Supplementary: Object Recognition with Continual Open Set Domain Adaptation for Home Robot

Ikki Kishida[1], Hong Chen[1], Masaki Baba[1], Jiren Jin[*1], Ayako Amma[2], and Hideki Nakayama[1]

[1] *Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan*
[2] *Toyota Research Institute - Advanced Development, Tokyo, Japan*
{kishida,chen,baba,jin,nakayama}@nlab.ci.i.u-tokyo.ac.jp[1], ayako.amma@tri-ad.global[2]

Figure 1: The appearance of HSR.

Body Height	1005 mm — 1350 mm
Maximum Velocity	0.8 km/h
4*Head Sensors	RGB-D sensor $\times$ 1
	Stereo camera $\times$ 1
	Wide-angle camera $\times$ 1
	Microphone array $\times$ 1

Table 1: The specifications of HSR.

1. Toyota Human Support Robot

The appearance and specifications of Toyota Human Support Robot (HSR) [32] are shown in Figure 1 and Table 1.

[*] now at Google.

2. Implementation Details

2.1. Continual learning

We conduct experiments on the COSDA-HR dataset in continual learning setting by 8 different methods: Naive, LwF [19], EWC [14], DGR [39], Rehearsal, iCaRL [32], BiC [50] and Cumulative. There are 10 episodes in the experiment. In each episode, there are new 16 classes. We resize all images into 224×224 except DGR which is resized into 128×128. The memory size is 2000 images for those methods which need to do rehearsal (i.e., Rehearsal, iCaRL and BiC). We train the model for 4 epochs in each episode and SGD optimizer is used with a fixed learning rate of 0.01 for all the methods except Naive and EWC. This is because their accuracies drop very fast when training long times. Thus, we decay the learning rate based on the number of the current episode. It can be formalized as follow:

$$\text{lr} = \text{lr}_{\text{init}} \times \alpha^n,$$

$$\text{epoch}_n = \lfloor \text{epoch}_{\text{base}} \times \alpha^n \rfloor.$$

n denotes the number of the current episode. lr and lr_{init} denotes the learning rate at the episode n and the initial learning rate, respectively. epoch_n and $\text{epoch}_{\text{base}}$ denotes the number of the epoch to run at the episode n and the base epoch number, which is set as 4. α controls the decayness of the learning rate and the number of the iterations for each episode, and we set α as 0.7 in our experiments.

2.2. Open-set recognition

We conduct experiments of the open-set recognition task in the COSDA-HR dataset with three methods: softmax thresholding, Openmax [3] and CROSR [47]. SGD optimizer is used with 0.9 momentum, and the initial learning rate and the L2 weight decay value are 0.001 and 0.0001, respectively. For training CROSR, we extend the ResNet-18 by inserting additional convolution and transpose convolution layers after conv3, conv4 and conv5. The additional convolution layers reduce the channel size to 32 and expand back to the original channel size, and we use this reduced intermediate feature to calculate Openmax. The transpose convolution layers are used to reconstruct the image. The classification loss (i.e., cross-entropy loss) and the L2 loss (i.e., image reconstruction loss) are used to update the model parameters.

Softmax thresholding recognizes the image as the unknown class when the highest prediction value from softmax is under 0.8.

Openmax. We use $\alpha = 1$, tail size $= 7$ to train Openmax. The model recognizes the image as the unknown class when the highest prediction value from Openmax is under 0.8.

CROSR. We use $\alpha = 2$, tail size $= 6$ to train CROSR. The model recognizes the image as the unknown class when the highest prediction value from Openmax is under 0.8.

2.3. Domain adaptation

We experiment with five recently proposed unsupervised domain adaptation methods for the domain adaptation task in the COSDA-HR dataset, including AFN [46], BSP [6], DAAN [49], DTA [18] and MRAN [54]. The hyper-parameters used for these methods are as follows:

AFN. We use SGD optimizer with the initial learning rate of 0.001 and 0.01 for the L2 weight regularization. The batch size is 32.

BSP. We use SGD optimizer with the inverse learning rate scheduler and the initial learning rate is 0.003. The L2 weight decay value is 0.0005 and the batch size is 36.

DAAN. We use SGD optimizer with the initial learning rate of 0.001 and 0.9 as the momentum. The batch size is 32 and the L2 weight decay value is 0.0005.

DTA. The learning rate 0.001 and the L2 weight decay 0.0005 are used in SGD optimizer with 0.9 momentum. The batch size is 128. The source consistency weight and the target consistency weight are 1 and 2, respectively. In addition, the class balance weight is 0.01 and the target VAT loss weight is 0.2.

MRAN. We use SGD optimizer with the initial learning rate 0.01 and momentum 0.9. The batch size is 32 and the L2 weight decay value is 0.0005. In addition, we use the different initial learning rate for the classifier, which is 0.1.

2.4. Open set domain adaptation

We take two types of approaches to the COSDA-HR dataset in open-set domain adaptation setting: open-set recognition + domain adaptation and open-set domain adaptation.

The first approach is to combine methods in open-set recognition and domain adaptation. As the method of domain adaptation, we use MRAN [54] that achieves the best score for the COSDA-HR dataset with domain adaptation setting. As the method of open-set recognition, we use softmax-thresholding and Openmax [3]. The open-set recognition is the post-processing method, thus the combining procedure is to 1. train the model with domain adaptation and 2. apply the method of open-set recognition to the trained model. We use the same training procedures for open-set recognition and domain adaptation as described in section 2.2 and 2.3.

The second approach is to use methods of open set domain adaptation and we apply two standard methods to the COSDA-HR dataset in open-set domain adaptation setting: OPDABP [36] and UDA [48]. The implementation and learning setting for both methods are based on the original codes.

OPDABP.We use SGD optimizer with 0.9 momentum and the initial learning rate is 0.001. The L2 weight decay value is 0.005 and the batch size is 128.

UDA. UDA consists of feature extractor (i.e., ImageNet-1k [33] pre-trained ResNet-18 [10]), two adversarial network and the classifier. We use SGD optimizer with 0.9 momentum, the L2 weight decay value is 0.0005 and the batch size is 36. The initial learning rate is 0.0001 for the feature extractor and 0.001 for the rest of the parts. The learning rate scheduler is based on the equation below:

$$\mathrm{lr} = \mathrm{lr}_{\mathrm{init}} * (1.0 + \gamma * \min(1.0, \frac{t}{t_{\max}}))^{-p}$$

$\mathrm{lr}_{\mathrm{init}}$ is the initial learning rate and t is the number of current iteration. We set $t_{\max}$, γ and p as 10000, 10 and 0.75, respectively.

2.5. Continual open set domain adaptation

We take two types of approaches to the COSDA-HR dataset in continual open set domain adaptation setting: continual learning + open-set recognition + domain adaptation and continual learning + open-set domain adaptation. In this setting, there are 10 episodes and each episode has new 16 classes.

As the first approach, we conduct the experiment with Bic [50] + MRAN [54] + Softmax and Bic + MRAN + Openmax [3]. Since methods of open-set recognition are post-processing methods, the combining procedure is to 1. train the model with Bic + MRAN and 2. apply the open-set recognition method to the trained model. To combine MRAN with Bic, SGD optimizer is used with momentum 0.9. The batch size is 32, the L2 weight decay value is 0.0005 and the memory size is 2000. The initial learning rate for the classifier and the rest of the parts are 0.001 and 0.0001, respectively. The learning rate is fixed through all episodes. For combining the method of open-set recognition, we use the same hyper-parameters and the training procedure described in section 2.2.

method	accuracy (%)
Naive	31.59
LwF [19]	29.61
EWC [14]	33.80
DGR [39]	8.40
Rehearsal	89.27
ICaRL [32]	82.70
BiC [50]	**90.18**
Cumulative (upper bound)	95.05

Table 2: The results of the COSDA-HR dataset in continual learning setting on $\text{Source}_{\text{test}}$.

As the second approach, we combine UDA with continual learning methods. SGD optimizer is used with momentum 0.9. The batch size is 32 and the L2 weight decay value is 0.0005. The initial learning rate for the feature extractor (i.e., ResNet-18) and the rest of the parts are 0.0001 and 0.001, respectively. The memory size is set as 2000 for Bic and Rehearsal. The learning rate is fixed through all episodes when combining Naive, Rehearsal, Bic and Cumulative. For combining LwF, we use the same learning rate scheduling described in section 2.1.

