Array 11 (2021) 100071

Contents lists available at ScienceDirect

Array

journal homepage: www.elsevier.com/journals/array/2590-0056/open-access-journal

Generative and self-supervised domain adaptation for one-stage object detection

Kazuma Fujii, Kazuhiko Kawamoto *

Chiba University, 1-33, Yayoicho, Inage Ward, Chiba-shi, Chiba, 263-8522, Japan

ARTICLE INFO

2010MSC:
00-01
99-00

Keywords:
Domain adaptation
Object detection
Unsupervised learning

ABSTRACT

Unsupervised cross-domain object detection has recently attracted considerable attention because of its ability to significantly reduce annotation costs. For two-stage detectors, several improvements have been made in feature-level adaptations. However, this approach is not suitable for one-stage detectors that do not have access to instance-level features. Although other approaches are often used for one-stage detectors, their performance is insufficient compared to domain adaptation methods for two-stage detectors. In this study, we propose a generative and self-supervised domain adaptation method for one-stage detectors. The proposed method is composed of an adversarial generative method and a self-supervision-based method. We tested our method on three evaluation datasets, and an improvement in the mean average precision was achieved using this method. We also confirmed the complementary effects of an adversarial generative method and a self-supervision-based method.

1. Introduction

Computer vision has attracted attention because of its applications in automated driving, video surveillance, anomaly detection, etc. In addition, the advent of deep learning has led to significant developments in computer vision. One of the typical tasks in this field is object detection.

In object detection, objects in an input image are classified, and then, localized using bounding boxes. Recent studies in this area have achieved remarkable results based on advancements in deep neural networks. Object detectors can be categorized as two-stage [1–3] or one-stage [4–7]. One-stage detectors are superior in terms of their inference speed.

Deep-learning-based object detectors are typically trained from a dataset that has many real-world images with instance-label annotations, such as Pascal visual object classes (VOC) [8]. However, the performance often decreases significantly when the training and test data have different distributions. One possible solution is to collect labeled data for a new domain, although this is a time-consuming approach. Another solution is domain adaptation. The goal of domain adaptation is to adapt a model from a label-rich domain (source domain) to a label-scarce domain (target domain). In particular, unsupervised domain adaptation assumes that there are no labels available in the target domain.

Unsupervised domain adaptation methods focused on visual tasks can be divided into four categories: discrepancy-based, adversarial discriminative, adversarial generative, and self-supervision-based methods [9]. Discrepancy-based methods are designed to reduce the difference between the source and target domain distributions [10,11]. Adversarial discriminative methods are designed to align features using the adversarial learning of feature extractors and domain classifiers [12,13]. Adversarial generative methods use target-like images with original source annotations obtained from image-to-image translations [14–16]. Self-supervision-based methods incorporate a self-supervised learning task in the target domain [17–19].

For two-stage detectors, domain adaptation approaches are mainly based on adversarial discriminative methods. They are often designed to align features at several levels, and have been improved in various manners. For example, image- and instance-level alignments have been used in Ref. [20]. However, it is difficult to assume such an approach in one-stage detectors, predicting bounding boxes and object classes concurrently.

For one-stage detectors, the authors of [16] proposed a domain

* Corresponding author.
E-mail addresses: fujisan8@chiba-u.jp (K. Fujii), kawa@faculty.chiba-u.jp (K. Kawamoto).

https://doi.org/10.1016/j.array.2021.100071
Received 14 February 2021; Received in revised form 26 April 2021; Accepted 26 May 2021
Available online 7 June 2021

transfer (DT) based on adversarial generative methods in a weakly supervised cross-domain setting. With DT, images with instance-level annotations are transferred from the source domain to the target domain. Under an unsupervised setting, the authors of [19] proposed weak self-training (WST) based on self-supervision-based methods. WST enables the training of unlabeled images, by reducing the negative effects of inaccurate pseudo-labels. However, the performances of these domain adaptation methods for one-stage detectors are insufficient compared with those for two-stage detectors.

Among the four categories of domain adaptation, adversarial generative and self-supervision-based methods can be easily applied to one-stage detectors. Moreover, these two methods have different advantages. An adversarial generative method can access accurate source labels; meanwhile, a self-supervision-based method can use the original target images. To take advantage of both methods, we propose an unsupervised domain adaptation method that combines an adversarial generative method with a self-supervision-based method. For each method, we use DT and WST, which have been shown to be effective for one-stage detectors [16,19]. We show that the two components complement each other, thereby improving the detection performance.

In summary, our main contributions are as follows.

