Domain2Vec: Domain Embedding for Unsupervised Domain Adaptation Supplementary Material

Xingchao Peng*[1], Yichen Li*[2], and Kate Saenko[1,3]

[1] Boston University, Boston, MA, USA
[2] Stanford University, Stanford, CA, USA
[3] MIT-IBM Watson AI Lab, Boston, MA, USA
{xpeng,saenko}@bu.edu, liyichen@stanford.edu

The supplementary material is organized as follows: Section A shows the comparison of our two datasets with the state-of-the-art cross-domain datasets. Section B describes the details of generating the TinyDA dataset. Section C shows the detailed information about DomainBank dataset. Section D introduces the detailed network framework for experiments on TinyDA dataset. Section E shows the additional experimental analysis. Section F shows the category information in the openset domain adaptation experiments.

A Comparison to modern datasets

Dataset	Year	Images	Classes	Domains	Description
Digit-Five	-	~100,000	10	5	digit
Office [1]	2010	4,110	31	3	office
Office-Caltech [2]	2012	2,533	10	4	office
CAD-Pascal [3]	2015	12,000	20	6	animal,vehicle
Office-Home [4]	2017	15,500	65	4	office, home
PACS [5]	2017	9,991	7	4	animal, stuff
Open MIC [6]	2018	16,156	-	-	museum
Syn2Real [7]	2018	280,157	12	3	animal,vehicle
DomainNet [8]	2019	569,010	345	6	clipart,sketch
TinyDA (Ours)	-	965,619	10 or 26	54	tiny images
DomainBank (Ours)	-	339,772	-	55	dataset

Table 1. A collection of most notable datasets to evaluate domain adaptation methods. Specifically, "Digit-Five" dataset indicates five most popular digit datasets (MNIST [9], MNIST-M [10], Synthetic Digits [10], SVHN, and USPS) which are widely used to evaluate domain adaptation models. Our dataset is challenging as it contains more images and domains than other datasets.

B TinyDA Generation

The images from TinyDA dataset are generated by blending different foreground shapes over patches randomly extracted from background images. In the first step, we select a foreground shape from the following five MNIST-style datasets: MNIST [11], USPS [12], EMNIST [13], KMNIST [14], QMNIST [15], and Fashion-MNIST [16]. Secondly, we choose a background pattern from the CIFAR10 [17]

* These authors contributed equally.

dataset or randomly cropped from BSD500 [18] dataset. Thirdly, we perform three different post-process to our rendered images: (1) replace the background with black patch, (2) replace the background with white path, (3) convert the generated images to grayscale images. These three post-processes, together with the original foreground images and the generated color images, form a dataset with five different modes, *i.e.* Black Background (*BB*), White Background (*WB*), GrayScale image (*GS*), Color (*Cr*) image, Original image(*Or*). In total, we generate a dataset with 54 domains and about one million MNIST-style training examples.

The image examples of our TinyDA dataset are shown in Table 2. Specifically, the upper and below table show the images generated with backgrounds from BSDS500 [18] and CIFAR10 [17], respectively. The image number of each domain in TinyDA dataset can be seen from Table 3.

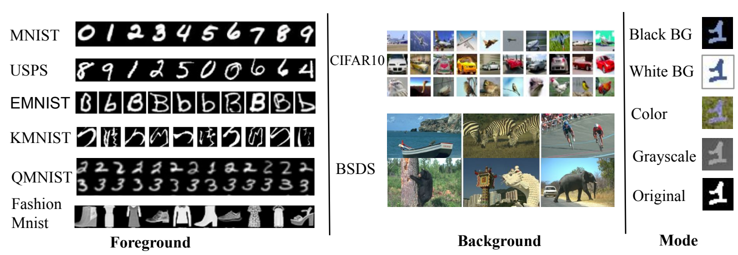


Fig. 1. Generation configuration for TinyDA dataset. We create our TinyDA dataset with six foregrounds, two backgrounds and five modes. The foreground images are from MNIST [11], USPS [12], EMNIST [13], KMNIST [14], QMNIST [15], FashionMNIST [16]. The background images are randomly sampled from CIFAR10 [17] or randomly cropped from BSDS500 [18] dataset. The five modes include "Black Background", "White Background", "Color", "GrayScale", and "Original".