3. Experiment

3.1. Continual learning

In this section, we perform a preliminary experiment to evaluate several continual learning methods on the closed source images in the COSDA-HR dataset. Specifically, we train the system on $\text{Source}_{\text{train}}$ and evaluate on $\text{Source}_{\text{test}}$. 160 classes are randomly split into 10 subsets to form a 10-step sequence of continual learning. Lwf [19], EWC [14] and DGR [39] are approaches without memory. Rehearsal, ICaRL [32], BiC [50] are approaches with memory. We set the size of the memory as 2000 in the above three methods. Naive is the regular training strategy without memory. Cumulative is the regular training strategy with unlimited memory and it is the upper-bound performance for this task.

As can be seen in Table 2, methods without memory achieve weak performances and methods with memory gives a strong performance that is closed to the performance of Cumulative which is the upper-bound performance. We find that DGR uses GAN in its method and the generator fails to learn to generate images, thus GAN is difficult to apply to the COSDA-HR dataset with current techniques. It represents that the continual learning setting without memory is challenging in the COSDA-HR dataset.

4. Analysis

4.1. Confusion matrix

Figure 2 shows the confusion matrix of Bic + UDA baseline model in the task of continual open set domain adaptation on the COSDA-HR dataset. As can be seen in Figure 2, many examples are misclassified as the unknown class. We need a more sophisticated method that detects the unknown class without degrading the performance on the closed set.

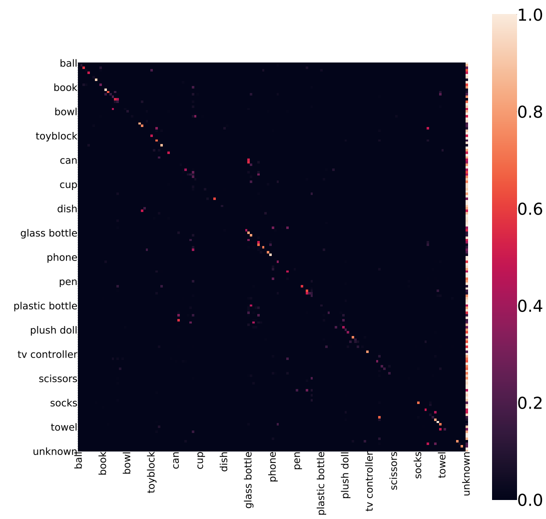

Figure 2: The confusion matrix of BiC + UDA baseline model in the COSDA-HR dataset. There are 16 super-category and the unknown class. Each super-category has 10 different classes, thus the number of total classes is 161. The confusion matrix is 161×161 and the super-category is denoted in each axis. Ten different classes belonging to the super-category are arranged between the neighbored super categories on the axis (e.g., [ball1, ball2, ... ball10, book1, book2,...]). Each row in the confusion matrix is normalized by the sum of its column.

References

[1] Rahaf Aljundi, Min Lin, Baptiste Goujaud, and Yoshua Bengio. Gradient based sample selection for online continual learning. *NIPS*, 2019.

[2] Mahsa Baktashmotlagh, Masoud Faraki, Tom Drummond, and Mathieu Salzmann. Learning factorized representations for open-set domain adaptation. *ICLR*, 2019.

[3] Abhijit Bendale and Terrance Boult. Towards open set deep networks. *CVPR*, 2016.

[4] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. Unsupervised pixel-level domain adaptation with generative adversarial networks. *CVPR*, 2017.

[5] Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-gem. *ICLR*, 2019.

[6] Xinyang Chen, Sinan Wang, Mingsheng Long, and Jianmin Wang. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. *ICML*, 2019.

[7] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.

[8] Zongyuan Ge, Sergey Demyanov, Zetao Chen, and Rahil Garnavi. Generative openmax for multi-class open set classification. *BMVC*, 2017.

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *NIPS*, 2014.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CVPR*, 2016.

[11] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. *CVPR*, 2017.

[12] Jing Jiang and ChengXiang Zhai. Instance weighting for domain adaptation in nlp. *ACL*, 2007.

[13] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. *CVPR*, 2019.

[14] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

[15] Simon Kornblith, Jon Shlens, and Quoc V. Le. Do better imagenet models transfer better? *CVPR*, 2019.

[16] Jogendra Nath Kundu, Naveen Venkat, Ambareesh Revanur, Rahul M. V., and R. Venkatesh Babu. Towards inheritable models for open-set domain adaptation. *CVPR*, 2020.

[17] K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view rgb-d object dataset. *ICRA*, 2011.

[18] Seungmin Lee, Dongwan Kim, Namil Kim, and Seong-Gyun Jeong. Drop to adapt: Learning discriminative features for unsupervised domain adaptation. *ICCV*, 2019.

