
Appendix for “Zero-Round Active Learning”

Anonymous Author(s)

Affiliation

Address

email

A Details of Datasets Used in Section 4

MNIST. [17] MNIST dataset contains a training set of 60,000 examples and a test set of 10,000 examples. The images are grayscale handwritten digits with size 28×28 . We resize the images to 32×32 in setting $SVHN \Rightarrow MNSIT$.

USPS. [1] USPS dataset is a digit dataset scanned from envelopes. It contains a total of 9,298 16×16 grayscale pixels. We resize them to 28×28 in both $MNIST \Rightarrow USPS$ and $USPS \Rightarrow MNIST$ setting.

SVHN. [25] SVHN is a real-world color house-number dataset containing 73,257 images for training and 26,032 images for testing. We use the version where all digits have been resized to 32×32 pixels.

CIFAR-10. [16] The CIFAR-10 is an image recognition dataset containing 60,000 32×32 3-channel images in 10 classes.

STL-10. [7] The STL-10 dataset consists of 13,000 color images of size 96×96 in 10 classes. We resize them to 32×32 in the experiments.

VISDA2017. [27] VISDA2017 dataset is designed for unsupervised domain adaptation challenge which contains more than 280K images across 12 object categories with large domain gap. The source domain are synthetic 2D images rendering of 3D models which the angles and lighting conditions are different. The target domain are photo-realistic or real-images. In the experiment, we resize all the images to 256×256 and crop at the center obtaining images with size 224×224 . An example of synthetic-real image pair is shown in Figure 1.

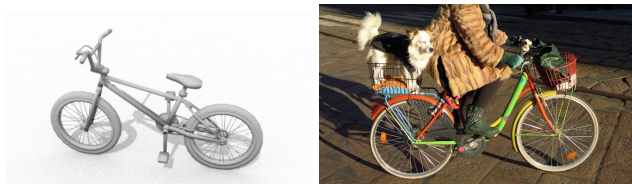


Figure 1: Example images in VISDA2017. The left is an image of source domain (synthetic) while the right is an image of target domain (real).

B Details of Models and Baseline Algorithms in Section 4

SVM We use Linear Support Vector Classification (SVC) implemented by scikit-learn [26] with L2 penalty and regularization parameter $C = 0.1$. Others remain as default.

24 **Logistic Regression** We use Logistic Regression implemented by scikit-learn [26]. We set the
 25 maximum number of iterations to be 1000.

26 **Small CNN** The small CNN model we used has two convolutional layers and two max pooling
 27 layers and three fully-connected layers. We use Adam optimizer with learning rate 10^{-3} , $\varepsilon = 10^{-7}$,
 28 batch size 32 for training the small CNN model.

29 **DeepSets Model** A DeepSets model can be represented as $f_{DS}(S) = \rho(\sum_{x \in S} \phi(x))$ where both
 30 ρ and ϕ are neural networks. In our experiments, both ρ and ϕ contain 3 linear layer with ELU
 31 activation, and we set the number of neurons to be 256 in each hidden layer, the dimension of set
 32 features which is the output of ϕ network to be 256. For training DeepSets models, we use Adam
 33 optimizer with learning rate 10^{-5} , batch size 32, $\beta_1 = 0.9$, and $\beta_2 = 0.99$.

34 **Baseline AL Techniques** We use BADGE, FASS, and GLISTER implemented by DISTIL¹. Specif-
 35 ically, we set batch size to be 32 for all of the three strategies, and learning rate to be 0.001 for
 36 glister.

37 C Other Implementation Details

38 **Domain Adaptation** We test our method with three state-of-the-art domain adaptation frameworks
 39 in this paper: CyCADA [13], UDA [37], AFN [43].

40 For CyCADA², we follow their official implementation where a source classifier is firstly trained
 41 using Adam optimizer with learning rate 10^{-4} , batch size 128, $\beta_1 = 0.9$, and $\beta_2 = 0.99$. Then,
 42 weights of this source classifier are used as the initial weights of target classifier to perform domain
 43 adaptation. Same optimizer are used for training target classifier. We set the k in Line 10 of Algorithm
 44 2 to be 10.

45 For UDA³, we use SGD optimizer with initial learning rate 0.1. We later decay the learning rate to
 46 0.001 after 10 epochs. And we set k to be 5.

47 For AFN⁴, we use SGD optimizer with learning rate 0.001 and weight decay 5×10^{-4} for training
 48 feature extractor, and SGD optimizer with learning rate 0.001, momentum 0.9 and weight decay
 49 5×10^{-4} for training class predictor. We set k to be 5.

50 When combined with all of the above three DA frameworks, the same Adam optimizer with learning
 51 rate 10^{-6} , $\beta_1 = 0.9$, and $\beta_2 = 0.99$ is used for DeepSets Loss back propagation.

52 **Data Selection** We apply stochastic greedy optimization [24] to solve Equation (4), and we set
 53 $\epsilon = 10^{-3}$.

54 References

- 55 [1] E Alpaydin and C Kaynak. Optical recognition of handwritten digits data set. *UCI Machine*
 56 *Learning Repository*, 1998.
- 57 [2] Jordan T Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agar-
 58 wal. Deep batch active learning by diverse, uncertain gradient lower bounds. *arXiv preprint*
 59 *arXiv:1906.03671*, 2019.
- 60 [3] Jordan T Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agar-
 61 wal. Deep batch active learning by diverse, uncertain gradient lower bounds. *arXiv preprint*
 62 *arXiv:1906.03671*, 2019.
- 63 [4] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jen-
 64 nifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*,
 65 79(1):151–175, 2010.

