Adaptiope: A Modern Benchmark for Unsupervised Domain Adaptation Supplementary Material

Tobias Ringwald Karlsruhe Institute of Technology tobias.ringwald@kit.edu

1. Introduction

Due to space limitations in the main paper, we provide further statistics and visualizations in this supplementary material. Section 2 shows indepth replacement statistics for the cleaned Amazon domain of our Refurbished Office-31 dataset. Section 3 provides further information about our Adaptiope dataset, the collection process of its synthetic domain and example images for 15 synthetic models in Figure 4 to 6. Furthermore, we show examples of label noise in the Amazon domain of Office-31 [3] in Figure 7. Finally, additional confusion matrices are shown in Figure 8.

2. Refurbished Statistics

We also provide further datasets statistics for our Refurbished Office-31 dataset. Table 1 depicts the image replacement rate per class for the cleaned Amazon domain. The rules according to which we replaced images were illustrated in the main paper. Additionally, the Dupes column shows how many duplicated images the original dataset contained. The *desktop computer* class was the most problematic in terms of ambiguous labels, as many images partially included objects from other classes such as *monitor*, *mouse, speaker* or *keyboard*.

3. Adaptiope

3.1. Dataset Statistics

The 123 classes of Adaptiope are: acoustic guitar, axe, backpack, baseball bat, bicycle, bicycle helmet, binoculars, bookcase, bottle, boxing gloves, brachiosaurus, calculator, car jack, cellphone, chainsaw, coat hanger, comb, compass, computer, computer mouse, cordless fixed phone, corkscrew, crown, dart, desk lamp, diving fins, drum set, electric guitar, electric shaver, fan, fighter jet, file cabinet, fire extinguisher, flat iron, game controller, glasses, golf club, grill, hair dryer, handcuffs, handgun, hand mixer, hard-wired fixed phone, hat, helicopter, hot glue gun, hourglass, hoverboard, ice cube tray, ice skates, inear headphones, keyboard, knife, ladder, laptop, lawn Rainer Stiefelhagen Karlsruhe Institute of Technology rainer.stiefelhagen@kit.edu



Figure 1: Examples for our Adaptiope dataset. From left to right: product, real life and synthetic images. Best viewed in the digital version.

mower, letter tray, magic lamp, microwave, mixing console, monitor, motorbike helmet, mug, nail clipper, network switch, notepad, office chair, over-ear headphones, pen, phonograph, pikachu, pipe wrench, pogo stick, power drill, power strip, printer, projector, puncher, purse, quadcopter, razor, rc car, rifle, ring binder, roller skates, rubber boat, ruler, scissors, scooter, screwdriver, sewing machine, shower head, skateboard, skeleton, sleeping bag, smartphone, smoking pipe, snow shovel, spatula, speakers, stand mixer, stapler, stethoscope, stroller, sword, syringe, tank, tape dispenser, telescope, tent, toilet brush, toothbrush, trash can, tyrannosaurus, umbrella, usb stick, vacuum cleaner, vr goggles, wallet, watering can, webcam, wheelchair and wristwatch.

As mentioned in the main paper, these classes are a superset of Office-31 [3] and can therefore be used as an extra



Figure 2: Rendering setup in Blender. Left: 100 randomly sampled viewpoints from the fibonacci sphere. Middle: positioning of object and camera. Right: bowl shaped shadow catcher surface for shadow casting.

domain or for research on cross-dataset UDA. Our Adaptiope dataset is perfectly balanced and contains 100 unique images per class and domain. We thus refrain from showing the common dataset histogram.

3.2. Synthetic Domain

For the synthetic domain of Adaptiope, we collected 615 models and raytraced them from 20 different viewpoints for a total of 12,300 images. For this, all models were manually positioned on a bowl-shaped shadow catcher surface (see Figure 2) and illuminated by a single light source. The bowl shape was chosen in order to prevent hard edges in the background while still allowing for shadows to be cast on a surface. After this manual setup, the rendering process was automatically completed by a Python script using the Blender API [1]. Camera viewpoints were generated by randomly sampling a 3D point from the upper half of the fibonacci sphere (see Figure 2) as origin for the camera which was then oriented towards the object's center of volume. As mentioned in the main paper, extreme viewpoints were manually filtered afterwards. Example viewpoints for the classes watering can, office chair and stroller are depicted in Figure 4 to 6.

3.3. Example Images

Finally, we show additional images for 52 out of the 123 classes of our Adaptiope dataset in Figure 1 and 3.

References

- [1] Blender Online Community. *Blender a 3D modelling and rendering package*. Blender Foundation, 2019.
- [2] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4893–4902, 2019.

Class Nama	Total	Dercentage	Dupes
Class Name	Peplaced	Perlaced (in %)	Dupes
	Replaced	Replaced (III %)	
back_pack	8	8.7	0
bike	20	24.4	0
bike_helmet	14	19.4	0
bookcase	19	23.2	0
bottle	20	55.6	0
calculator	2	2.1	0
desk_chair	33	36.3	1
desk_lamp	47	48.5	0
desktop_computer	84	86.6	15
file_cabinet	30	37.0	3
headphones	11	11.1	1
keyboard	44	44.0	2
laptop_computer	11	11.0	11
letter_tray	32	32.7	0
mobile_phone	4	4.0	1
monitor	13	13.1	4
mouse	13	13.0	3
mug	15	16.0	1
paper_notebook	27	28.1	3
pen	26	27.4	1
phone	31	33.3	12
printer	3	3.0	0
projector	14	14.3	3
punchers	68	69.4	0
ring_binder	40	44.4	2
ruler	25	33.3	4
scissors	55	55.0	0
speaker	25	25.3	0
stapler	36	36.4	0
tape_dispenser	38	39.6	3
trash_can	26	40.6	1
Overall	834	29.6	71

Table 1: Image replacement statistics per class for the Amazon domain of our Refurbished Office-31 dataset.

[3] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European conference on computer vision*, pages 213–226. Springer, 2010.



Figure 3: Additional examples for 44 out of the 123 classes of our Adaptiope dataset. From left to right: product, real life and synthetic images. Please note that the images were resized to a 1:1 aspect ratio for displaying purposes, our dataset contains the normally sized images. Best viewed in the digital version.



(a) Model 1

(b) Model 2





(d) Model 4



Figure 4: All 100 synthetic images for the watering can class. Images are grouped into the 5 model groups with 20 different viewpoints each. Best viewed in the digital version.





(d) Model 4



Figure 5: All 100 synthetic images for the office chair class. Images are grouped into the 5 model groups with 20 different viewpoints each. Best viewed in the digital version.





(a) Model 1





(c) Model 3

(d) Model 4



Figure 6: All 100 synthetic images for the stroller class. Images are grouped into the 5 model groups with 20 different viewpoints each. Best viewed in the digital version.



Figure 7: Example images for the *desktop computer* and *puncher* classes demonstrating label noise in the Amazon domain of the Office-31 [3] dataset. Most images are only vaguely related to the category due to the nature of the data collection process via web scraping, *e.g.* mainboards instead of full computers or punch cards instead of punchers. Overall, 69.4% of the *puncher* images and 86.6% of the *desktop computer* images had to be replaced for our Refurbished Office-31 dataset.



Figure 8: Confusion matrices for the W \rightarrow A and W \rightarrow A_{ref} tasks using CAN [2]. The refurbished Amazon domain exhibits a much more distinct diagonal due to the rectification of annotation errors. Best viewed in the digital version.