[19] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017.

[20] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection (best student paper award). *ICCV*, 2017.

[21] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. *CVPR*, 2014.

[22] Hong Liu, Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Qiang Yang. Separate to adapt: Open set domain adaptation via progressive separation. *CVPR*, 2019.

[23] Mohammad Reza Loghmani, Barbara Caputo, and Markus Vincze. Recognizing objects in-the-wild: Where do we stand? *ICRA*, 2018.

[24] Vincenzo Lomonaco and Davide Maltoni. Core50: a new dataset and benchmark for continuous object recognition. *Proceedings of the 1st Annual Conference on Robot Learning*, 78, 2017.

[25] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I Jordan. Learning transferable features with deep adaptation networks. *arXiv*, 2015.

[26] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *NIPS*, 2018.

[27] David Lopez-Paz and Marc Aurelio Ranzato. Gradient episodic memory for continual learning. *NIPS*, 2017.

[28] Yingwei Pan, Ting Yao, Yehao Li, Chong-Wah Ngo, and Tao Mei. Exploring category-agnostic clusters for open-set domain adaptation. *CVPR*, 2020.

[29] Pau Panareda Busto and Juergen Gall. Open set domain adaptation. *ICCV*, 2017.

[30] G. Pasquale, C. Ciliberto, L. Rosasco, and L. Natale. Object identification from few examples by improving the invariance of a deep convolutional neural network. *IROS*, 2016.

[31] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. *AAAI*, 2018.

[32] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. *CVPR*, 2017.

[33] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 2015.

[34] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. *ECCV*, 2010.

[35] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. *CVPR*, 2018.

[36] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. *ECCV*, 2018.

[37] Qi She, Fan Feng, Xinyue Hao, Qihan Yang, Chuanlin Lan, Vincenzo Lomonaco, Xuesong Shi, Zhengwei Wang, Yao Guo, Yimin Zhang, Fei Qiao, and Rosa H. M. Chan. Openloris-object: A dataset and benchmark towards lifelong object recognition. *arXiv*, 2019.

[38] Jian Shen, Yanru Qu, Weinan Zhang, and Yong Yu. Wasserstein distance guided representation learning for domain adaptation. *AAAI*, 2018.

[39] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. *NIPS*, 2017.

[40] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. *ECCV*, 2016.

[41] Dan Teng and Sakyasingha Dasgupta. Continual learning via online leverage score sampling. *arXiv*, 2019.

[42] TOYOTA MOTOR CORPORATION. *Toyota shifts home helper robot R&D into high gear with new developer community and upgraded prototype*. `https://global.toyota/en/detail/8709541`. Accessed 7 Nov, 2020.

[43] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. *CVPR*, 2017.

[44] Yifan Wu, Ezra Winston, Divyansh Kaushik, and Zachary Lipton. Domain adaptation with asymmetrically-relaxed distribution alignment. *arXiv*, 2019.

[45] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. *CVPR*, 2017.

[46] Ruijia Xu, Guanbin Li, Jihan Yang, and Liang Lin. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. *ICCV*, 2019.

[47] Ryota Yoshihashi, Wen Shao, Rei Kawakami, Shaodi You, Makoto Iida, and Takeshi Naemura. Classification-reconstruction learning for open-set recognition. *CVPR*, 2019.

[48] Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan. Universal domain adaptation. *CVPR*, 2019.

[49] Chaohui Yu, Jindong Wang, Yiqiang Chen, and Meiyu Huang. Transfer learning with dynamic adversarial adaptation network. *arXiv*, 2019.

[50] Lijuan Wang Yuancheng Ye Zicheng Liu Yandong Guo Yun Fu Yue Wu, Yinpeng Chen. Large scale incremental learning. *arXiv*, 2019.

[51] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *BMVC*, 2016.

[52] Werner Zellinger, Thomas Grubinger, Edwin Lughofer, Thomas Natschläger, and Susanne Saminger-Platz. Central moment discrepancy (cmd) for domain-invariant representation learning. *arXiv*, 2017.

[53] Han Zhao, Remi Tachet des Combes, Kun Zhang, and Geoffrey J Gordon. On learning invariant representation for domain adaptation. *arXiv*, 2019.

[54] Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Jingwu Chen, Zhiping Shi, Wenjuan Wu, and Qing He. Multi-representation adaptation network for cross-domain image classification. *Neural Networks*, 119:214–221, 2019.