CAD Models to Real-World Images: A Practical Approach to Unsupervised Domain Adaptation in Industrial Object Classification

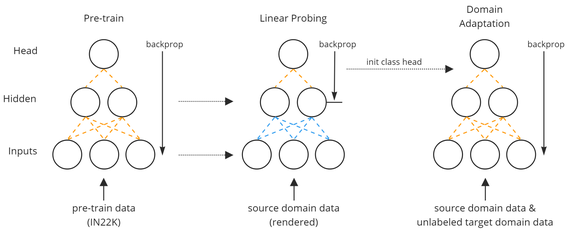
Dennis Ritter[1], Mike Hemberger[2], Marc Hönig[3], Volker Stopp[3], Erik Rodner[4], and Kristian Hildebrand[1]

[1] Berliner Hochschule für Technik, [2] nyris GmbH, [3] topex GmbH
[4] KI-Werkstatt/FB2, University of Applied Sciences Berlin

Abstract

In this paper, we systematically analyze **unsupervised domain adaptation pipelines** for object classification in a challenging **industrial setting**. We emphasize critical design considerations when utilizing category-labeled CAD models for classifying real-world images. Our domain adaptation pipeline **outperforms SoTA** performance on the **VisDA-2017 benchmark**. We conclude by providing **practical guidelines** for practitioners seeking to implement cutting-edge unsupervised domain adaptation techniques. Our code is available at github.com/dritter-bht/synthnet-transfer-learning

Pipeline



Contribution

- **Competitive two-stage transfer learning UDA pipeline** for object classification in a challenging industrial setting
- **First multi-domain (CAD, real) industrial image dataset** comprising 102 machine parts
- **Outperforming SoTA on VisDA-2017** classification benchmark
- **Creating benchmark performance on our new topex-printer dataset**

Motivation

- **Machine-specific specialists are often required to quickly identify components**, making it challenging for customers to independently recognize their machine parts
- **High cost of creating labeled images** for each component of complex machines often makes training automatic recognition systems unfeasible
- **Companies own the computer-aided design (CAD) data of the parts**, which can be rendered with any parameters and in any quantity

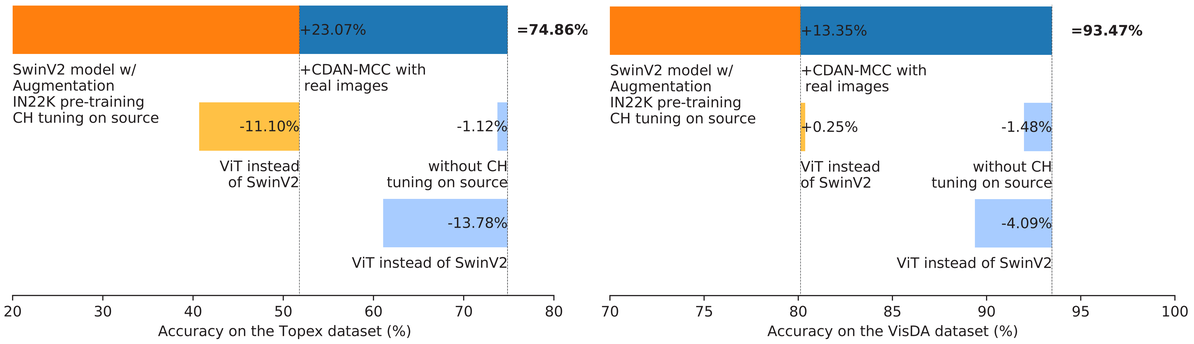
Topex-Printer Dataset

We present a **two-domain-dataset for image-based machine part identification**. It comprises **102 parts from a labeling machine** and is designed to mimic **real-world complexities**, including distinguishing **closely related classes**. The dataset consists of **3,264 CAD-rendered images** (32 per part) and **6,146 real images** (6 to 137 per part) for domain adaptation and testing purposes.

Source Domain
Synthetic 3D-object renders

Target Domain
Real photos

Results



Results of our DA pipeline for the (Left): **VisDA** and our new (Right) Topex-Printer dataset for different model and pre-training choices. Blue bars highlight results obtained using **UDA** with additional target images. It can be seen, that the **UDA stage training profits from class head tuning** in stage one and using the SwinV2 model instead of ViT.

Method		Pl	Bcl	Bus	Car	Hrs	Knf	Mcy	Per	Plt	Skb	Trn	Tck	Mean
TVT	ViT	92.9	85.6	77.5	60.5	93.6	98.2	89.3	76.4	93.6	92.0	91.7	55.7	83.9
CDTRANS	ViT	97.1	90.5	82.4	77.5	96.6	96.1	93.6	88.6	97.9	86.9	90.3	62.8	88.4
SDAT	ViT	98.4	90.9	85.4	82.1	98.5	97.6	96.3	86.1	96.2	96.7	92.9	56.8	89.8
MIC	ViT	**99.0**	93.3	86.5	87.6	**98.9**	**99.0**	**97.2**	**89.8**	98.9	**98.9**	96.5	68.0	92.8
Ours w/o UDA	ViT	96.48	71.82	90.14	**99.20**	94.66	77.71	87.28	44.45	95.12	83.64	94.05	40.76	80.54
Ours	ViT	94.82	93.49	92.80	95.89	90.95	88.51	77.46	75.42	96.27	97.32	94.74	88.03	89.38
Ours w/o UDA	Swin	97.09	80.48	85.35	98.12	92.39	83.54	94.85	19.89	89.13	78.89	**97.03**	55.18	80.12
Ours	Swin	97.96	**95.15**	**95.81**	98.64	98.34	95.68	80.12	83.87	**99.39**	94.68	96.61	**93.85**	**93.47**

Comparing accuracy to other literature on VisDA-2017 classification benchmark. **Our approach outperforms recent SoTA in mean accuracy**.

References

1. Kim, D., Wang, K., Sclaroff, S., Saenko, K.: A broad study of pre-training for domain generalization and adaptation. In: ECCV (2022)
2. Kumar, A., Raghunathan, A., Jones, R.M., Ma, T., Liang, P.: Fine-tuning can distort pretrained features and underperform out-of-distribution. In: ICLR (2022)
3. Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. In: NeurIPS (2017)
4. Jin, Y., Wang, X., Long, M., Wang, J.: Minimum class confusion for versatile domain adaptation. In: ECCV (2019)
5. Hoyer, L., Dai, D., Wang, H., Van Gool, L.: Mic: Masked image consistency for contextenhanced domain adaptation. In: CVPR (2023)
6. Peng, X., Usman, B., Kaushik, N., Wang, D., Hoffman, J., Saenko, K.: Visda: A synthetic-to-real benchmark for visual domain adaptation. In: CVPR-W (2018)

BHT Berliner Hochschule für Technik
htw Hochschule für Technik und Wirtschaft Berlin
University of Applied Sciences
nyris
topex
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung