Unsupervised Domain Adaptation via Structurally Regularized Deep Clustering

Hui Tang, Ke Chen, and Kui Jia*
South China University of Technology
381 Wushan Road, Tianhe District, Guangzhou, Guangdong, China
eehuitang@mail.scut.edu.cn, {chenk,kuijia}@scut.edu.cn

A. Other implementation details

Other implementation details are as follows: 1) the momentum is set to 0.9; 2) the weight decay is set to 0.0001; 3) the batch size is set to 64; 4) the number of training epochs is set to 200; 5) for each trial, we follow [13] and use the best-performing clustering model as the test model; 6) data augmentations of random crop and horizontal flip are applied during training; 7) the number of the task-specific FC layers of the base network is set to 2 (i.e. $2048 \rightarrow 512 \rightarrow K$), where the first FC layer is the so-called bottleneck layer [3, 13, 27]; 8) we perform discriminative clustering in the bottleneck feature space as additional regularization; 9) we implement our experiments in PyTorch.

Our proposed SRDC simultaneously learns parameters of the feature embedding function $\boldsymbol{\theta}$, the classifier $\boldsymbol{\vartheta}$, and the learnable cluster centers $\{\boldsymbol{\mu}_k\}_{k=1}^K$ by minimizing the structurally regularized deep clustering objective (11). Note that we re-initialize $\{\boldsymbol{\mu}_k\}_{k=1}^K$ at the start of each training epoch based on the current cluster assignments of $\{\boldsymbol{z}_i^t\}_{i=1}^{n_t}$ together with labeled source $\{\boldsymbol{z}_j^s\}_{j=1}^{n_s}$; the introduced auxiliary distributions $q_{i,k}^t = \widetilde{q}_{i,k}^t = \mathrm{I}[k = \hat{y}_i^t]$ for $i \in \{1, 2, \dots, n_t\}$ and $k \in \{1, 2, \dots, K\}$ at the first training epoch, where $\hat{y}_i^t$ is the assigned class label by standard K-means clustering on the embedded target features $\{\boldsymbol{z}_i^t\}_{i=1}^{n_t}$; the weights $\{w_j^s\}_{j=1}^{n_s}$ are set to 1 at the first training epoch. For the K-means, the target cluster centers are initialized as the class centroids of the source data. Training algorithm of SRDC is given in Algorithm 1.

Algorithm 1 Training algorithm for SRDC, E denotes the training epoch, I denotes the training iteration, B_t and B_s denote the mini-batches.

Input: unlabeled target samples $\mathcal{T} = \{\boldsymbol{x}_i^t\}_{i=1}^{n_t}$; labeled source samples $\mathcal{S} = \{(\boldsymbol{x}_j^s, y_j^s)\}_{j=1}^{n_s}$
Output: $\boldsymbol{\theta}, \boldsymbol{\vartheta}, \{\boldsymbol{\mu}_k\}_{k=1}^K$
Initialize: $\boldsymbol{\theta}, \boldsymbol{\vartheta}, \{\boldsymbol{\mu}_k\}_{k=1}^K$, $q_{i,k}^t = \widetilde{q}_{i,k}^t = \mathrm{I}[k = \hat{y}_i^t]$ for $i \in \{1, 2, \dots, n_t\}$ and $k \in \{1, 2, \dots, K\}$, $w_j^s = 1$ for $j \in \{1, 2, \dots, n_s\}$, $E = 1$
while not converge **do**
    for $I \leftarrow 1, MAX\_ITER$ **do**
        Sample B_t and B_s from $\mathcal{T}$ and $\mathcal{S}$
        if E != 1 **then**
            Compute q_{ik}^t and $\widetilde{q}_{ik}^t$ by using (2)
        end if
        Update $\boldsymbol{\theta}, \boldsymbol{\vartheta}, \{\boldsymbol{\mu}_k\}_{k=1}^K$ by minimizing (11) on B_t and B_s
    end for
    Compute $\{\boldsymbol{c}_k^t\}_{k=1}^K$ by standard K-means clustering
    Compute $w_j^s = 1, j \in \{1, 2, \dots, n_s\}$ by using (12)
    Initialize: $\{\boldsymbol{\mu}_k\}_{k=1}^K$
    E = E + 1
end while

B. More comparisons

B.1. Comparisons on Office-31

Comparisons with existing methods on Office-31 [19] using ResNet-50 [7] as the base network are shown in Table A, where results of existing methods are quoted from their respective papers or the works of [2, 11, 13, 16]. We can see that SRDC outperforms all compared methods on almost all transfer tasks, verifying the effectiveness of SRDC.

*Corresponding author.