.

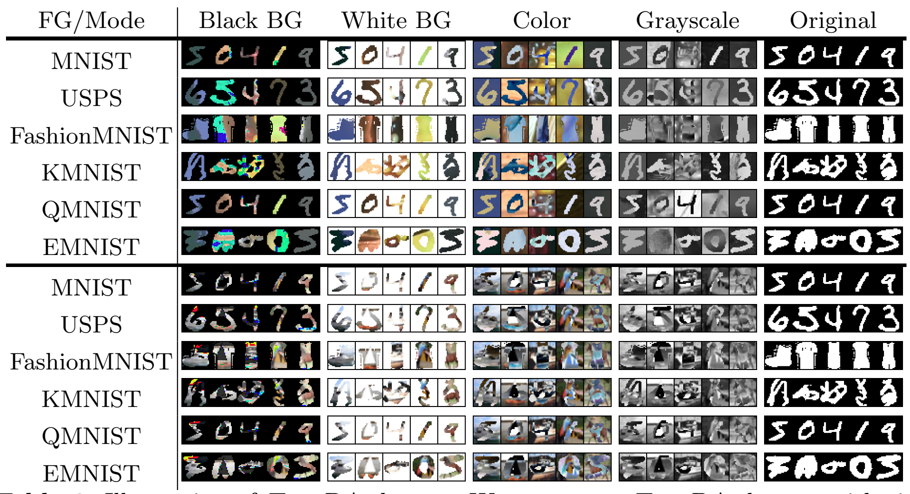


Table 2. Illustration of TinyDA dataset. We create our TinyDA dataset with six foregrounds, two backgrounds, and five modes. The upper and below table show the images generated with backgrounds from BSDS500 [18] and CIFAR10 [17], respectively.

FG/Mode	Black BG	White BG	Color	Grayscale	Original
MNIST	40,000	40,000	40,000	40,000	20,000
USPS	14,582	14,582	14,582	14,582	7291
FashionMNIST	40,000	40,000	40,000	40,000	20,000
KMNIST	40,000	40,000	40,000	40,000	20,000
QMNIST	40,000	40,000	40,000	40,000	20,000
EMNIST	40,000	40,000	40,000	40,000	20,000

Table 3. Number of images in each domain of TinyDA dataset.

