Supplementary Material: Bridging Adversarial and Statistical Domain Transfer via Spectral Adaptation Networks

Christoph Raab[0000−0001−9921−2668], Philipp Väth, Peter Meier, and Frank-Michael Schleif[0000−0002−7539−1283]

University of Applied Sciences Würzburg-Schweinfurt, Würzburg, Germany
{christoph.raab,frank-michael.schleif}@fhws.de

1 Introduction

This document contains additional results of the ACCV2020 publication *Bridging Adversarial and Statistical Domain Transfer via Spectral Adaptation Networks* not included in the main document due to space issues. This document consists of the following content:

- Sec. 2 analyzes the subspace similarity of our ASAN and related methods.
- Sec. 3 presents the ablation study showing the need for combing Spectral Normalization and our Relevance Spectral Loss.
- Sec. 4 evaluates the influence of the Hyperparameter k on the performance of ASAN.
- Sec. 5 presents additional performance results on the Visda [1] dataset.
- Sec. 6 presents additional results of the convergence and spectral analysis.
- Sec. 7 contains additional results of the feature analysis.
- Sec. 8 shows example images taken from the benchmark datasets used in the experiments.

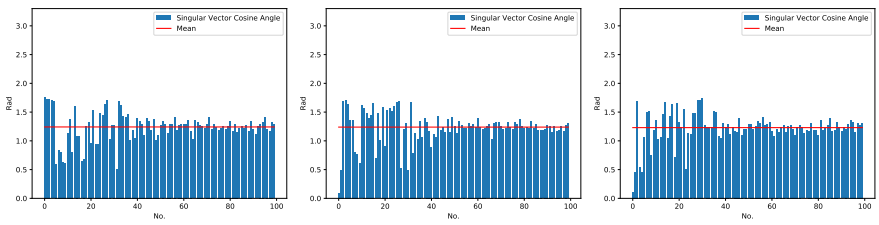
Note that the techniques used for the analysis in Sec. 6 and 7 are the same as the techniques used to obtain the results in the original manuscript. However, they are applied to different datasets to emphasize the usefulness and efficiency of the approach beyond the experiments shown in the main document.

2 Additional Results of Subspace Analysis

The transferability of features from source to target domain associated with the singular values σ_i and t_i is measurable by the cosine angle of associated singular vectors $\mathbf{v}_i$ and $\mathbf{r}_i$. While a high angle represents good transferability, this characteristic is associated with the most significant singular values in DANN [2]-type networks. This leads to devastating signals from mid-size singular values, containing necessary discriminative information for a classification task [3].

We follow the suggestion of [3] and analyze the cosine angle of the trained DANN [2], CDAN [4], and our proposed ASAN to measure the transferability.

The results are presented in Fig. 1 and show the cosine angle in rad of the singular vectors of the associated singular values between the source and target data in descending order with respect to the magnitude of the singular values. The data is obtained using **A**→**W** images from Office-31 [5] dataset. The plot indicates that DANN [2], as already stated in [3], has a too high focus on easily transferable features due to the high cosine similarity of the singular vector of the largest singular vectors compared to the rest of the spectrum. This makes DANN [2] vulnerable to neglect discriminative features [3]. The CDAN [4] model is more prone to this problem, showing a low angle on the first singular vectors. However, the ASAN model has a better overall distribution of cosine angles and distributes transferability better over the spectrums analyzed range. The red line shows the mean value, which is very similar in all three architectures, supporting the claim that our ASAN distributes transferability over the spectrum more successfully than compared approaches. Following the interpretation of [3], the ASAN model should, therefore, better express discriminative features over two domains. This finding reflects the main papers results by observing a better mean accuracy presented in the results section.



(a) Rad. of DANN [2] Bottleneck-Feature Subspace (b) Rad. of CDAN [4] Bottleneck-Feature Subspace (c) Rad. of ASAN Bottleneck-Feature Subspace

Fig. 1: Cosine angle in rad of singular vectors associated to the first 100 largest singular values between the source and target bottleneck feature output of the trained DANN [2], CDAN [4], and our ASAN. The data is obtained using **A**→**W** images from the Office-31 [5] dataset. The red line shows the mean rad. The plot indicates a better distribution of subspace similarities between the singular vectors of associated large and mid-size singular values of ASAN features. Best viewed on a computer display.