- We propose an unsupervised domain adaptation approach for one-stage detectors. Our method consists of an adversarial generative method and a self-supervision-based method.
- We show that an adversarial generative method and a self-supervision-based method complement each other.
- The proposed method achieves an improvement in the mean average precision on three benchmark datasets.

2. Related work

In this section, we review the literature on object detection and domain adaptation.

2.1. Object detection

The development of deep convolutional neural networks (CNN) has improved the performance of object detection. Two-stage detectors (such as R-CNN [1], Fast R-CNN [2], and Faster R-CNN [3]) extract region proposals, and then, classify them. One of the advantages of the two-stage detectors is that the classifier can be customized to suit a specific task [21]. In contrast, one-stage detectors, such as You Only Look Once (YOLO) [4] and Single Shot MultiBox Detector (SSD) [5], achieve significant improvements in the inference speed using a single-stage network. Furthermore, recent studies [6,7] have improved both the accuracy and inference speed.

In this study, we tested our method on an SSD, which is a representative one-stage detector. SSD has a simple architecture, and is well balanced in terms of inference speed and performance. Furthermore, because SSD has been used in related studies [16,19], we can make a fair comparison.

2.2. Domain adaptation

The goal of domain adaptation is to adapt the information learned from the source domain for use in the target domain. The authors of [9] divided unsupervised domain adaptation methods for visual tasks into four categories: discrepancy-based methods, adversarial discriminative methods, adversarial generative methods, and self-supervision-based methods.

Discrepancy-based methods [10,11] manage to reduce the discrepancy between the feature distributions of the source and target domains. Adversarial discriminative methods [12,13] employ adversarial learning to align features. Discrepancy-based methods and adversarial discriminative methods are also called feature-level adaptations, because they aim to obtain domain-invariant features. Although these two categories often perform well, they require architecture-specific design.

Adversarial generative methods [14–16] are known as pixel-level adaptations. They are based on generative adversarial nets [22], and produce target-like training data from the source images. This category can be applied regardless of the architecture type, as it only changes the training data. However, the performance is highly dependent on the quality of image generation.

Self-supervision-based methods [17–19] employ self-supervised tasks, such as reconstruction, image rotation prediction, and self-training. This category brings the source and target domains closer by adding auxiliary tasks to the target images. However, the performance is limited compared to the other categories.

2.3. Domain adaptive object detection

Recently, numerous studies have been proposed to address the problem of domain shifts in object detection. For two-stage detectors, adversarial discriminative methods are often used [20,23,24], and have shown very good performance in recent studies [25–27]. In many cases, they are designed to align features at several levels, including the instance-level. One-stage detectors do not have access to instance-level features, as they predict bounding boxes and object classes simultaneously. Therefore, adversarial discriminative methods are not sufficient for one-stage detectors.

However, adversarial generative methods and self-supervision-based methods are suitable for both two-stage and one-stage detectors. The authors of [16] proposed the DT method, which is based on adversarial generative methods. DT transfers images with instance-level annotations from the source domain to the target domain using CycleGAN [28], and trains a detector on domain-adapted images. The authors of [19] proposed the use of WST and adversarial background score regularization (BSR), which are based on self-supervision-based methods. With WST, reliable detections on unlabeled images were chosen, and pseudo-instance-level annotations were generated. BSR reduces the domain shift by extracting discriminative features for target backgrounds. In this study, we combine an adversarial generative method and a self-supervision-based method for a one-stage detector.

3. Method

In this study, we propose a generative and self-supervised domain adaptation method for a one-stage detector. Fig. 1 shows an overview of our method. Our method is based on SSD [5], and combines two methods: DT [16], which is an adversarial generative method, and WST [19], which is a self-supervision-based method.

In this section, we first formulate the problem, and then, explain the effect of combining the two methods and the details of the proposed method.

3.1. Problem setting

Let $\mathbf{x}$ and y denote an input image and a label, respectively. We assume that the source data $\{(\mathbf{x}_s^i, y_s^i)\}_{i=1}^{N_s}$ are drawn from the source domain