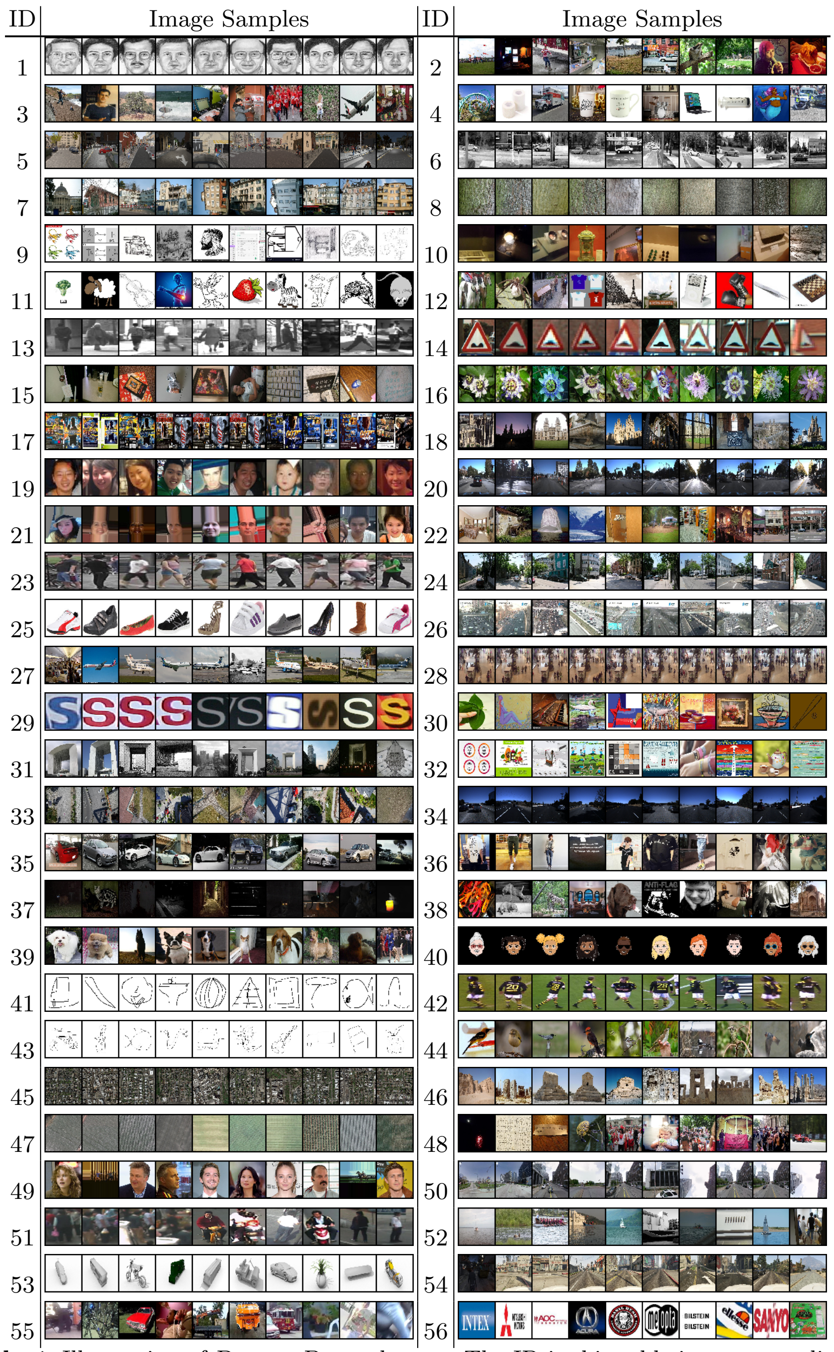


Table 4. Illustration of DomainBank dataset. The ID is this table is corresponding to the ID in Table 5.

ID	Dataset Name	Image#	Description	ID	Dataset Name	Image#	Description
1	CUFSF [19]	1,194	face_sketch	2	COCO [20]	10,000	real
3	PASCAL [21]	10,000	realh	4	DomainNet [8]	10,000	real
5	SYNTHIA [22]	10,000	street_syn	6	UIUC CAR [23]	1,220	car
7	ZuDuB [24]	210	building	8	Bark-101 [25]	2,586	bark
9	DomainNet [8]	10,000	sketch	10	Open-MIC [6]	10,000	indoor
11	DomainNet [8]	10,000	clipart	12	Caltech256 [26]	10,000	real
13	Ped. Detection [27]	10,000	pedestrian	14	Traffic Sign [28]	4,053	traffic
15	UKBench [29]	10,000	indoor_stuff	16	Oxford Flower [30]	8,189	flower
17	Caltech Games [31]	7,660	game_cover	18	Oxford Buildings [32]	5,063	building
19	GFW Face [33]	3,236	face	20	Driving [34]	9,420	road
21	MegaAge [35]	10,000	face	22	ADE20K [36]	10,000	indoor
23	Ped. Color [37]	10,000	pedestrian	24	LabelMeFacade [38]	395	building
25	UT Zappos50K [39]	10,000	shoes	26	TRANCOS [40]	1,244	traffic
27	FGVC [40]	10,000	aeroplane	28	Mall Dataset [41]	2,000	mall
29	Chars74K [42]	7,705	character	30	DomainNet [8]	10,000	painting
31	Paris Dataset [43]	3,187	street	32	DomainNet [8]	10,000	infograph
33	DroneDataset [44]	400	drone	34	Boxy [45]	2,148	road
35	Stanford Car [46]	8,144	car	36	DeepFashion2 [47]	10,000	fashion
37	ExDark [48]	6,619	dark	38	LaMem [49]	10,000	memorial
39	Stanford Dog [50]	10,000	dog	40	Cartoon Set [51]	9,999	cartoon
41	DomainNet [8]	10,000	quick_draw	42	Football [52]	771	football
43	Sketch Objects [53]	10,000	sketch	44	CUB200 [54]	10,000	bird
45	CITY-OSM [55]	914	drone_view	46	Arch Style [56]	4,630	building
47	UCM Land [57]	2,100	satellite	48	Privacy Attribute [58]	4,157	stuff
49	IMDB-WIKI [59]	10,000	face	50	Street View [60]	6,594	street
51	PPSS [61]	1,458	pedestrian	52	Sketch Retrieval [62]	1,213	sketch
53	VisDA [63]	10,000	syn	54	GTA [64]	5,000	syn
55	Youtube BBox [65]	10,000	real	56	Logo-2k+ [66]	10,000	logo

Table 5. Detailed information about our DomainBank dataset.