3 Ablation Study

In this section, we present the ablation study results to verify the effectiveness of our approach shown in Tab. 1. The building block of our ASAN network is the CDAN base network. The Spectral-Normalization is integrated into the discriminator network, and our proposed Relevance Spectral Loss is learned based on

bottleneck features. All parts are evaluated separately on the Office-31 dataset with the same experimental setup as in the main paper (see Sec. 4 in main paper for details).

The results show that the CDAN as base-networks has the worst performance, due to no additional adaptation techniques besides the inherent adaptation of CDAN. Integrating SN into CDAN helps to stabilize the discriminator gradients, and thereby, CDAN learns a better invariant representation. Learning our RSL with CDAN (CDAN+RSL), we can report that the result is marginally better as CDAN+SN. While not stabilizing the gradients, the RSL aligns the spectra of the source and target and therefore has increased performance compared to CDAN. Finally, combining RSL, SN, and CDAN (named as ASAN) leads to the best results because it combines the advantages of the just discussed techniques: the feature extractor receives stable gradients from the discriminator while the spectra are aligned.

Table 1: Result of the Ablation study as mean prediction **accuracy** with standard deviation on the **Office-31** dataset over three random runs.

Dataset	**A→W**	**D→W**	**W→D**	**A→D**	**D→A**	**W→A**	Avg.
CDAN	93.1±0.2	98.2±0.2	**100±0**	89.8±0.3	70.1±0.4	68.0±0.4	86.6
CDAN + SN	95.3±0.2	98.9±0.1	**100±0**	**94.7±0.3**	72.6±0.2	71.7±0.2	88.9
CDAN + RSL	95.1±0.0	**99.1±0.0**	**100.0±0.0**	93.0±0.0	73.8±0.0	73.2±0.0	89.0
ASAN (CDAN+RSL+SN)	**95.6±0.4**	98.8±0.2	**100±0**	94.4±0.9	**74.7±0.3**	**74.0±0.9**	**90.0**

4 Hyperparameter Tuning and Behavior

The influence in the prediction performance of ASANs only hyperparameter k over four datasets is shown in Fig. 2.

As already introduced in the main paper, k specifies the smallest source singular values group size, which are shrunk during learning. Again, because we inspect the smallest part of the source spectrum, the tuning of this parameter is different from common hyperparameter tuning: First, we define a range of spectral indices that we define as small, e.g., the smallest 20 singular values. Next, the hyperparameter is optimized via random search [6] in the domain adaptation setting [7], and finally, based on the best result, k is chosen accordingly. Note that all results in the main and supplementary paper are based on the tuning on the Office-31 dataset **A→B**.

We can report that the best value is $k = 11$, and as the plot indicates, minimizing the smallest parts of the spectra almost every time leads to better results. At *Art vs Clipart* the effect is significant, with a performance boost of over 20%. Our results confirm the results of [3] that the smallest part of the source spectrum contains domain-specific information and is at risk of harming the adaptation process.

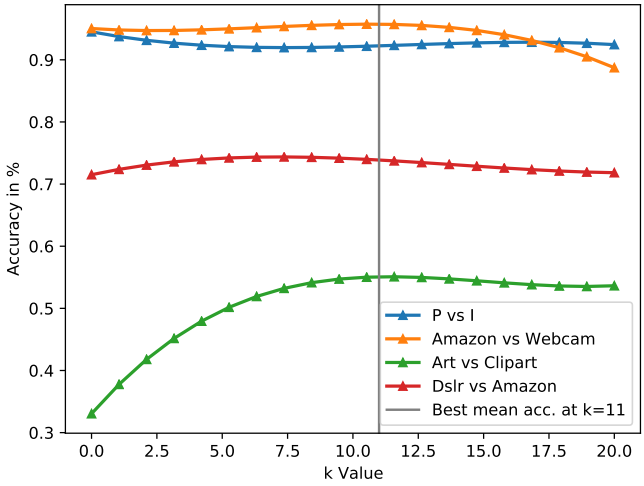


Fig. 2: Influence of the Hyperparameter k on the prediction performance of ASAN in accuracy with the best mean performance at $k = 11$. Best viewed on a computer display.

5 Additional Experimental Results on VisDa

This section introduces the VisDa dataset [1] and discusses the performance results of ASAN compared to other networks.