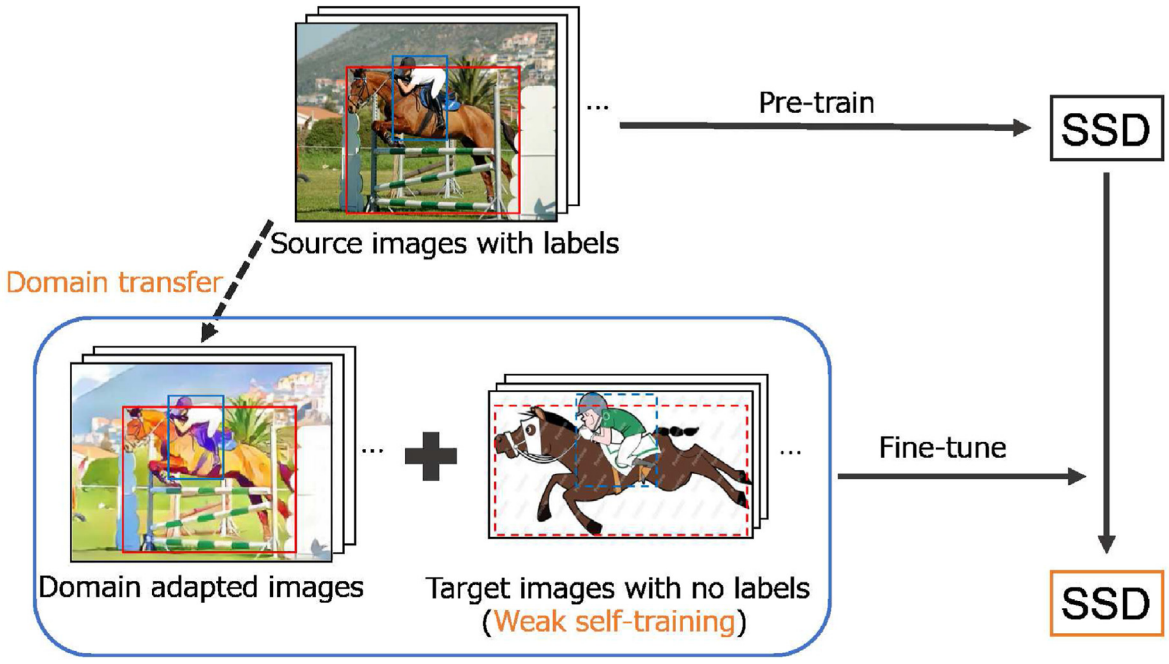


Fig. 1. Overview of the proposed approach.

X_s, and the target data $\{(\mathbf{x}_t^i, y_t^i)\}_{i=1}^{N_t}$ are drawn from the target domain X_t, where N_s. N_t is the number of source and target samples, respectively. We denote the distribution of domain X as $P(X)$ and $P(X_s) \neq P(X_t)$. Therefore, the source and target data have different distributions, as shown in Fig. 2(a). We do not have access to the target labels $\{y_t^i\}_{i=1}^{N_t}$ because we address unsupervised domain adaptation.

3.2. Generative and self-supervised domain adaptation

In an ideal scenario where the target labels are available, supervised learning of the target data is possible, as shown in Fig. 2(b). We approach the unsupervised domain adaptation for one-stage detectors by bringing the learning setting closer to the ideal case.

We propose generative and self-supervised domain adaptation composed of an adversarial generative method and a self-supervision-based method. In the adversarial generative method (Fig. 2(c)), the source images were converted to target-like images. Although the distribution of the transferred images does not perfectly match that of the target images, the transferred images with source labels enable supervised learning. In the self-supervision-based method (Fig. 2(d)), self-supervised tasks were employed on the target images. Although supervised learning cannot be applied, this method enables the training of the original target images. These two contrasting methods are expected to complement each other. The proposed method (Fig. 2(e)) is close to the ideal case (Fig. 2(b)) in terms of supervised learning near the target domain and using the original target images for training.

In this study, we applied DT [16] to the adversarial generative method and WST [19] for the self-supervision-based method. DT uses CycleGAN [28] to transform the source images into target-like images. WST enables self-supervised learning by generating pseudo-labels on the target images. The pseudo-labels are assigned to reliable detections, considering the detection results of neighboring regions. Then, WST trains detectors using pseudo-labels, while reducing the effect of false negatives using weak negative mining.

3.3. Training method

The proposed method can be divided into three steps.

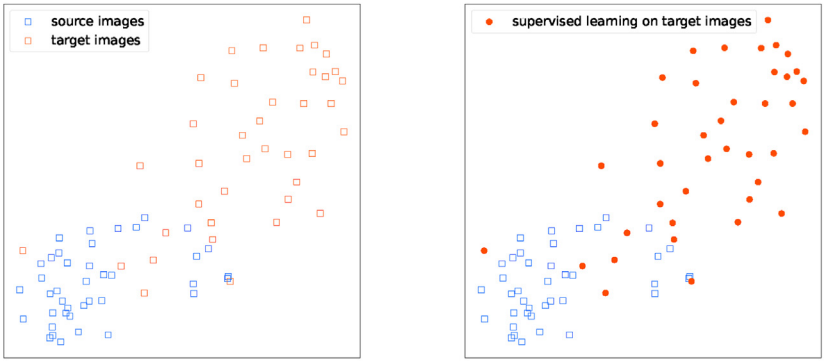
Pre-training SSD: We pre-train the model using the source data $\{(\mathbf{x}_s^i, y_s^i)\}_{i=1}^{N_s}$. The training loss of SSD [5] can be written as follows.

$$L(b, c, l, g) = \frac{1}{N}(L_{conf}(b, c) + \alpha L_{loc}(b, l, g)). \quad (1)$$

where b is a matched default box, c is the confidence of multiple classes, l is a predicted box, g is a ground-truth box, N is the number of matched default boxes, α is the weight, L_{conf} is the confidence loss, and L_{loc} is the localization loss.