B.2. Comparisons on ImageCLEF-DA

Comparisons with existing methods on ImageCLEF-DA [1] using ResNet-50 [7] as the base network are reported in Table B, where results of existing methods are quoted from their respective papers or the work of [13, 16]. To compare our proposed SRDC with the state-of-the-art method CAN [8] on ImageCLEF-DA, we report results of CAN obtained by running the official code (i.e. available at the website of https://github.com/kgl-prml/Contrastive-Adaptation-Network-for-Unsupervised-Domain-Adaptation). We can see that SRDC exceeds all compared methods including CAN on all transfer tasks by a large margin, confirming the efficacy of SRDC.

Method	A → W	D → W	W → D	A → D	D → A	W → A	Avg
Source Model [7]	77.8±0.2	96.9±0.1	99.3±0.1	82.1±0.2	64.5±0.2	66.1±0.2	81.1
RTN [14]	84.5±0.2	96.8±0.1	99.4±0.1	77.5±0.3	66.2±0.2	64.8±0.3	81.6
DAN [12]	81.3±0.3	97.2±0.0	99.8±0.0	83.1±0.2	66.3±0.0	66.3±0.1	82.3
DANN [6]	81.7±0.2	98.0±0.2	99.8±0.0	83.9±0.7	66.4±0.2	66.0±0.3	82.6
ADDA [24]	86.2±0.5	96.2±0.3	98.4±0.3	77.8±0.3	69.5±0.4	68.9±0.5	82.9
JAN-A [15]	86.0±0.4	96.7±0.3	99.7±0.1	85.1±0.4	69.2±0.4	70.7±0.5	84.6
MADA [16]	90.0±0.1	97.4±0.1	99.6±0.1	87.8±0.2	70.3±0.3	66.4±0.3	85.2
VADA [22]	86.5±0.5	98.2±0.4	99.7±0.2	86.7±0.4	70.1±0.4	70.5±0.4	85.4
SimNet [17]	88.6±0.5	98.2±0.2	99.7±0.2	85.3±0.3	73.4±0.8	71.8±0.6	86.2
GTA [21]	89.5±0.5	97.9±0.3	99.8±0.4	87.7±0.5	72.8±0.3	71.4±0.4	86.5
MSTN [26]	91.3	98.9	100.0	90.4	72.7	65.6	86.5
MCD [20]	88.6±0.2	98.5±0.1	**100.0**±0.0	92.2±0.2	69.5±0.1	69.7±0.3	86.5
SAFN+ENT [27]	90.1±0.8	98.6±0.2	99.8±0.0	90.7±0.5	73.0±0.2	70.2±0.3	87.1
DAAA [9]	86.8±0.2	99.3±0.1	100.0±0.0	88.8±0.4	74.3±0.2	73.9±0.2	87.2
iCAN [28]	92.5	98.8	**100.0**	90.1	72.1	69.9	87.2
rRevGrad+CAT [4]	94.4±0.1	98.0±0.2	100.0±0.0	90.8±1.8	72.2±0.6	70.2±0.1	87.6
CDAN+E [13]	94.1±0.1	98.6±0.1	**100.0**±0.0	92.9±0.2	71.0±0.3	69.3±0.3	87.7
MSTN+DSBN [2]	92.7	99.0	100.0	92.2	71.7	74.4	88.3
TADA [25]	94.3±0.3	98.7±0.1	99.8±0.2	91.6±0.3	72.9±0.2	73.0±0.3	88.4
TAT [11]	92.5±0.3	**99.3**±0.1	**100.0**±0.0	93.2±0.2	73.1±0.3	72.1±0.3	88.4
SymNets [30]	90.8±0.1	98.8±0.3	**100.0**±0.0	93.9±0.5	74.6±0.6	72.5±0.5	88.4
BSP+CDAN [3]	93.3±0.2	98.2±0.2	**100.0**±0.0	93.0±0.2	73.6±0.3	72.6±0.3	88.5
MDD [29]	94.5±0.3	98.4±0.1	**100.0**±0.0	93.5±0.2	74.6±0.3	72.2±0.1	88.9
DADA [23]	92.3±0.1	99.2±0.1	**100.0**±0.0	93.9±0.2	74.4±0.1	74.2±0.1	89.0
CADA-P [10]	**97.0**±0.2	**99.3**±0.1	**100.0**±0.0	95.6±0.1	71.5±0.2	73.1±0.3	89.5
CAN [8]	94.5±0.3	99.1±0.2	99.8±0.2	95.0±0.3	**78.0**±0.3	77.0±0.3	90.6
SRDC	95.7±0.2	99.2±0.1	**100.0**±0.0	**95.8**±0.2	76.7±0.3	**77.1**±0.1	**90.8**

Table A. Results (%) on Office-31 (ResNet-50).

