Supplementary Material: Neural Data Server: A Large-Scale Search Engine for Transfer Learning Data

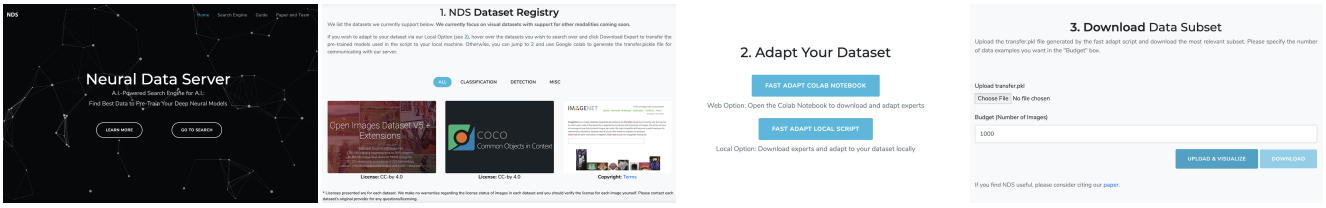
Xi Yan[1,2*] David Acuna[1,2,3*] Sanja Fidler[1,2,3]
[1]University of Toronto [2]Vector Institute [3]NVIDIA
xi.yan@mail.utoronto.ca, {davidj, fidler}@cs.toronto.edu

1. Appendix

In the Appendix, we provide additional details and results for our Neural Data Server.

1.1. Web Interface

Our NDS is running as a web-service at http://aidemos.cs.toronto.edu/nds/. We are inviting interested readers to try it and give us feedback. A snapshot of the website is provided in Figure 1.



aidemos.cs.toronto.edu/nds/ Dataset Registry Adapt Experts Download Recommended Data

Figure 1: Our Neural Data Server web-service. Note that NDS does not host datasets, but rather links to datasets hosted by original providers.

1.2. Additional Results

We visually assess domain confusion in Figures 2, 3, 4. We randomly select 9 images per cluster and display the top 8 clusters corresponding to the experts with the highest proxy task performance in miniModaNet, Cityscapes, and VOC-Pascal. We can observe that the images from the top clusters do indeed reflect the types of objects one encounters for autonomous driving, fashion, and general scenes corresponding to the respective target (client) datasets, showcasing the plausibility of our NDS.

We further extend Table 2 and 3 in the main paper by showing detailed instance segmentation results for fine-tuning on the Cityscapes dataset. We report the performance measured by the COCO-style mask AP (averaged over IoU thresholds) for the 8 object categories. Table 4 reports the mask AP by sampling 23K, 47K, and 59K images from COCO to be used for pre-training for Cityscapes, and Table 5 reports the mask AP by sampling 118K, 200K images from OpenImages for pre-training.

Self-Supervised Pretraining: We evaluate NDS in a scenario where a client uses self-supervised learning to pretrain on the selected server data. We follow the same setup as described in Section 4.2, except that rather than pretraining using classification labels, clients ignore the availability of the labels and pretrain using two self-supervised learning approaches: MoCo [5] and RotNet [4]. In Table 1, we pretrain on the selected data subset using MoCo, an approach recently proposed by He *et al.*, where the model is trained on the pretext task of instance discrimination. In Table 2, we use [4] to pretrain our model on the pretext task of predicting image rotation. We observe that in the case of MoCo, pretraining on NDS selected subset does not always yield better performance than pretraining on a randomly sampled subset of the same size. In the case of RotNet, pretraining on NDS selected subset has a slight gain over the baseline of uniform sampling. These results suggest that the optimal dataset for pretraining using self-supervised learning may be dependent on the pretext task. More formal studies on the relationship connecting training data, pretraining task, and transferring performance is required.

*authors contributed equally

Pretrain. Sel.	Method	Target Dataset				
		Stanf. Dogs	Stanf. Cars	Oxford-IIIT Pets	Flowers 102	CUB200 Birds
0%	Random Init.	23.66	18.60	32.35	48.02	25.06
100%	Entire Dataset	41.64	46.83	56.34	67.17	35.28
20%	Uniform Sample	41.01	44.16	56.01	64.42	34.41
	NDS	39.72	43.56	54.62	65.90	34.57

Table 1: MoCo Pretraining: Top-1 classification accuracy on five client datasets (columns) pretrained on the different subsets of data (rows) on the pretext task of instance discrimination.