DT: We trained CycleGAN on the source images $\{\mathbf{x}_s^i\}_{i=1}^{N_s}$ and target images $\{\mathbf{x}_t^i\}_{i=1}^{N_t}$. Using the trained CycleGAN, we converted the source images $\{\mathbf{x}_s^i\}_{i=1}^{N_s}$, which are used in pre-training, and obtained domain-adapted images $\{\mathbf{x}_{s\to t}^i\}_{i=1}^{N_s}$ that accompany labels $\{y_s^i\}_{i=1}^{N_s}$. Examples of domain-adapted images are shown in Fig. 3. The real-world images are transferred to each target domain.

Fine-tuning: The model was fine-tuned using domain-adapted data $\{(\mathbf{x}_{s\to t}^i, y_s^i)\}_{i=1}^{N_s}$ and target images $\{\mathbf{x}_t^i\}_{i=1}^{N_t}$. During fine-tuning, the training batch consists of half domain-adapted data and half target images. We applied the loss function of the SSD, as shown in Eq. (1), for the domain-adapted data and the WST for the target images.



(a) Source and target samples (b) Ideal case

(c) Adversarial generative method (d) Self-supervision-based method

(e) Proposed method

Fig. 2. Visualizations of input images and learning strategies. Blue: source images; red: target images; purple: transferred images from the source domain to the target domain. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4. Experiments

4.1. Datasets and evaluation

In our experiments, we used the Pascal VOC dataset [8] as the source domain with the Clipart1k, Watercolor2k, or Comic2k datasets [16] as the target domains. Examples of target images are shown in Fig. 4.

Pascal VOC is a real-world image dataset, providing instance-level annotations. VOC2007-trainval and VOC2012-trainval datasets have 16,551 images with 20 object classes. Clipart1k is a graphical image dataset, and has the same classes as Pascal VOC. It provides 500 images for the training set and another 500 images for the test set. Watercolor2k and Comic2k are unrealistic datasets, and have six classes in Pascal VOC. Each dataset provided 1000 images for the training set and 1000 images for the test set. We trained a model without using the labels of the target images, because we tackled unsupervised domain adaptation.

For all experiments, we evaluated different methods on the target test data using average precision (AP) and mean average precision (mAP) as indicators.

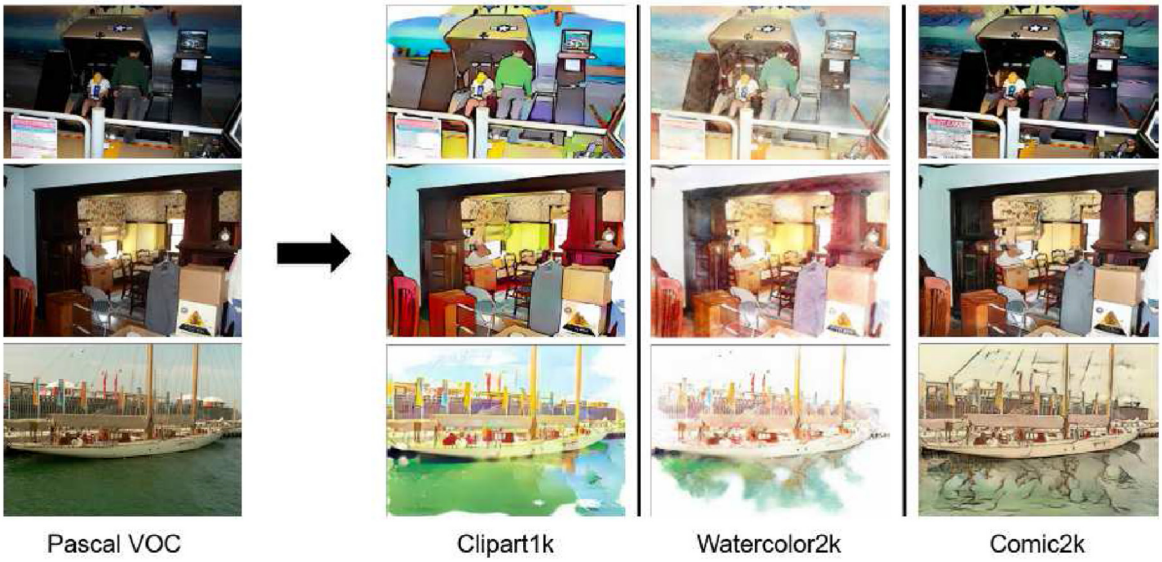


Fig. 3. Examples of domain-adapted images. The images in Pascal VOC [8] are transferred to the Clipart1k, Watercolor2k, and Comic2k [16] domains.