The VisDa [1] dataset is one of the largest domain adaptation benchmark datasets with over 280k images divided into the three subcategories training, validation and testing. The task is to train on 152,397 synthetic images and validate the obtained network on 55,388 real images. Both training and validation datasets consist of twelve classes with a minimum of approximately 6k examples and a maximum of approximately 16k class examples in the training domain. Some selected sample images are shown in Tab. 6 in Sec. 8.

The shown results are obtained using the same experimental setup as in the main paper (see Sec. 4 in main paper for details) and shown in Tab. 2. While

Table 2: **Mean** prediction **accuracy** with standard deviation on the **VisDa** dataset over three random runs. Results separated by baseline network ResNet-50 and ResNet-101 .

Method (ResNet-50)	Dataset Synthetic → Real	Method (ResNet-101)	Dataset Synthetic → Real
CDAN [8]	70.0	CDAN [8]	73.7
TAT [9]	71.9	BSP [3]	75.9
TransNorm [10]	71.6	SAFN [11]	**76.1**
ASAN	**74.0**	ASAN	74.7

using the ResNet-50 baseline, we can obtain very competitive results and very good performance improvement to the CDAN-baseline.

However, using ResNet-101, we are not able to report such an improvement in performance. There are two main reasons for this: first, the shrinkage parameter k is optimized using real images and discards domain specific information given on rich, detailed images. However, this richness is not present in synthetic images and leads to a drastic reduction of domain-specific information. Second, the parameter is optimized on the ResNet-50 bottleneck space while used in ResNet-101 with a different bottleneck space composition. Therefore, the occurrence of *domain-specific* in the source spectrum is different, which leads to a wrong application of the shrinkage mechanism. In future work, the k parameter will be optimized on the ResNet-101 network.

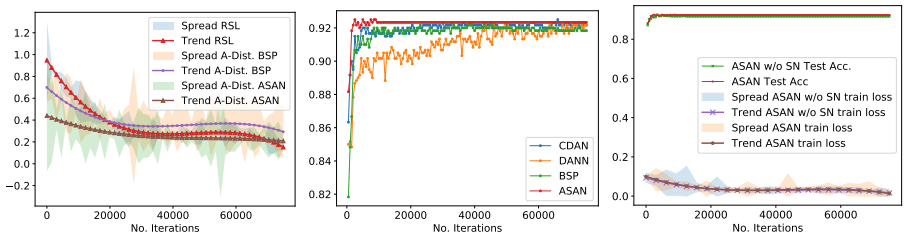
6 Additional Results of Convergence and Spectral Analysis

This sections present an additional analysis of convergence and spectral properties of the ASAN network against related methods. The data is obtained by learning and evaluating on $\mathbf{P}\rightarrow\mathbf{I}$ from the Image-Clef dataset.

The plot in Fig. 3a shows the performance of the proposed Relevance Spectral Loss (RSL) against BSP [3]. As in the main paper, we rely on the $\mathcal{A}$-Distance [12, 13] for the quality analysis of the invariant representation. The $\mathcal{A}$-Distance is defined as $\mathcal{A} = 2(2 - 1\varepsilon)$, where ε is the error of the trained domain classifier. The results confirm the results presented in the main proposal: the $\mathcal{A}$-Distance trend of ASAN (brown) is overall lower in value as the trend of BSP [3] (purple), allowing the statement that ASAN learns a better invariant representation faster than BSP [3]. The fluctuation around both trend lines, green for ASAN and orange for BSP [3], confirms this statement by having a spread generally oriented to lower $\mathcal{A}$-Distance values compared to BSP [3]. Additionally, the RSL is effectively reduced by ASAN during learning. This confirms that the similarity of spectra of both domains in the bottleneck space $\mathcal{F}$ are aligned, which is the reason for a better invariant representation of ASAN. Note that by the definition of the $\mathcal{A}$-Distance, negative values are possible [12, 13], indicating that the domain classifier has a lower performance than random guessing favorable for domain adaptation, because the source and target domain data are completely indistinguishable for the domain classifier.

The stability properties of the proposed ASAN in the main paper are also confirmed given different datasets. The plot is presented in Fig. 3b and shows that ASAN has the best overall accuracy. Further, the plots show that ASANs crucial advantage, besides the performance, is the convergence in an optimum, providing higher performance in accuracy than others. Once achieved, the optimum does not change afterward, while BSP [3], CDAN [4], and DANN [2] are fluctuating during the continuous learning process and, in the worse case, descend in performance.