Pretrain. Sel.	Method	Target Dataset				
		Stanf. Dogs	Stanf. Cars	Oxford-IIIT Pets	Flowers 102	CUB200 Birds
0%	Random Init.	23.66	18.60	32.35	48.02	25.06
100%	Entire Dataset	47.83	55.87	67.54	78.99	44.25
20%	Uniform Sample	42.74	42.82	60.25	72.59	39.30
	NDS	43.33	43.34	61.49	72.85	40.47

Table 2: RotNet Pretraining: Top-1 classification accuracy on five client datasets (columns) pretrained on the different subsets of data (rows) on the pretext task of predicting image rotations.

Dataset	Images	Class	Task	Evaluation Metric
Downsampled ImageNet [1]	1281167	1000	classification	-
OpenImages [9]	1743042	601(bbox) / 300(mask)	detection	-
COCO [10]	118287	80	detection	-
VOC2007 [3]	5011(trainval) / 4962(test)	20	detection	mAP
miniModaNet [14]	1000(train) / 1000(val)	13	detection	mAP
Cityscapes [2]	2975(train) / 500(val)	8	detection	mAP
Stanford Dogs [7]	12000(train) / 8580(val)	120	classification	Top-1
Stanford Cars [8]	8144(train) / 8041 (val)	196	classification	Top-1
Oxford-IIIT Pets [12]	3680(train) / 3369(val)	37	classifiation	Top-1
Flowers 102 [11]	2040(train) / 6149(val)	102	classification	Top-1
CUB200 Birds [13]	5994(train) / 5794(val)	200	classification	Top-1

Table 3: Summary of the number of images, categories, and evaluation metrics for datasets used in our experiments. We used 10 datasets (3 server datasets and 7 client datasets) to evaluate NDS.

Data (# Images)	Method	AP^{bb}	AP^{bb}_{50}	AP	AP_{50}	car	truck	rider	bicycle	person	bus	mcycle	train
0	ImageNet Initialization	36.2	62.3	32.0	57.6	49.9	30.8	23.2	17.1	30.0	52.4	17.9	35.2
23K	Uniform Sampling	38.1	64.9	34.3	60.0	50.0	34.2	24.7	19.4	32.8	52.0	18.9	42.1
	NDS	40.7	66.0	36.1	61.0	51.3	35.4	25.9	20.4	33.9	56.9	20.8	44.0
47K	Uniform Sampling	39.8	65.5	34.4	60.0	50.7	31.8	25.4	18.3	33.3	55.2	21.2	38.9
	NDS	42.2	68.1	36.7	62.3	51.8	36.9	26.4	19.8	33.8	59.2	22.1	44.0
59K	Uniform Sampling	39.5	64.9	34.9	60.4	50.8	34.8	26.3	18.9	33.2	55.5	20.8	38.7
	NDS	41.7	66.6	36.7	61.9	51.7	37.2	26.9	19.6	34.2	56.7	22.5	44.5
118K	Full COCO	41.8	66.5	36.5	62.3	51.5	37.2	26.6	20.0	34.0	56.0	22.3	44.2

Table 4: Transfer to instance segmentation with Mask R-CNN [6] on Cityscapes by selecting images from COCO.

Data (# Images)	Method	AP^{bb}	AP^{bb}_{50}	AP	AP_{50}	car	truck	rider	bicycle	person	bus	mcycle	train
0	ImageNet Initialization	36.2	62.3	32.0	57.6	49.9	30.8	23.2	17.1	30.0	52.4	17.9	35.2
118K	Uniform Sampling	37.5	62.5	32.8	57.2	49.6	33.2	23.3	18.0	30.8	52.9	17.4	37.1
	NDS	39.9	65.1	35.1	59.8	51.6	36.7	24.2	18.3	32.4	56.4	18.0	42.8
200K	Uniform Sampling	37.8	63.1	32.9	57.8	49.7	31.7	23.8	17.8	31.0	51.8	18.4	38.8
	NDS	40.7	65.8	36.1	61.2	51.4	38.2	24.2	17.9	32.3	57.8	19.7	47.3

Table 5: Transfer to instance segmentation with Mask R-CNN [6] on Cityscapes by selecting images from OpenImages.