Methods	I → P	P → I	I → C	C → I	C → P	P → C	Avg
Source Model [7]	74.8±0.3	83.9±0.1	91.5±0.3	78.0±0.2	65.5±0.3	91.2±0.3	80.7
DAN [12]	74.5±0.4	82.2±0.2	92.8±0.2	86.3±0.4	69.2±0.4	89.8±0.4	82.5
RTN [14]	75.6±0.3	86.8±0.1	95.3±0.1	86.9±0.3	72.7±0.3	92.2±0.4	84.9
DANN [6]	75.0±0.6	86.0±0.3	96.2±0.4	87.0±0.5	74.3±0.5	91.5±0.6	85.0
MADA [16]	75.0±0.3	87.9±0.2	96.0±0.3	88.8±0.3	75.2±0.2	92.2±0.3	85.8
JAN [15]	76.8±0.4	88.0±0.2	94.7±0.2	89.5±0.3	74.2±0.3	91.7±0.3	85.8
rRevGrad+CAT [4]	77.2±0.2	91.0±0.3	95.5±0.3	91.3±0.3	75.3±0.6	93.6±0.5	87.3
iCAN [28]	79.5	89.7	94.7	89.9	78.5	92.0	87.4
CDAN+E [13]	77.7±0.3	90.7±0.2	97.7±0.3	91.3±0.3	74.2±0.2	94.3±0.3	87.7
CAN [8]	77.2±0.6	90.3±0.5	96.0±0.2	90.9±0.3	78.0±0.6	95.6±0.6	88.0
CADA-P [10]	78.0	90.5	96.7	92.0	77.2	95.5	88.3
TAT [11]	78.8±0.2	92.0±0.2	97.5±0.3	92.0±0.3	78.2±0.4	94.7±0.4	88.9
SAFN+ENT [27]	79.3±0.1	93.3±0.4	96.3±0.4	91.7±0.0	77.6±0.1	95.3±0.1	88.9
SymNets [30]	80.2±0.3	93.6±0.2	97.0±0.3	93.4±0.3	78.7±0.3	96.4±0.1	89.9
SRDC	**80.8**±0.3	**94.7**±0.2	**97.8**±0.2	**94.1**±0.2	**80.0**±0.3	**97.7**±0.1	**90.9**

Table B. Results (%) on ImageCLEF-DA (ResNet-50). Note that results of CAN are obtained by running the official code.

Methods	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
Source Model [7]	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
DAN [12]	43.6	57.0	67.9	45.8	56.5	60.4	44.0	43.6	67.7	63.1	51.5	74.3	56.3
DANN [6]	45.6	59.3	70.1	47.0	58.5	60.9	46.1	43.7	68.5	63.2	51.8	76.8	57.6
JAN [15]	45.9	61.2	68.9	50.4	59.7	61.0	45.8	43.4	70.3	63.9	52.4	76.8	58.3
SE [5]	48.8	61.8	72.8	54.1	63.2	65.1	50.6	49.2	72.3	66.1	55.9	78.7	61.5
DWT-MEC [18]	50.3	72.1	77.0	59.6	69.3	70.2	58.3	48.1	77.3	69.3	53.6	82.0	65.6
CDAN+E [13]	50.7	70.6	76.0	57.6	70.0	70.0	57.4	50.9	77.3	70.9	56.7	81.6	65.8
TAT [11]	51.6	69.5	75.4	59.4	69.5	68.6	59.5	50.5	76.8	70.9	56.6	81.6	65.8
BSP+CDAN [3]	52.0	68.6	76.1	58.0	70.3	70.2	58.6	50.2	77.6	72.2	59.3	81.9	66.3
SAFN [27]	52.0	71.7	76.3	64.2	69.9	71.9	63.7	51.4	77.1	70.9	57.1	81.5	67.3
TADA [25]	53.1	72.3	77.2	59.1	71.2	72.1	59.7	53.1	78.4	72.4	60.0	82.9	67.6
SymNets [30]	47.7	72.9	78.5	64.2	71.3	74.2	64.2	48.8	79.5	74.5	52.6	82.7	67.6
MDD [29]	54.9	73.7	77.8	60.0	71.4	71.8	61.2	53.6	78.1	72.5	60.2	82.3	68.1
CAN [8]	**58.5**	75.3	75.1	61.7	74.5	70.1	61.3	**54.6**	75.9	72.4	58.3	82.4	68.3
CADA-P [10]	56.9	**76.4**	80.7	61.3	75.2	75.2	63.2	54.5	80.7	73.9	**61.5**	84.1	70.2
SRDC	52.3	76.3	**81.0**	**69.5**	**76.2**	**78.0**	**68.7**	53.8	**81.7**	**76.3**	57.1	**85.0**	**71.3**

Table C. Results (%) on Office-Home (ResNet-50). Note that results of CAN are obtained by running the official code.

