{M}_{t-1}, \mathcal{M}_t^{\text{im}}$.
- 18: Reconstruct \mathcal{M}_{t-1} background representation: $\hat{\mathbf{p}}_r^{c,t-1} = \mathbf{p}_r^{c,t-1} \cdot \mathbf{p}_r^{b,\text{im}}$.
- 19: Get distillation prediction by concatenating: $\bar{\mathbf{p}}_r = \text{concat}(\hat{\mathbf{p}}_r^{c,t-1}, \mathbf{p}_r^{c,\text{im}})$.
- 20: Get ROI features \mathbf{z} from \mathcal{M}_t .
- 21: Get cross-head prediction from \mathcal{M}_{t-1} and $\mathcal{M}_t^{\text{im}}$: $\mathbf{p}^{ch,t-1} = \text{softmax}(\mathcal{H}_{t-1}(\mathbf{z}))$, $\mathbf{p}^{ch,\text{im}} = \text{softmax}(\mathcal{H}_t^{\text{im}}(\mathbf{z}))$.
- 22: Reconstruct cross-head prediction: $\hat{\mathbf{p}}_r^{ch,c,t-1} = \mathbf{p}_r^{ch,c,t-1} \cdot \mathbf{p}_r^{ch,b,\text{im}}$.
- 23: Get joint cross-head prediction: $\bar{\mathbf{p}}_r^{ch} = \text{concat}(\hat{\mathbf{p}}_r^{ch,c,t-1}, \mathbf{p}_r^{ch,c,\text{im}})$.
- 24: $\mathcal{L}_{\text{FaCHD}} = L_{\text{KL}}(\bar{\mathbf{p}}_r, \bar{\mathbf{p}}_r^{ch})$.
- 25: **end for**
- 26: **end for**

Outputs: Future-Aware Decoupled Cross-Head Distillation loss: $\mathcal{L}_{\text{FaCHD}}$.

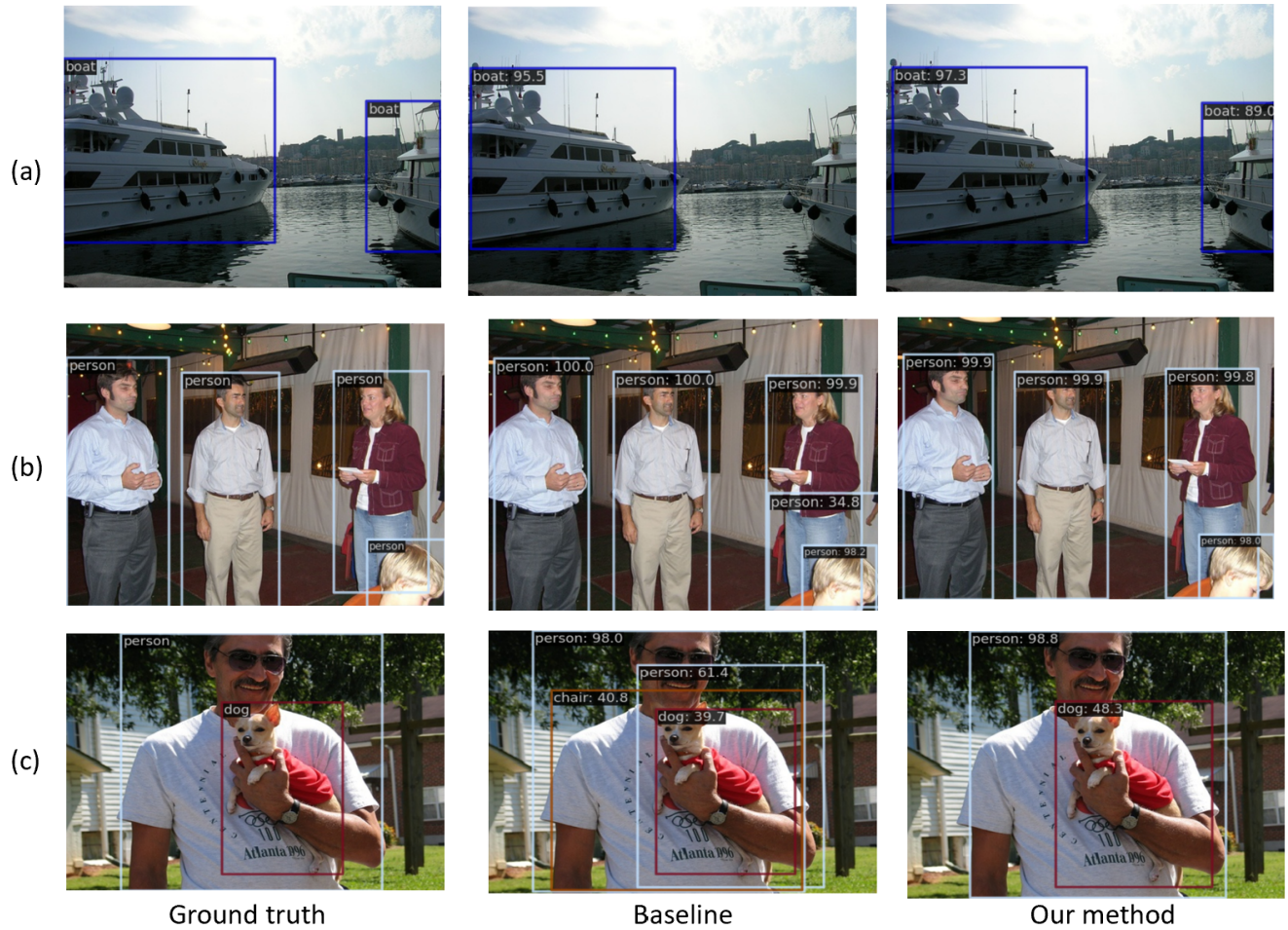


Figure 3. Qualitative visualization of detection predictions from the baseline and our method on PASCAL VOC 2007 (10+10) setting