References

[1] Patryk Chrabaszcz, Ilya Loshchilov, and Frank Hutter. A downsampled variant of imagenet as an alternative to the cifar datasets, 2017. 2

[2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. *CVPR*, pages 3213–3223, 2016. 2

[3] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John M. Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88:303–338, 2009. 2

[4] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. In *ICLR*, 2018. 1

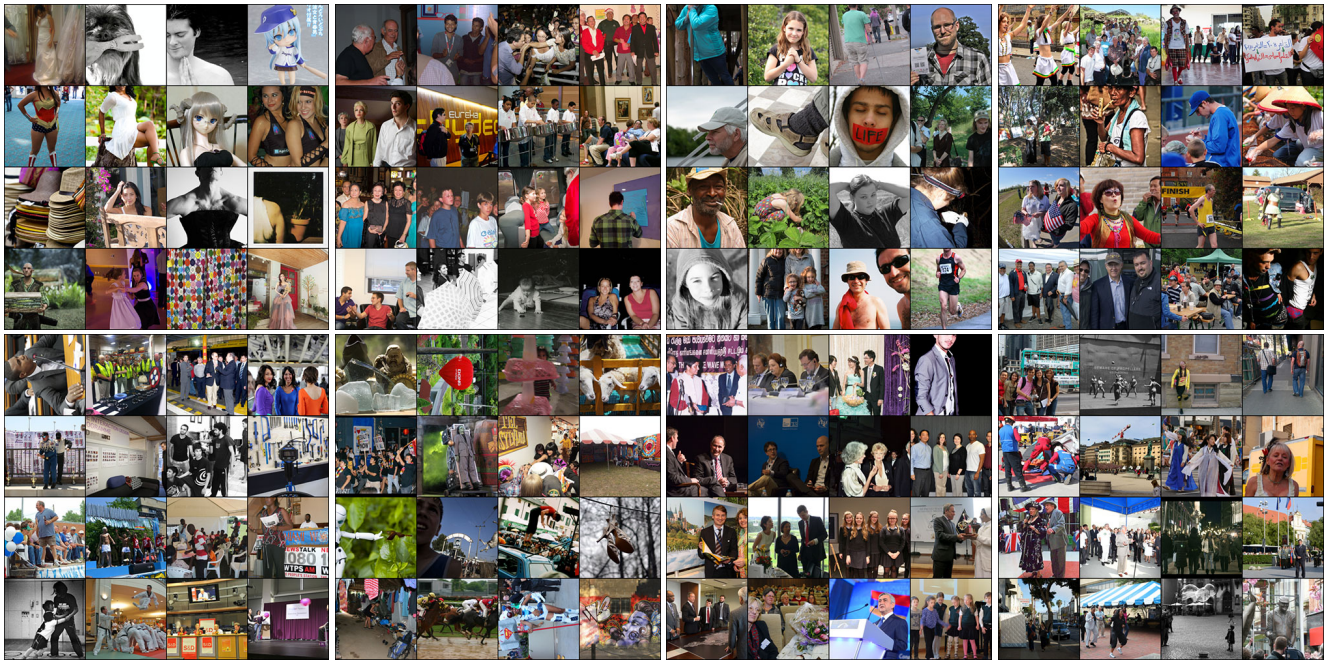


Figure 2: Top 8 clusters from COCO+OpenImages corresponding to the best performing expert adapted on miniModaNet.

Figure 3: Top 8 clusters from COCO+OpenImages corresponding to the best performing expert adapted on Cityscapes.

[5] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning, 2019. 1

[6] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask r-cnn. *ICCV*, pages 2980–2988, 2017. 2

[7] Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao, and Li Fei-Fei. Novel dataset for fine-grained image categorization. In *CVPR Workshop on Fine-Grained Visual Categorization*, Colorado Springs, CO, June 2011. 2

[8] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *IEEE Workshop on 3D Representation and Recognition (3dRR-13)*, Sydney, Australia, 2013. 2

Figure 4: Top 8 clusters from COCO+OpenImages corresponding to the best performing expert adapted on PASCAL VOC.

[9] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Tom Duerig, and Vittorio Ferrari. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *arXiv:1811.00982*, 2018. 2

[10] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014. 2

[11] M-E. Nilsback and A. Zisserman. Automated flower classification over a large number of classes. In *Proc. of the Indian Conference on Computer Vision, Graphics and Image Processing*, Dec 2008. 2

[12] O. M. Parkhi, A. Vedaldi, A. Zisserman, and C. V. Jawahar. Cats and dogs. In *CVPR*, 2012. 2

[13] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 2

[14] Shuai Zheng, Fan Yang, M. Hadi Kiapour, and Robinson Piramuthu. Modanet: A large-scale street fashion dataset with polygon annotations. In *ACM Multimedia*, 2018. 2