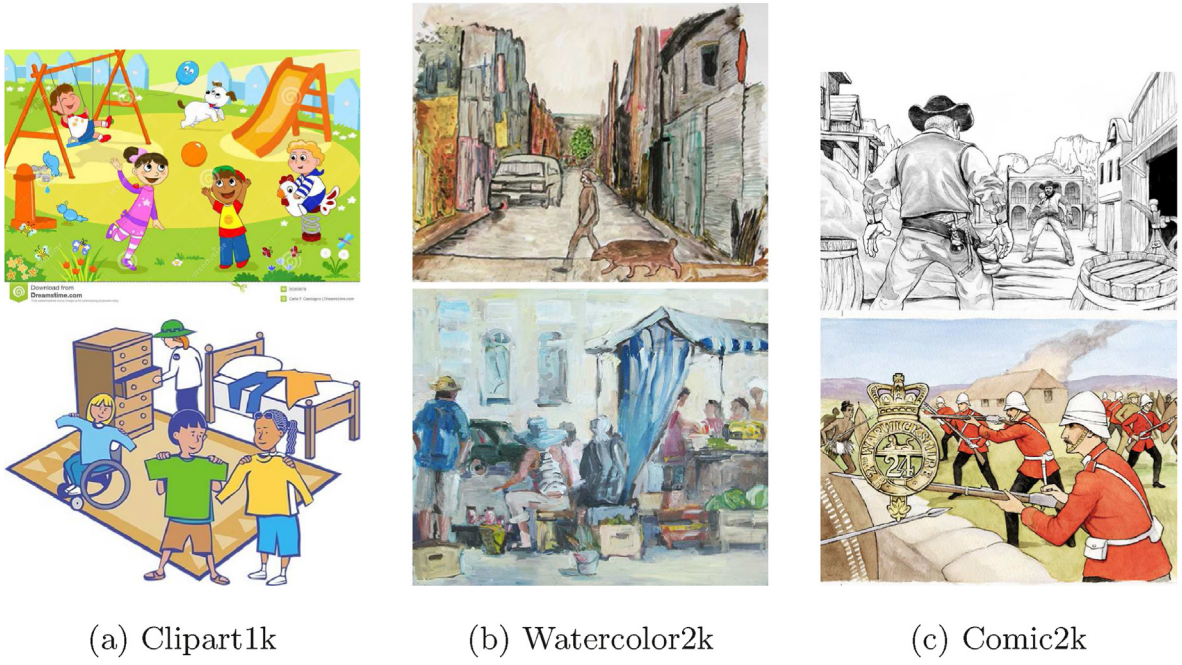


Fig. 4. Example images in target domains.

4.2. Experimental setup

Details of training: For all experiments, SSD300 [5] was used as a base detector. Following the original study [5], we trained the model using the source data for 120,000 iterations. We used this model as the baseline for the experiments.

We also trained CycleGAN using source images and each target image for 20 epochs, following the original study [28]. Using the trained CycleGAN, we obtained domain-adapted images from the source images.

We then fine-tuned all layers of the model using our method. The baseline was applied as the initial weight. Each batch was composed of 32 images—16 from the domain-adapted images and 16 from the target images. The model was trained for 3000 iterations with a learning rate of 1.0×10^{-5}.

Comparison: We compared our method with the baseline [5], DT [16] and WST + BSR [19] approaches. To quantify the relative contributions of the adversarial generative method and self-supervision-based method with our method, we trained and tested the model using DT or

Table 1
Results on adaptation from Pascal VOC to Clipart1k. AP (%) was evaluated for the target images. Column "C" indicates which categories the method belongs to, where "G" and "S" denote the adversarial generative method and self-supervision-based method, respectively.