The plot in Fig. 3c shows that our proposed RSL transfer-loss is substantially responsible for achieving high accuracy. Simultaneously, Spectral Normalization [14] supports the adaptation process via stability during training as the ASAN and ASAN w/o (without Spectral Normalization [14]) showing similar overall learning. The ASAN w/o achieves similar accuracy from an early stopping perspective. However, the Spectral Normalization [14] process within ASAN stabilizes the network in a high optimum. The combination is leading to the overall best prediction performance. The latter can again be confirmed by comparing the accuracy of Fig. 3b and Fig. 3c and in the experiments section of the main paper, where ASAN performs considerably better than SDAN [15]. SDAN consists of the same base-network and Spectral Normalization [14] without further spectral alignments necessary for rich domain adaptation, as presented with our ASAN model.



(a) A-Distance & RSL (b) Target Acc. (c) Source loss & Test Acc.

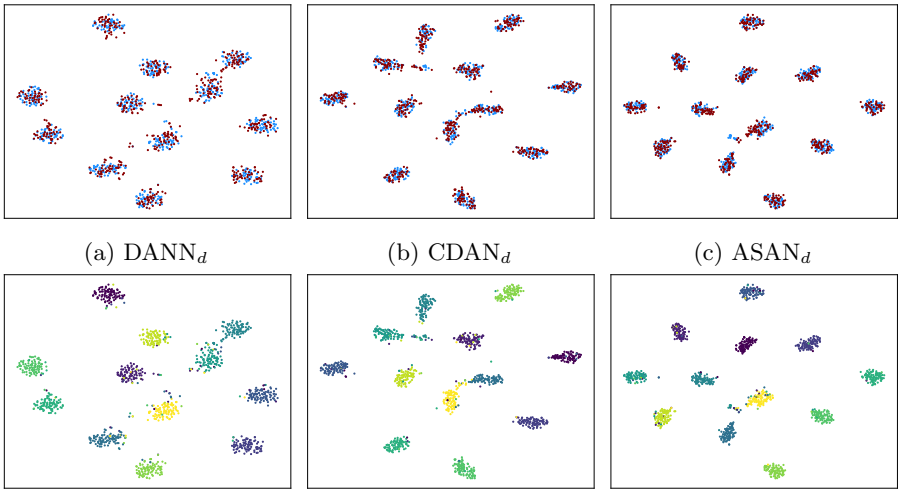
Fig. 3: Learning process of our ASAN compared to related networks over time given **P**→**I** images from Image-Clef dataset. Best viewed on computer display.

7 Additional Results of Feature Analysis

This sections presents an additional analysis of feature properties of the ASAN network against related methods. The data is obtain by learning and evaluating on **P**→**I** from the Image-Clef dataset.

The result is presented in Fig. 4 and is split into two parts: the top row (Fig. 4a - 4c) is a scatter plot of the bottleneck features of trained DANN [2], CDAN [4], and our ASAN colored with ground truth domain labels. Blue shows the source, and red shows the target domain. ASAN shows the superiority of creating a domain invariant representation by learning more compact clusters and observable separates the obtained clusters better compared to DANN [2] and CDAN [4]. The bottom row (Fig. 4d - 4f) shows the same representation but with classification labels. Due to dense and separated clusters, the obtained representation is easily classified by a neural network. However, again, the main proposals limitation also takes part in the plot, but is the same for DANN [2] and CDAN [4]. The limitation is the lack of approximating the joint distribution

of labels and features, leading to some samples assigned to the wrong cluster. See the main manuscript for a detailed discussion. Nevertheless, this effect is much weaker with ASAN than with CDAN [4] and DANN [2].



(a) DANN$_d$ (b) CDAN$_d$ (c) ASAN$_d$

(d) DANN$_c$ (e) CDAN$_c$ (f) ASAN$_c$

Fig. 4: T-SNE [16] of bottleneck features of selected networks given **P**→**I** images from the Image-Clef dataset. <Name>$_d$ and <Name>$_c$ show the outputs with ground truth domain and classification labels respectively. For the first row, blue shows the source, and red shows the target domain. Best viewed in color.

8 Examples from Benchmark Datasets

To provide an overview of the dataset characteristics and their differences among each other, we present some selected examples of the datasets Office-31 [5] in Tab. 3, Image-Clef in Tab. 4, Office-Home [17] in Tab. 5 and VisDa [1] in Tab. 6. With this, the focus of the sampling process lies in showing the difficulties between the dataset domains.