C DomainBank Dataset

The images of DomainBank dataset are sampled from 56 existing popular computer vision datasets. We choose the dataset with different image modalities, illuminations, camera perspectives *etc.* to increase the diversity of the domains. More details about our DomainBank benchmark are shown in Table 4 and Table 5. In total, we collect 339,772 images with image-level and domain-level annotations. Different from TinyDA, the categories of different domains in DomainBank are not identical. This property makes DomainBank a good testbed for Openset Domain Adaptation and Partial Domain Adaptation.

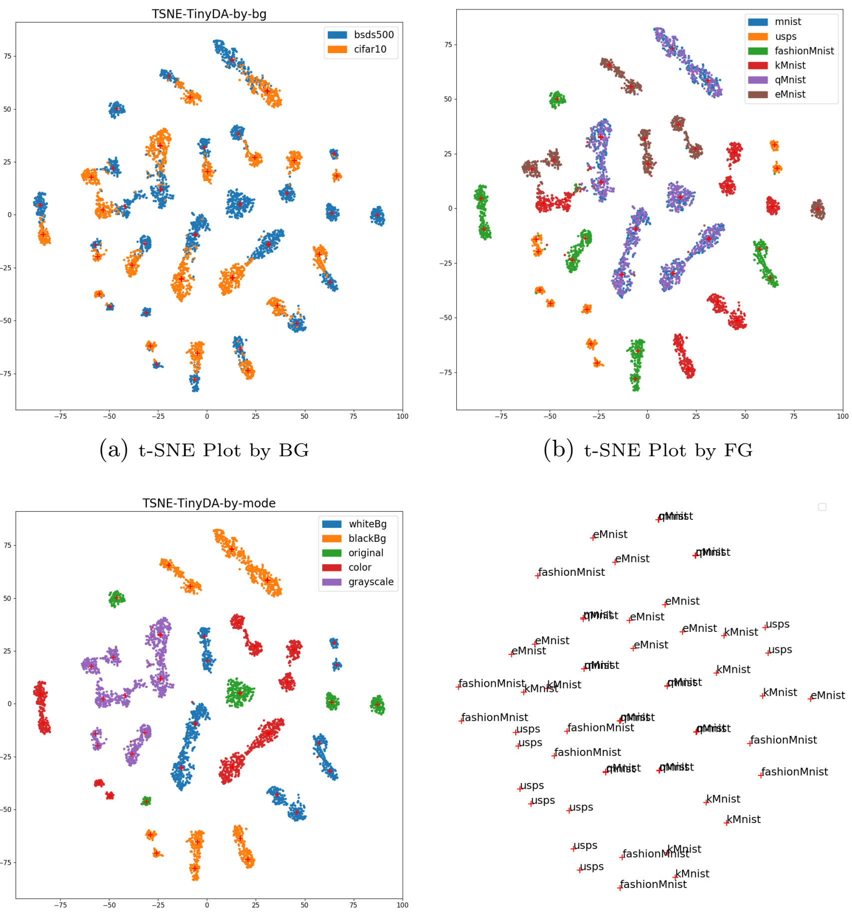
D Model architecture

The detailed network architecture for TinyDA dataset is shown in Table 6.

layer	configuration
	Feature Generator
1	Conv2D (3, 64, 5, 1, 2), BN, ReLU, MaxPool
2	Conv2D (64, 64, 5, 1, 2), BN, ReLU, MaxPool
3	Conv2D (64, 128, 5, 1, 2), BN, ReLU
	Disentangler
1	FC (8192, 3072), BN, ReLU
2	DropOut (0.5), FC (3072, 2048), BN, ReLU
	Domain Classifier
1	FC (2048, 256), LeakyReLU
2	FC (256, 56), LeakyReLU
	Classifier
1	FC (2048, 10 or 26), BN, Softmax
	Reconstructor
1	FC (4096, 8192)
	Mutual Information Estimator
fc1_x	FC (2048, 512)
fc1_y	FC (2048, 512), LeakyReLU
2	FC (512,1)

Table 6. Model Architecture for experiments on TinyDA dataset. For each convolution layer, we list the input dimension, output dimension, kernel size, stride, and padding. For the fully-connected layer, we provide the input and output dimensions. For drop-out layers, we provide the probability of an element to be zeroed.