Method	C	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
Base [5]		23.3	56.6	17.9	17.3	14.5	39.4	33.3	7.2	43.4	11.5	28.6	11.1	26.4	48.1	35.6	27.3	2.7	22.0	26.0	23.5	25.8
DT [16]	G	23.3	60.1	24.9	**41.5**	**26.4**	53.0	44.0	4.1	45.3	**51.5**	39.5	11.6	**40.4**	62.2	61.1	37.1	**20.9**	**39.6**	38.4	36.0	38.0
WST + BSR [19]	S	28.0	**64.5**	23.9	19.0	21.9	**64.3**	43.5	**16.4**	42.2	25.9	30.5	7.9	25.5	**67.6**	54.5	36.4	10.3	31.2	**57.4**	43.5	35.7
DT	G	26.3	56.3	24.3	27.7	26.2	49.8	45.3	5.0	**49.6**	49.2	**41.1**	15.2	32.3	55.5	59.5	**39.3**	14.8	33.2	39.5	48.1	36.9
WST	S	24.2	55.6	18.2	20.4	18.4	41.9	38.7	5.6	45.5	18.0	32.5	7.1	29.0	53.5	45.0	26.4	3.9	25.6	28.6	30.7	28.5
DT + WST(proposed)	G + S	**28.2**	61.5	**25.1**	28.9	23.6	57.0	**46.7**	6.8	48.7	49.6	37.0	**16.6**	34.5	60.2	**63.3**	38.5	13.6	36.6	42.4	**48.9**	**38.4**

WST alone.

4.3. Results

Results on Clipart1k: A comparison of the performance on Clipart1k is presented in Table 1. Our method outperformed the other methods in terms of AP on the six classes, and improved the mAP by 12.6% from that of the baseline and by 0.4% from that of the existing methods. Applying DT or WST alone also outperformed the baseline. In particular, DT improved the mAP by 11.1% from that of the baseline.

Results on Watercolor2k: A comparison of the performance on Watercolor2k is presented in Table 2. Our method outperformed the other methods for AP in all three classes, improving the mAP by 4.1% over that of the baseline and by 0.5% over that of the existing methods. Applying DT or WST alone also outperformed the baseline. However, DT improved the mAP by only 0.2% from that of the baseline.

Results on Comic2k: A comparison of the performance on Comic2k is presented in Table 3. Our method outperformed the other methods for AP in three classes, improving the mAP by 10.7% over that of the baseline and by 2.4% over that of the existing methods. Applying DT or WST alone also outperformed the baseline. Specifically, DT improved mAP by 9.7% from that of the baseline.

4.4. Qualitative results

The qualitative results are shown in Fig. 5. We found that the proposed method detected more objects correctly compared to the baseline. Furthermore, objects detected by DT but not by WST were detected by the proposed method, and vice versa.

4.5. Discussion

Complementary effect: The experiments show that our method is effective for unsupervised domain adaptation. Based on the improved accuracy compared to using DT or WST alone, the adversarial generative method and self-supervision-based method are considered to complement each other.

Measuring domain distances: To quantitatively evaluate the performance of each target dataset, we computed the FID [29] between the source and target images and between the domain-adapted and target images, as shown in Table 4. The FID measures the difference between two distributions in the high-dimensional feature space of the Inception-v3 model [30], and indicates the similarity between the two groups. The smaller the FID between the source and target images, the better the performance of the baseline, as shown in Tables 1–3.

DT performance: We found that DT is ineffective for Watercolor2k as compared to Clipart1k and Comic2k. Although the FID between the domain-adapted and target images did not differ significantly among the datasets, the distance difference on Watercolor2k was the smallest, which led to DT's poor performance. In contrast, DT performs better for Clipart1k and Comic2k, where the distance differences are larger. Thus, the effectiveness of the adversarial generative method depends on the target dataset.

WST performance: WST showed improvements for all three datasets. The results suggest that the self-supervision-based method is robust to variations in the target domain.

Table 2
AP for the adaptation from Pascal VOC to Watercolor2k(%).

Method	C	bike	bird	car	cat	dog	person	mAP
Base [5]		81.2	45.6	39.8	29.5	27.1	57.7	46.8
DT [16]	G	82.8	47.0	40.2	**34.6**	**35.3**	62.5	50.4
WST + BSR [19]	S	75.6	45.8	**49.3**	34.1	30.3	64.1	49.9
DT	G	80.7	45.1	41.6	29.0	27.2	58.3	47.0
WST	S	76.0	**47.1**	42.7	30.3	29.6	65.7	48.6
DT + WST(proposed)	G + S	**88.6**	**47.1**	44.1	30.6	28.1	**67.1**	**50.9**

Table 3
AP for the adaptation from Pascal VOC to Comic2k(%).