Intuitively, it is easy to see that the Office-31 [5] dataset is the easiest one among all four datasets. This is also confirmed in the experiments section of the main proposal. While Office-31 [5] contains objects from different angles and light settings, it is still observable that these images belong to the same class. The Image-Clef dataset is more complicated. However, the algorithms still show a consistently good performance. We assume that the main reason for the good performance is that the photographs are still made in some sense in the real world. However, when it comes to Office-Home [17], we observe that some

Table 3: Example images from the Office-31 [5] datasets. The objects shown capture the domain shift via difficult examples. Row caption is class name and column caption is domain name.

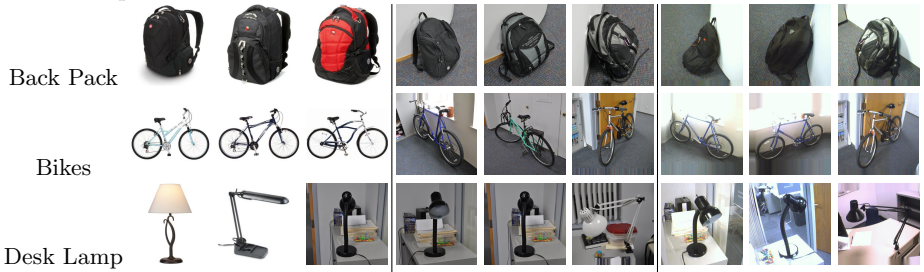


Table 4: Example images from the Image-Clef datasets. The objects shown capture the domain shift via difficult examples. Row caption is class name and column caption is domain name.

domains like Art show drastic changes in the appearances (style and presentation) of the images to the domain Real-World. Also, in comparison to Product, there are partially some broad domain shifts, which is the overall reason why all tested algorithms have worse performance on Office-Home [17] in comparison to Office-31 [5] and Image-Clef. On the other hand, the VisDa [1] dataset creates an entirely different scenario because the objects in the train images are entirely synthetic, while the validation dataset contains real images with partly non-trivial representations of an object. However, due to the small class size and large sample size, the networks are usually better on VisDa [1] as on Office-Home [17].

Table 5: Example images from the Office-Home [17] datasets. The objects shown capture the domain shift via difficult examples. Row caption is class name and column caption is domain name.

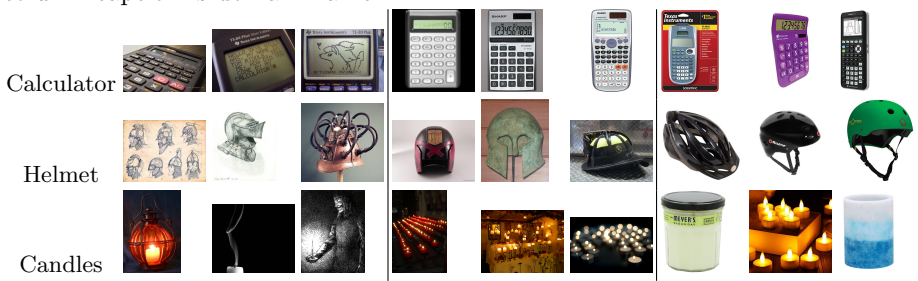
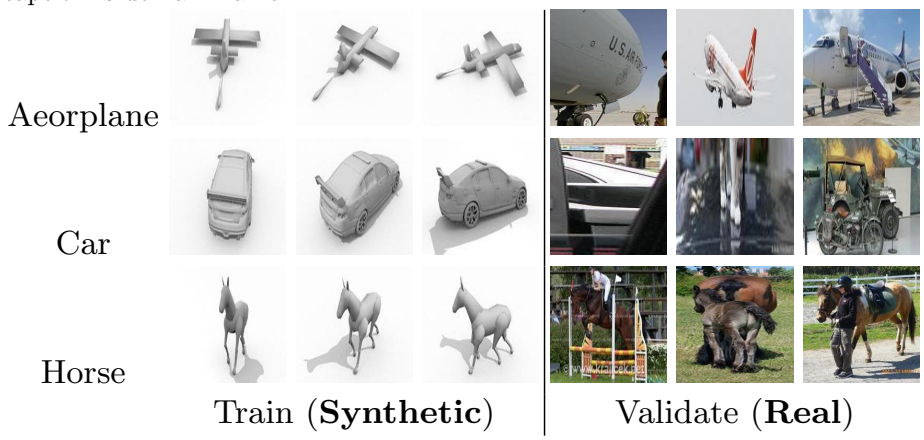


Table 6: Example images from the VisDa [1] datasets. The objects shown capture the domain shift via difficult examples. Row caption is class name and column caption is domain name.