E Additional experimental results



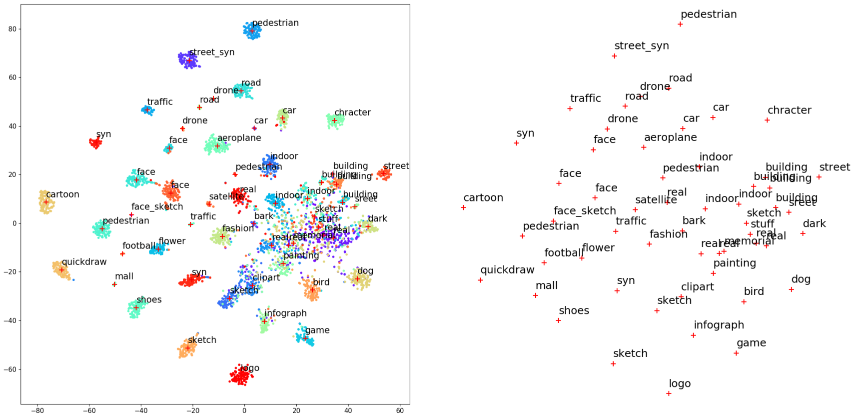
(a) t-SNE Plot by BG (b) t-SNE Plot by FG

(c) t-SNE Plot by Mode (d) Deep Embedding

Fig. 2. Deep domain embedding results of our DOMAIN2VEC model on TINYDA dataset: (**a**) t-SNE plot of the embedding result (color indicates different background); (**b**)t-SNE plot of the embedding result (color indicates different foreground); (**c**) t-SNE plot of the embedding result (color indicates different mode); (**d**) Deep embedding result. (Best viewed in color. Zoom in to see details.)

F Category information

For openset domain adaptation experiments, we choose the "aeroplane", "bus", "horse", "motorcycle", "plant", "train", and "truck" as the common categories



(a) t-SNE Plot by Domain (b) Deep Embedding of DomainBank

Fig. 3. Deep domain embedding results of our Domain2Vec model on DomainBank dataset: (**a**) t-SNE plot of the embedding result (color indicates different domain); (**d**) Deep embedding result. (Best viewed in color. Zoom in to see details.)

across the four domains. We set “bicycle”, “car” “knife”, “person”, “skateboard” as the unknown categories.

References

1. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting visual category models to new domains. In: European conference on computer vision, Springer (2010) 213–226
2. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 2066–2073
3. Peng, X., Sun, B., Ali, K., Saenko, K.: Learning deep object detectors from 3d models. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1278–1286
4. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing network for unsupervised domain adaptation. In: (IEEE) Conference on Computer Vision and Pattern Recognition (CVPR). (2017)
5. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.: Deeper, broader and artier domain generalization. In: International Conference on Computer Vision. (2017)
6. Koniusz, P., Tas, Y., Zhang, H., Harandi, M., Porikli, F., Zhang, R.: Museum exhibit identification challenge for the supervised domain adaptation and beyond. In: The European Conference on Computer Vision (ECCV). (September 2018)
7. Peng, X., Usman, B., Saito, K., Kaushik, N., Hoffman, J., Saenko, K.: Syn2real: A new benchmark forsynthetic-to-real visual domain adaptation. CoRR **abs/1806.09755** (2018)
8. Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B.: Moment matching for multi-source domain adaptation. In: Proceedings of the IEEE International Conference on Computer Vision. (2019) 1406–1415

9. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11) (1998) 2278–2324
10. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In Bach, F., Blei, D., eds.: Proceedings of the 32nd International Conference on Machine Learning. Volume 37 of Proceedings of Machine Learning Research., Lille, France, PMLR (07–09 Jul 2015) 1180–1189
11. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L.: Backpropagation applied to handwritten zip code recognition. Neural Computation (1989)
12. Hull, J.J.: A database for handwritten text recognition research. IEEE Transactions on pattern analysis and machine intelligence **16**(5) (1994) 550–554
13. Cohen, G., Afshar, S., Tapson, J., van Schaik, A.: Emnist: an extension of mnist to handwritten letters. arXiv preprint arXiv:1702.05373 (2017)
14. Clanuwat, T., Bober-Irizar, M., Kitamoto, A., Lamb, A., Yamamoto, K., Ha, D.: Deep learning for classical japanese literature. arXiv preprint arXiv:1812.01718 (2018)
15. Yadav, C., Bottou, L.: Cold case: The lost mnist digits. arXiv preprint arXiv:1905.10498 (2019)
16. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747 (2017)
17. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images. Technical report, Citeseer (2009)
18. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5) (May 2011) 898–916
19. Zhang, W., Wang, X., Tang, X.: Coupled information-theoretic encoding for face photo-sketch recognition. In: CVPR 2011, IEEE (2011) 513–520
20. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. In: European conference on computer vision, Springer (2014) 740–755
21. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International Journal of Computer Vision **88**(2) (June 2010) 303–338
22. Zolfaghari Bengar, J., Gonzalez-Garcia, A., Villalonga, G., Raducanu, B., Aghdam, H.H., Mozerov, M., Lopez, A.M., van de Weijer, J.: Temporal coherence for active learning in videos. In: The IEEE International Conference in Computer Vision, Workshops (ICCV Workshops). (2019)
23. UIUC: Uiuc car dataset, http://l2r.cs.uiuc.edu/ cogcomp/data/car/
24. Shao, H., Svoboda, T., Van Gool, L.: Zubud: Zurich buildings database for image based recognition. (01 2003)
25. Ratajczak, R., Bertrand, S., Crispim-Junior, C., Tougne, L.: Efficient bark recognition in the wild. (02 2019)
26. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. (2007)
27. Enzweiler, M., Gavrila, D.M.: Monocular pedestrian detection: Survey and experiments. IEEE transactions on pattern analysis and machine intelligence **31**(12) (2008) 2179–2195
28. Mathias, M., Timofte, R., Benenson, R., Van Gool, L.: Traffic sign recognition—how far are we from the solution? In: The 2013 international joint conference on Neural networks (IJCNN), IEEE (2013) 1–8

29. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Volume 2., Ieee (2006) 2161–2168
30. Nilsback, M.E., Zisserman, A.: A visual vocabulary for flower classification. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Volume 2., IEEE (2006) 1447–1454
31. Aly, M., Welinder, P., Munich, M., Perona, P.: Towards automated large scale discovery of image families. In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE (2009) 9–16
32. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: 2007 IEEE conference on computer vision and pattern recognition, IEEE (2007) 1–8
33. He, Y., Cao, K., Li, C., Loy, C.C.: Merge or not? learning to group faces via imitation learning. In: Thirty-Second AAAI Conference on Artificial Intelligence. (2018)
34. Udacity: https://github.com/udacity/self-driving-car
35. Zhang, Y., Liu, L., Li, C., Loy, C.C.: Quantifying facial age by posterior of age comparisons. In: British Machine Vision Conference (BMVC). (2017)
36. Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., Torralba, A.: Semantic understanding of scenes through the ade20k dataset. International Journal of Computer Vision **127**(3) (2019) 302–321
37. Cheng, Z., Li, X., Loy, C.C.: Pedestrian color naming via convolutional neural network. In: Asian Conference on Computer Vision, Springer (2016) 35–51
38. Fröhlich, B., Rodner, E., Denzler, J.: A fast approach for pixelwise labeling of facade images. In: Proceedings of the International Conference on Pattern Recognition (ICPR 2010). (2010)
39. Yu, A., Grauman, K.: Fine-grained visual comparisons with local learning. In: Computer Vision and Pattern Recognition (CVPR). (Jun 2014)
40. Guerrero-Gómez-Olmedo, R.: Extremely overlapping vehicle counting. In: Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA). (2015)
41. Change Loy, C., Gong, S., Xiang, T.: From semi-supervised to transfer counting of crowds. In: Proceedings of the IEEE International Conference on Computer Vision. (2013) 2256–2263
42. De Campos, T.E., Babu, B.R., Varma, M., et al.: Character recognition in natural images. VISAPP (2) **7** (2009)
43. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Lost in quantization: Improving particular object retrieval in large scale image databases. In: 2008 IEEE conference on computer vision and pattern recognition, IEEE (2008) 1–8
44. of Computer Graphics, I., Vision: http://dronedataset.icg.tugraz.at/
45. Behrendt, K.: Boxy vehicle detection in large images. In: Proceedings of the IEEE International Conference on Computer Vision. (2019)
46. Krause, J., Stark, M., Deng, J., Fei-Fei, L.: 3d object representations for fine-grained categorization. In: Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on, IEEE (2013) 554–561
47. Ge, Y., Zhang, R., Wu, L., Wang, X., Tang, X., Luo, P.: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. CVPR (2019)
48. Loh, Y.P., Chan, C.S.: Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding **178** (2019) 30–42