¹<https://github.com/decile-team/distil>

²https://github.com/jhoffman/cycada_release

³https://github.com/yueatsprograms/uda_release

⁴<https://github.com/jihanyang/AFN>

- [5] Colin Campbell, Nello Cristianini, Alex Smola, et al. Query learning with large margin classifiers. In *ICML*, volume 20, page 0, 2000.
- [6] Rita Chattopadhyay, Wei Fan, Ian Davidson, Sethuraman Panchanathan, and Jieping Ye. Joint transfer and batch-mode active learning. In *International conference on machine learning*, pages 253–261. PMLR, 2013.
- [7] Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 215–223. JMLR Workshop and Conference Proceedings, 2011.
- [8] Shai Fine, Ran Gilad-Bachrach, and Eli Shamir. Query by committee, linear separation and random walks. *Theoretical Computer Science*, 284(1):25–51, 2002.
- [9] Yoav Freund, H Sebastian Seung, Eli Shamir, and Naftali Tishby. Selective sampling using the query by committee algorithm. *Machine learning*, 28(2):133–168, 1997.
- [10] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015.
- [11] 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.
- [12] Thore Graepel and Ralf Herbrich. The kernel gibbs sampler. In *NIPS*, pages 514–520. Citeseer, 2000.
- [13] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pages 1989–1998. PMLR, 2018.
- [14] Krishnateja Killamsetty, Durga Sivasubramanian, Ganesh Ramakrishnan, and Rishabh Iyer. Glisten: Generalization based data subset selection for efficient and robust learning. *arXiv preprint arXiv:2012.10630*, 2020.
- [15] Andreas Kirsch, Joost Van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. *arXiv preprint arXiv:1906.08158*, 2019.
- [16] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.
- [17] Yann LeCun. The mnist database of handwritten digits. <http://yann.lecun.com/exdb/mnist/>, 1998.
- [18] Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10285–10295, 2019.
- [19] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [20] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.
- [21] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *arXiv preprint arXiv:1705.10667*, 2017.
- [22] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pages 2208–2217. PMLR, 2017.
- [23] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

- [24] Baharan Mirzasoleiman, Ashwinkumar Badanidiyuru, Amin Karbasi, Jan Vondrák, and Andreas Krause. Lazier than lazy greedy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.
- [25] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011.
- [26] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.
- [27] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge, 2017.
- [28] Piyush Rai, Avishek Saha, Hal Daumé III, and Suresh Venkatasubramanian. Domain adaptation meets active learning. In *Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*, pages 27–32, 2010.
- [29] Avishek Saha, Piyush Rai, Hal Daumé, Suresh Venkatasubramanian, and Scott L DuVall. Active supervised domain adaptation. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 97–112. Springer, 2011.
- [30] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Adversarial dropout regularization. *arXiv preprint arXiv:1711.01575*, 2017.
- [31] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3723–3732, 2018.
- [32] Greg Schohn and David Cohn. Less is more: Active learning with support vector machines. In *ICML*, volume 2, page 6. Citeseer, 2000.
- [33] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.
- [34] H Sebastian Seung, Manfred Oppel, and Haim Sompolsky. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 287–294, 1992.
- [35] Jian Shen, Yanru Qu, Weinan Zhang, and Yong Yu. Wasserstein distance guided representation learning for domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- [36] Jong-Chyi Su, Yi-Hsuan Tsai, Kihyuk Sohn, Buyu Liu, Subhransu Maji, and Manmohan Chandraker. Active adversarial domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 739–748, 2020.
- [37] Yu Sun, Eric Tzeng, Trevor Darrell, and Alexei A. Efros. Unsupervised domain adaptation through self-supervision, 2019.
- [38] Simon Tong and Daphne Koller. Support vector machine active learning with applications to text classification. *Journal of machine learning research*, 2(Nov):45–66, 2001.
- [39] Eric Tzeng, Judy Hoffman, Trevor Darrell, and Kate Saenko. Simultaneous deep transfer across domains and tasks. In *Proceedings of the IEEE international conference on computer vision*, pages 4068–4076, 2015.
- [40] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*, 2014.
- [41] Tianhao Wang, Si Chen, and Ruoxi Jia. One-round active learning. *arXiv preprint arXiv:2104.11843*, 2021.

- 159 [42] Kai Wei, Rishabh Iyer, and Jeff Bilmes. Submodularity in data subset selection and active
160 learning. In *International Conference on Machine Learning*, pages 1954–1963. PMLR, 2015.
- 161 [43] Ruijia Xu, Guanbin Li, Jihan Yang, and Liang Lin. Larger norm more transferable: An
162 adaptive feature norm approach for unsupervised domain adaptation. In *The IEEE International
163 Conference on Computer Vision (ICCV)*, October 2019.
- 164 [44] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov,
165 and Alexander Smola. Deep sets. *arXiv preprint arXiv:1703.06114*, 2017.
- 166 [45] Han Zou, Yuxun Zhou, Jianfei Yang, Huihan Liu, Hari Prasanna Das, and Costas J Spanos.
167 Consensus adversarial domain adaptation. In *Proceedings of the AAAI conference on artificial
168 intelligence*, volume 33, pages 5997–6004, 2019.