References

1. Peng, X., Usman, B., Kaushik, N., Wang, D., Hoffman, J., Saenko, K.: VisDA: A Synthetic-to-Real Benchmark for Visual Domain Adaptation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Volume 2018-June., IEEE (2018) 2134–2139
2. Ganin, Y., Lempitsky, V.S.: Unsupervised Domain Adaptation by Backpropagation. In: Proceedings of the 32nd International Conference on Machine Learning, ICLR 2015, Lille, France, 6-11 July 2015. (2015) 1180–1189
3. Chen, X., Wang, S., Long, M., Wang, J.: Transferability vs. Discriminability: Batch Spectral Penalization for Adversarial Domain Adaptation. In Chaudhuri, K., Salakhutdinov, R., eds.: Proceedings of the 36th International Conference on Machine Learning. Volume 97 of Proceedings of Machine Learning Research., Long Beach, California, USA, PMLR (2019) 1081–1090
4. Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional Adversarial Domain Adaptation. In: Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada. (2018) 1647–1657
5. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting Visual Category Models to New Domains. In Daniilidis, K., Maragos, P., Paragios, N., eds.: Computer Vision – ECCV 2010, Berlin, Heidelberg, Springer Berlin Heidelberg (2010) 213–226
6. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. Journal of Machine Learning Research **13** (2012) 281–305
7. Zhong, E., Fan, W., Yang, Q., Verscheure, O., Ren, J.: Cross validation framework to choose amongst models and datasets for transfer learning. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **6323 LNAI** (2010) 547–562
8. Chen, X., Wang, S., Fu, B., Long, M., Wang, J.: Catastrophic Forgetting Meets Negative Transfer: Batch Spectral Shrinkage for Safe Transfer Learning. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E., Garnett, R., eds.: Advances in Neural Information Processing Systems 32. Curran Associates, Inc. (2019) 1908–1918
9. Liu, H., Long, M., Wang, J., Jordan, M.: Transferable Adversarial Training: A General Approach to Adapting Deep Classifiers. In Chaudhuri, K., Salakhutdinov, R., eds.: Proceedings of the 36th International Conference on Machine Learning. Volume 97 of Proceedings of Machine Learning Research., Long Beach, California, USA, PMLR (2019) 4013–4022
10. Wang, X., Jin, Y., Long, M., Wang, J., Jordan, M.I.: Transferable Normalization: Towards Improving Transferability of Deep Neural Networks. In Wallach, H., Larochelle, H., Beygelzimer, A., d\textquotesingle Alché-Buc, F., Fox, E., Garnett, R., eds.: Advances in Neural Information Processing Systems 32. Curran Associates, Inc. (2019) 1951–1961
11. Xu, R., Li, G., Yang, J., Lin, L.: Larger Norm More Transferable: An Adaptive Feature Norm Approach for Unsupervised Domain Adaptation. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Volume 2019-Octob., IEEE (2019) 1426–1435
12. Kifer, D., Ben-David, S., Gehrke, J.: Detecting change in data streams. In: Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30. VLDB '04, VLDB Endowment (2004) 180–191

13. Ben-David, S., Blitzer, J., Crammer, K., Pereira, F.: Analysis of Representations for Domain Adaptation. In Schölkopf, B., Platt, J.C., Hoffman, T., eds.: Advances in Neural Information Processing Systems 19. MIT Press (2007) 137–144
14. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral Normalization for Generative Adversarial Networks. In: International Conference on Learning Representations. (2018)
15. Zhao, L., Liu, Y.: Spectral Normalization for Domain Adaptation. Information **11** (2020)
16. Laurens van der Maaten, Geoffrey E., H.: Visualizing Data using t-SNE. Journal of Machine Learning Research **164** (2008) 10
17. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep Hashing Network for Unsupervised Domain Adaptation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Volume 2017-Janua., IEEE (2017) 5385–5394