B.3. Comparisons on Office-Home

Comparisons with existing methods on Office-Home using ResNet-50 [7] as the base network are reported in Table C, where results of existing methods are quoted from their respective papers or the works of [13, 18]. To compare our proposed SRDC with the state-of-the-art method CAN [8] on Office-Home, we report results of CAN obtained by running the official code (i.e. available at the website of https://github.com/kgl-prml/Contrastive-Adaptation-Network-for-Unsupervised-Domain-Adaptation). We can observe that SRDC achieves much better results than all compared methods including CAN on almost all transfer tasks, affirming the usefulness of SRDC.

References

[1] The imageclef-da dataset is available at `http://imageclef.org/2014/adaptation`.

[2] W. Chang, T. You, S. Seo, S. Kwak, and B. Han. Domain-specific batch normalization for unsupervised domain adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7346–7354, June 2019.

[3] Xinyang Chen, Sinan Wang, Mingsheng Long, and Jianmin Wang. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1081–1090, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

[4] Z. Deng, Y. Luo, and J. Zhu. Cluster alignment with a teacher for unsupervised domain adaptation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9943–9952, Oct 2019.

[5] Geoff French, Michal Mackiewicz, and Mark Fisher. Self-ensembling for visual domain adaptation. In *International Conference on Learning Representations*, 2018.

[6] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.*, 17(1):2096–2030, Jan. 2016.

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016.

[8] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4888–4897, June 2019.

[9] Guoliang Kang, Liang Zheng, Yan Yan, and Yi Yang. Deep adversarial attention alignment for unsupervised domain adaptation: The benefit of target expectation maximization. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 420–436, Cham, 2018. Springer International Publishing.

[10] V. K. Kurmi, S. Kumar, and V. P. Namboodiri. Attending to discriminative certainty for domain adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 491–500, June 2019.

[11] Hong Liu, Mingsheng Long, Jianmin Wang, and Michael Jordan. Transferable adversarial training: A general approach to adapting deep classifiers. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4013–4022, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

[12] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I. Jordan. Learning transferable features with deep adaptation networks. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 97–105. JMLR.org, 2015.

[13] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan. Conditional adversarial domain adaptation. In *Proceedings of the 32Nd International Conference*

on Neural Information Processing Systems, NIPS'18, pages 1647–1657, USA, 2018. Curran Associates Inc.

[14] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I. Jordan. Unsupervised domain adaptation with residual transfer networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 136–144, USA, 2016. Curran Associates Inc.

[15] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I. Jordan. Deep transfer learning with joint adaptation networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pages 2208–2217. JMLR.org, 2017.

[16] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. In *Association for the Advancement of Artificial Intelligence (AAAI)*, pages 3934–3941, 2018.

[17] P. O. Pinheiro. Unsupervised domain adaptation with similarity learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8004–8013, June 2018.

[18] S. Roy, A. Siarohin, E. Sangineto, S. R. Bulò, N. Sebe, and E. Ricci. Unsupervised domain adaptation using feature-whitening and consensus loss. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9463–9472, June 2019.

[19] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *Proceedings of the 11th European Conference on Computer Vision: Part IV*, ECCV'10, pages 213–226, Berlin, Heidelberg, 2010. Springer-Verlag.

[20] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3723–3732, June 2018.

[21] S. Sankaranarayanan, Y. Balaji, C. D. Castillo, and R. Chellappa. Generate to adapt: Aligning domains using generative adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8503–8512, June 2018.

[22] Rui Shu, Hung Bui, Hirokazu Narui, and Stefano Ermon. A DIRT-t approach to unsupervised domain adaptation. In *International Conference on Learning Representations*, 2018.

[23] Hui Tang and Kui Jia. Discriminative adversarial domain adaptation. In *Association for the Advancement of Artificial Intelligence (AAAI)*, 2020.

[24] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell. Adversarial discriminative domain adaptation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2962–2971, July 2017.

[25] Ximei Wang, Liang Li, Weirui Ye, Mingsheng Long, and Jianmin Wang. Transferable attention for domain adaptation. In *Association for the Advancement of Artificial Intelligence (AAAI)*, 2019.

[26] Shaoan Xie, Zibin Zheng, Liang Chen, and Chuan Chen. Learning semantic representations for unsupervised domain adaptation. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5423–5432, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

[27] R. Xu, G. Li, J. Yang, and L. Lin. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1426–1435, Oct 2019.

[28] W. Zhang, W. Ouyang, W. Li, and D. Xu. Collaborative and adversarial network for unsupervised domain adaptation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3801–3809, June 2018.

[29] Yuchen Zhang, Tianle Liu, Mingsheng Long, and Michael Jordan. Bridging theory and algorithm for domain adaptation. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7404–7413, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

[30] Y. Zhang, H. Tang, K. Jia, and M. Tan. Domain-symmetric networks for adversarial domain adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5026–5035, June 2019.