Unsupervised Domain Attention Adaptation Network for Caricature Attribute Recognition - *Supplementary Material*

Wen Ji[1†], Kelei He[2,3†*], Jing Huo[1*], Zheng Gu[1], and Yang Gao[1,3]
{hkl,huojing,gaoy}@nju.edu.cn
{jiwen,guzheng}@smail.nju.edu.cn

[1] State Key Laboratory for Novel Software Technology, Nanjing, China
[2] Medical School of Nanjing University, Nanjing, China
[3] National Institute of Healthcare Data Science at Nanjing University, Nanjing, China

We further evaluate the effectiveness of the hyper-parameters, report the comparison of DAAN with state-of-the-art methods on three small benchmark datasets, and show the visualization results of class activation maps in this document.

1 Parameter Analysis

The performance of DAAN-A with respect to different weights of λ_a and DAAN-F with respect to different weights of λ_f is shown in Fig. 1. The results show the method is sensitive to the weight of λ_f, and is not sensitive to the weight of λ_a. The DAAN-F achieves the best performance of 0.7094 in F1 with weight $\lambda_f = 0.1$, and drops a little when further decreasing it. The experiments partially demonstrate that the features are not very efficient to the adaptation method, compared with the attribute-aware attention. The features contain noise may harmful to the network when it occupies a large portion.

Table 1. The performance of DAAN-F with respect to different λ_f and DAAN-A with respect to different λ_a.

Method		Param.	Avg. Acc	Avg. F1
DAAN	**-F**	$\lambda_f = 1$	0.8127	0.7028
		$\lambda_f = 0.1$	0.8169	0.7094
		$\lambda_f = 0.01$	0.8176	0.7089
	-A	$\lambda_a = 1$	0.8143	0.7106
		$\lambda_a = 0.1$	0.8147	0.7107
		$\lambda_a = 0.01$	0.8144	0.7103

[†]These authors contributed equally as co-first authors.
[*]Co-corresponding authors.

2 Comparison with the State-of-the-art Methods on Small Benchmark Datasets

To evaluate the generalization ability of DAAN, we compare it with several state-of-the-arts methods on small benchmark datasets, i.e., MNIST [4], USPS [3], SVHN [7]. For a fair comparison, we evaluate DAAN in the task of MNIST → USPS, USPS → MNIST, SVHN → MNIST. As reported in Table 2, DAAN is compared to six state-of-the-art methods, i.e., DAN [6], DANN [1], CoGAN [5], ADDA [9], CyCADA [2] and MCD [8]. Specifically, CyCADA also has an image-to-image generation module (i.e., CycleGAN) for cross-domain generation. As suggest by the table, DAAN performs best in the task of MNIST → USPS and USPS → MNIST, but slighty worse than MCD method in the task of SVHN → MNIST. We conclude that the image-to-image translation methods, including DAAN and CyCADA may more suitable for complex vision tasks. Since the image generation cannot generate totally good samples, the performance will be drop when facing failed generated samples, and thus leading to a not extremely high accuracy. Therefore, we think image generation is not necessary for simple domain adaptation tasks. For the difficult computer vision tasks, image generation helps narrow the gap between domains, and thus often helpful to improve the performance.

Table 2. Comparison with the state-of-the-art methods for performance in average Accuracy on small benchmark datasets.

Method	MNIST → USPS	USPS → MNIST	SVHN → MNIST
Source Only	82.2	69.6	67.1
DAN [6]	-	-	71.1
DANN [1]	77.1	73.0	71.1
CoGAN [5]	91.2	89.1	-
ADDA [9]	89.4	90.1	76.0
CyCADA pixel only [2]	95.6	96.4	70.3
CyCADA pixel+feat [2]	95.6	96.5	90.4
MCD [8]	94.2	94.1	**96.2**
DAAN-LFA (ours)	**97.1**	**97.0**	93.8
Target Only	96.3	99.2	99.2

3 Visualization Results of CAMs for DAAN

More visualization results of class activation maps (CAMs) for DAAN in target domain (i.e., caricatures) is shown in Fig. 1. The figure show consistently good quality of CAMs through different attributes. Moreover, the CAMs are accurate on caricatures even without the labeling of the images, and are robust in either whole-face attributes (e.g., 'Makeup') or regional face attributes (e.g., 'Toothy' and 'Large Forehead').