49. Khosla, A., Raju, A.S., Torralba, A., Oliva, A.: Understanding and predicting image memorability at a large scale. In: International Conference on Computer Vision (ICCV). (2015)
50. Khosla, A., Jayadevaprakash, N., Yao, B., Fei-Fei, L.: Novel dataset for fine-grained image categorization. In: First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO (June 2011)
51. Royer, A., Bousmalis, K., Gouws, S., Bertsch, F., Mosseri, I., Cole, F., Murphy, K.: Xgan: Unsupervised image-to-image translation for many-to-many mappings. In: Domain Adaptation for Visual Understanding. Springer (2020) 33–49
52. Kazemi, V., Sullivan, J.: Using richer models for articulated pose estimation of footballers. In: BMVC. (2012)
53. Eitz, M., Hays, J., Alexa, M.: How do humans sketch objects? ACM Trans. Graph. (Proc. SIGGRAPH) **31**(4) (2012) 44:1–44:10
54. Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S., Perona, P.: Caltech-ucsd birds 200. (2010)
55. Kaiser, P., Wegner, J.D., Lucchi, A., Jaggi, M., Hofmann, T., Schindler, K.: Learning aerial image segmentation from online maps. IEEE Transactions on Geoscience and Remote Sensing **55**(11) (2017) 6054–6068
56. Xu, Z., Tao, D., Zhang, Y., Wu, J., Tsoi, A.C.: Architectural style classification using multinomial latent logistic regression. In: European Conference on Computer Vision, Springer (2014) 600–615
57. Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. (2010) 270–279
58. Orekondy, T., Schiele, B., Fritz, M.: Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In: IEEE International Conference on Computer Vision (ICCV). (2017)
59. Rothe, R., Timofte, R., Gool, L.V.: Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision **126**(2-4) (2018) 144–157
60. Zamir, A., Shah, M.: Image geo-localization based on multiple nearest neighbor feature matching using generalized graphs (2014)
61. Luo, P., Wang, X., Tang, X.: Pedestrian parsing via deep decompositional network. In: Proceedings of the IEEE international conference on computer vision. (2013) 2648–2655
62. Eitz, M., Hildebrand, K., Boubekeur, T., Alexa, M.: Sketch-based image retrieval: Benchmark and bag-of-features descriptors. IEEE Transactions on Visualization and Computer Graphics **17**(11) (2011) 1624–1636
63. Peng, X., Usman, B., Kaushik, N., Hoffman, J., Wang, D., Saenko, K.: Visda: The visual domain adaptation challenge. arXiv preprint arXiv:1710.06924 (2017)
64. Richter, S.R., Vineet, V., Roth, S., Koltun, V.: Playing for data: Ground truth from computer games. In Leibe, B., Matas, J., Sebe, N., Welling, M., eds.: European Conference on Computer Vision (ECCV). Volume 9906 of LNCS., Springer International Publishing (2016) 102–118
65. Real, E., Shlens, J., Mazzocchi, S., Pan, X., Vanhoucke, V.: Youtube-boundingboxes: A large high-precision human-annotated data set for object detection in video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2017) 5296–5305

66. Wang, J., Min, W., Hou, S., Ma, S., Zheng, Y., Wang, H., Jiang, S.: Logo-2k+: A large-scale logo dataset for scalable logo classification. arXiv preprint arXiv:1911.07924 (2019)