Supplementary Material for "Few-shot Adaptive Object Detection with Cross-Domain CutMix"

Yuzuru Nakamura[1*†], Yasunori Ishii[1*], Yuki Maruyama[1], and Takayoshi Yamashita[2]

[1] Panasonic Holdings Corporation,
3-1-1 Yagumo-naka-machi, Moriguchi City, Osaka 570-8501, Japan
{nakamura.yuzuru,ishii.yasunori,maruyama.yuuki}@jp.panasonic.com
[2] Chubu University, 1200 Matsumoto-cho, Kasugai City, Aichi 487-8501, Japan
takayoshi@isc.chubu.ac.jp

A Pseudocode of Method

We show the pseudocode of the procedure for applying OCDC and OCDCDL from the target domain to the source domain in Algorithm A.1. While we show the procedure in the direction from the target domain to the source domain, a similar procedure is used from the source domain to the target domain, with the labels S and T swapped.

B Additional Results

When pasting an object in image A into image B, the overlap between the objects in image A and image B is considered. For consideration, the overlap ratio between the pasted object and the pasting object is calculated. If the overlap is exceeded more than threshold γ, the pasting is not performed. Table B.1 shows the performance comparison result. When γ is made smaller, the allowance for overlapping objects become stricter, and the number of objects that cannot be pasted increases. There is no big difference overall even if γ is changed. Therefore, we calculate the average of mAP for each target sample and the difference between the average of mAP and the mAP for each γ setting in Table B.2. The greater the difference between each γ in the positive direction, the more beneficial γ is for improving accuracy in that target samples. For example, when target samples label is full, the average of mAP is 75.7 %, and differences from the average of mAP are 0.2 point at $\gamma = 0.1$, 0.4 point at $\gamma = 0.2$, -0.1 point at $\gamma = 0.5$, -0.6 point at $\gamma = 0.75$, respectively. When the average difference points calculated for all target samples labels at each γ are added, the total is 1.3 points at $\gamma = 0.25$, indicating a higher performance. Thus, we set the optimal γ to 0.25 in the experiments.

* indicates equal contribution.
† indicates the corresponding author.

Algorithm A.1: The procedure of our proposed method

Input: Images of source and target domain I_S, I_T, bounding boxes of source and target domain B_S, B_T, domain identification labels $\mathcal{D}_S$, $\mathcal{D}_T$ in a batch.

Output: Pasted images of source and target domain $\hat{I_S}$, $\hat{I_T}$, added bounding boxes of source and target domain $\hat{B_S}$, $\hat{B_T}$, replaced domain identification labels $\hat{\mathcal{D}_S}$, $\hat{\mathcal{D}_T}$ in a batch.

1 **Difinition:** (x, y, w, h): upper left position, width, and height of b, γ: overlap threshold, (W, H): width and height of image.
2 **forall** bounding box of target domain image $b_T \in B_T$ **do**
3   // selecting a bounding box matched size criteria
4   **if** $16 < w_{b_T} < W_{I_S}$ *and* $16 < h_{b_T} < H_{I_S}$ **then**
5   // setting $w_{\hat{b_T}}, h_{\hat{b_T}}$, and left upper position $x_{\hat{b_T}}, y_{\hat{b_T}}$
6     $s_{rand} \sim \mathcal{U}(0.7, 1.3)$
7     $w_{\hat{b_T}} \leftarrow w_{b_T} * s_{rand}$
8     $h_{\hat{b_T}} \leftarrow h_{b_T} * s_{rand}$
9     $x_{\hat{b_T}} \sim \mathcal{U}(0, W_{I_S} - w_{\hat{b_T}})$
10     $y_{\hat{b_T}} \sim \mathcal{U}(0, H_{I_S} - h_{\hat{b_T}})$
11   **forall** bounding box of source domain image $b_S \in B_S$ **do**
12   // calculating intersection area between $\hat{b_T}$ and b_S
13       $\mathcal{R}_{inter} \leftarrow intersection(\hat{b_T}, b_S)$
14   // calculating area of b_S
15       $\mathcal{R}_{Sarea} \leftarrow area(b_S)$
16       $\mathcal{R}_{overlap} \leftarrow \mathcal{R}_{inter}\ /\ \mathcal{R}_{Sarea}$
17   **if** $\mathcal{R}_{overlap} > \gamma$ **then**
18   go to $w_{\hat{b_T}}, h_{\hat{b_T}}, x_{\hat{b_T}}, y_{\hat{b_T}}$ setting (line 6)
19   // cropping region $\mathcal{A}_T$ from I_T
20     $\mathcal{A}_T \leftarrow crop(I_T, b_T)$
21   // resizing region $\mathcal{A}_T$ to $w_{\hat{b_T}}, h_{\hat{b_T}}$
22     $\hat{\mathcal{A}_T} \leftarrow resize(\mathcal{A}_T, w_{\hat{b_T}}, h_{\hat{b_T}})$
23   // pasting $\hat{\mathcal{A}_T}$ on $\hat{b_T}$ region of I_S
24     $\hat{I_S} \leftarrow paste(I_S, \hat{\mathcal{A}_T}, \hat{b_T})$
25   // adding $\hat{b_T}$ to B_S
26     $\hat{B_S} \leftarrow add(B_S, \hat{b_T})$
27   // switching domain identification labels corresponding to $\hat{b_T}$
28     $\hat{\mathcal{D}_S} \leftarrow switch(\mathcal{D}_S, \hat{b_T}\ /\ 16)$

Table B.1. Results on the overlap threshold γ

Target Samples	γ	Person	Bicycle	Car	mAP
Full	0.1	**77.8**	62.8	**87.1**	75.9
	0.25	**77.8**	**63.5**	86.9	**76.1**
	0.5	77.4	62.6	86.8	75.6
	0.75	76.8	61.5	87.0	75.1
1/2	0.1	78.2	**64.1**	**87.4**	**76.6**
	0.25	78.3	62.6	87.2	76.1
	0.5	78.3	62.4	87.2	76.0
	0.75	**78.4**	63.1	87.2	76.3
1/4	0.1	76.7	**61.4**	86.8	75.0
	0.25	76.9	59.9	86.9	74.5
	0.5	77.4	61.0	86.9	**75.1**
	0.75	**77.6**	59.4	**87.1**	74.7
1/8	0.1	75.0	59.5	**85.8**	73.4
	0.25	**75.4**	**60.9**	85.7	**74.0**
	0.5	74.8	58.9	85.5	73.1
	0.75	75.1	58.8	85.6	73.2
1/16	0.1	72.3	56.2	84.2	70.9
	0.25	72.2	**57.9**	84.5	**71.5**
	0.5	72.3	54.3	**84.8**	70.5
	0.75	**72.4**	55.4	84.5	70.8
1/32	0.1	70.4	53.7	82.9	69.0
	0.25	71.1	53.8	82.0	69.3
	0.5	70.2	54.6	83.2	69.3
	0.75	**71.3**	**55.8**	**83.5**	**70.2**
1/64	0.1	68.0	**51.9**	81.6	67.2
	0.25	68.5	51.6	82.3	**67.5**
	0.5	**68.7**	49.7	**82.4**	66.9
	0.75	68.0	50.4	82.0	66.8

Table B.2. Results on the average of mAP and the difference from the average of mAP

Target Samples	the average of mAP	γ	the difference from the average of mAP
Full	75.7	0.1	+0.2
		0.25	+**0.4**
		0.5	-0.1
		0.75	-0.6
1/2	76.2	0.1	+**0.3**
		0.25	-0.2
		0.5	-0.2
		0.75	±0.0
1/4	74.8	0.1	+0.2
		0.25	-0.3
		0.5	+**0.3**
		0.75	-0.1
1/8	73.4	0.1	±0.0
		0.25	+**0.6**
		0.5	-0.3
		0.75	-0.2
1/16	70.9	0.1	±0.0
		0.25	+**0.6**
		0.5	-0.4
		0.75	-0.1
1/32	69.4	0.1	-0.4
		0.25	-0.2
		0.5	-0.1
		0.75	+**0.7**
1/64	67.1	0.1	+0.1
		0.25	+**0.4**
		0.5	-0.2
		0.75	-0.3