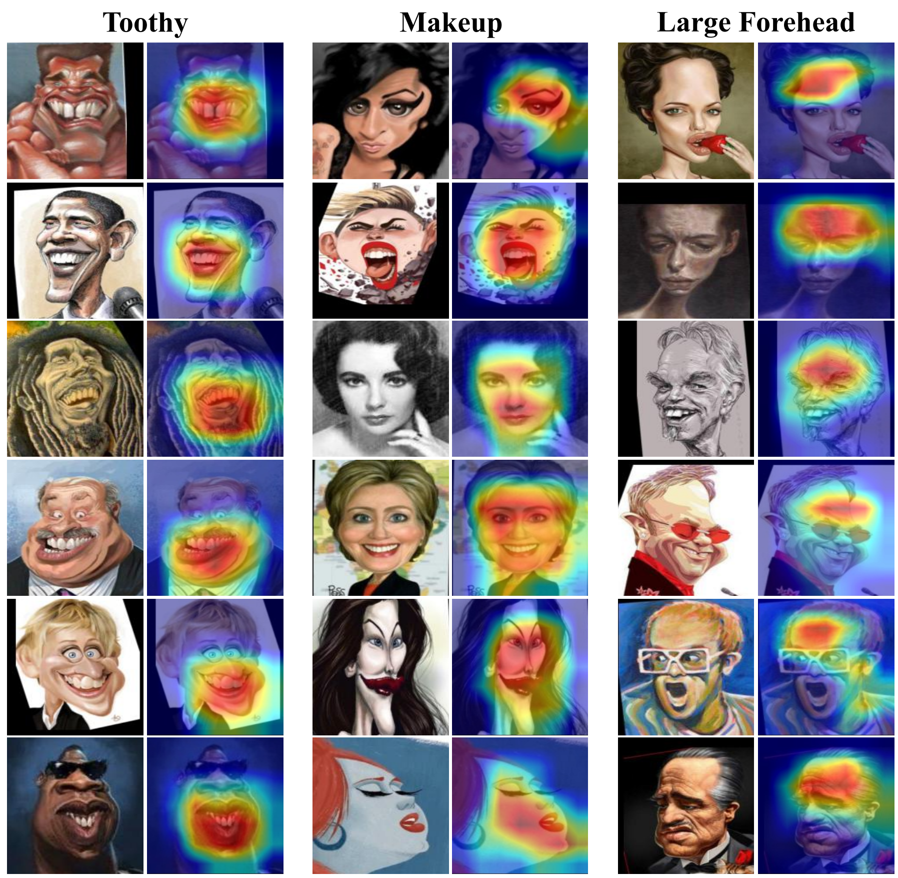

Fig. 1. More visualization results of CAMs for DAAN in three attributes of 'Toothy', 'Makeup' and 'Large Forehead'.

References

1. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International conference on machine learning. pp. 1180–1189 (2015)
2. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: International conference on machine learning. pp. 1989–1998 (2018)

3. Hull, J.J.: A database for handwritten text recognition research. IEEE Transactions on pattern analysis and machine intelligence **16**(5), 550–554 (1994)
4. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
5. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. In: Advances in neural information processing systems. pp. 469–477 (2016)
6. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: International conference on machine learning. pp. 97–105 (2015)
7. Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., Ng, A.Y.: Reading digits in natural images with unsupervised feature learning (2011)
8. Saito, K., Watanabe, K., Ushiku, Y., Harada, T.: Maximum classifier discrepancy for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3723–3732 (2018)
9. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7167–7176 (2017)