Method	C	bike	bird	car	cat	dog	person	mAP
Base [5]		40.5	10.1	22.1	10.2	11.8	34.5	21.5
DT [16]	G	43.6	13.6	30.2	**16.0**	**26.9**	48.3	29.8
WST + BSR [19]	S	50.6	13.6	31.0	7.5	16.4	41.4	26.8
DT	G	49.4	**16.3**	32.7	14.7	22.6	51.2	31.2
WST	S	45.6	9.9	27.3	9.8	12.6	50.4	25.9
DT + WST(proposed)	G + S	**52.7**	15.0	**35.8**	13.0	20.8	**56.1**	**32.2**

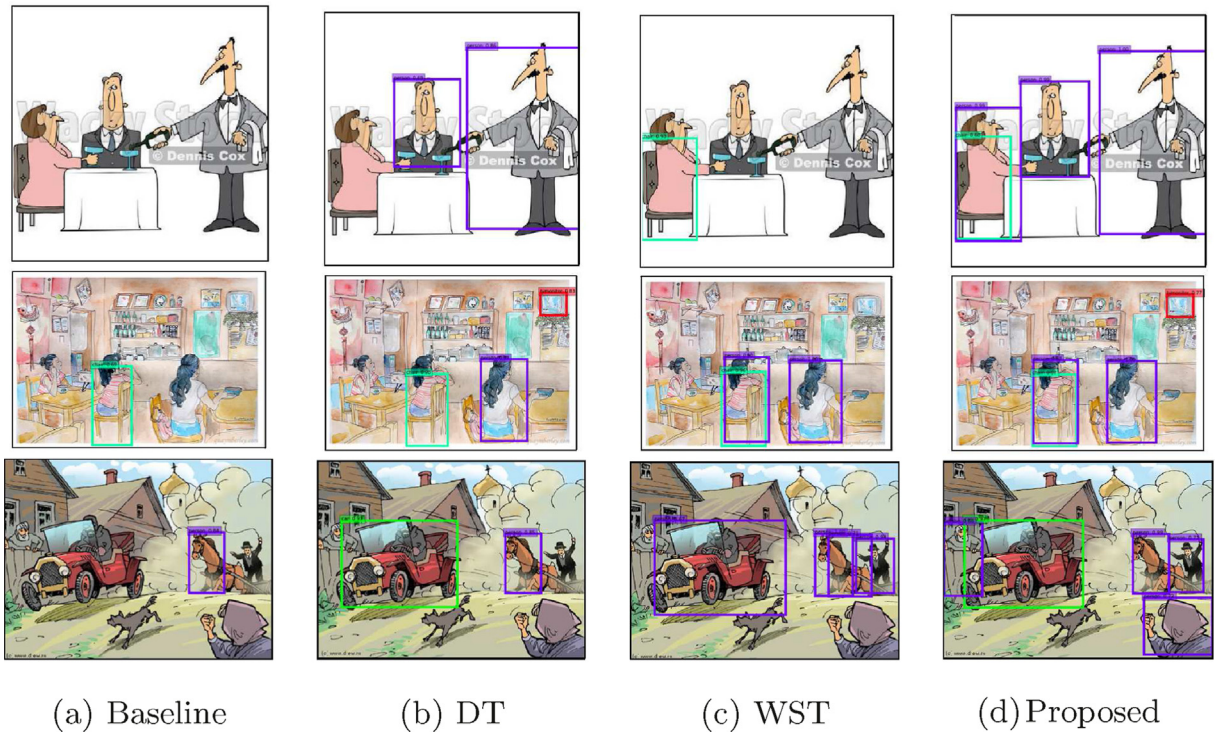


Fig. 5. Examples of detection results on the target domain, from top to bottom: Clipart1k, Watercolor2k, and Comic2k [16].

Table 4
Fréchet inception distance (FID) [29] between the source and target images and between the domain-adapted and target images. The source images represent the Pascal VOC dataset, the target images represent the respective target dataset, and the domain-adapted images represent the Pascal VOC dataset transferred to the respective target domain via DT. In addition, we computed the difference between the two FIDs.

Target dataset	source images ↔ target images	domain-adapted images ↔ target images	distance difference
Clipart1k	145.2	78.6	66.6
Watercolor2k	124.3	75.9	48.4
Comic2k	160.8	76.3	84.5

5. Conclusion

In this study, we the addressed unsupervised domain adaptation for one-stage detectors. To take advantage of both the adversarial generative method and self-supervision-based method, we introduced a generative and self-supervised domain adaptation method. Specifically, we proposed a learning strategy for SSDs by applying DT and WST.

Our experiments show that the proposed method improves the domain adaptation performance on three benchmark datasets. Furthermore, we confirmed that the two components of our method complement each other.

Credit author statement

Kazuma Fujii: Conceptualization, Methodology, Software, Writing - Original Draft; Kazuhiko Kawamoto: Supervision, Writing - Review & Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Sumitomo Construction Machinery CO., LTD.

References

[1] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR; 2014. p. 580–7.
[2] Girshick R. Fast r-cnn. In: ICCV; 2015. p. 1440–8.
[3] Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. In: NeurIPS; 2015. p. 91–9.
[4] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: CVPR; 2016. p. 779–88.
[5] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC. Ssd: single shot multibox detector. In: ECCV. Springer; 2016. p. 21–37.
[6] J. Redmon, A. Farhadi, Yolov3: an incremental improvement, arXiv preprint arXiv: 1804.02767.
[7] Zhao Q, Sheng T, Wang Y, Tang Z, Chen Y, Cai L, Ling H, M2det. A single-shot object detector based on multi-level feature pyramid network. In: AAAI, vol. 33; 2019. p. 9259–66.
[8] Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. IJCV 2010;88(2):303–38.
[9] S. Zhao, X. Yue, S. Zhang, B. Li, H. Zhao, B. Wu, R. Krishna, J. E. Gonzalez, A. L. Sangiovanni-Vincentelli, S. A. Seshia, et al., A review of single-source deep unsupervised visual domain adaptation, TNNLS.
[10] Long M, Cao Y, Wang J, Jordan M. Learning transferable features with deep adaptation networks. In: ICML, PMLR; 2015. p. 97–105.
[11] W. Zellinger, T. Grubinger, E. Lughofer, T. Natschläger, S. Saminger-Platz, Central moment discrepancy (cmd) for domain-invariant representation learning, arXiv preprint arXiv:1702.08811.
[12] Tzeng E, Hoffman J, Saenko K, Darrell T. Adversarial discriminative domain adaptation. In: CVPR; 2017. p. 7167–76.
[13] Long M, Cao Z, Wang J, Jordan MI. Conditional adversarial domain adaptation. In: NeurIPS; 2018. p. 1640–50.
[14] Bousmalis K, Silberman N, Dohan D, Erhan D, Krishnan D. Unsupervised pixel-level domain adaptation with generative adversarial networks. In: CVPR; 2017. p. 3722–31.
[15] Hoffman J, Tzeng E, Park T, Zhu J-Y, Isola P, Saenko K, Efros A, Darrell T. Cycada: cycle-consistent adversarial domain adaptation. In: ICML, PMLR; 2018. p. 1989–98.
[16] Inoue N, Furuta R, Yamasaki T, Aizawa K. Cross-domain weakly-supervised object detection through progressive domain adaptation. In: CVPR; 2018. p. 5001–9.
[17] Ghifary M, Kleijn WB, Zhang M, Balduzzi D, Li W. Deep reconstruction-classification networks for unsupervised domain adaptation. In: ECCV. Springer; 2016. p. 597–613.
[18] Xu J, Xiao L, López AM. Self-supervised domain adaptation for computer vision tasks. IEEE Access 2019;7:156694–706.
[19] Kim S, Choi J, Kim T, Kim C. Self-training and adversarial background regularization for unsupervised domain adaptive one-stage object detection. In: ICCV; 2019. p. 6092–101.
[20] Chen Y, Li W, Sakaridis C, Dai D, Van Gool L. Domain adaptive faster r-cnn for object detection in the wild. In: CVPR; 2018. p. 3339–48.
[21] Pérez-Hernández F, Tabik S, Lamas A, Olmos R, Fujita H, Herrera F. Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: application in video surveillance. Knowl Base Syst 2020; 194:105590.
[22] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: NeurIPS; 2014. p. 2672–80.
[23] Xie R, Yu F, Wang J, Wang Y, Zhang L. Multi-level domain adaptive learning for cross-domain detection. In: ICCV workshops; 2019.
[24] M. Fu, Z. Xie, W. Li, L. Duan, Deeply aligned adaptation for cross-domain object detection, arXiv preprint arXiv:2004.02093.
[25] Xu C-D, Zhao X-R, Jin X, Wei X-S. Exploring categorical regularization for domain adaptive object detection. In: CVPR; 2020.
[26] Chen C, Zheng Z, Ding X, Huang Y, Dou Q. Harmonizing transferability and discriminability for adapting object detectors. In: CVPR; 2020.
[27] Zheng Y, Huang D, Liu S, Wang Y. Cross-domain object detection through coarse-to-fine feature adaptation. In: CVPR; 2020.
[28] Zhu J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV; 2017. p. 2223–32.
[29] Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: NeurIPS; 2017. p. 6626–37.
[30] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: CVPR; 2016. p. 2818–26.