Supplementary Material
Metric imitation by manifold transfer for efficient vision applications

Dengxin Dai[1] Till Kroeger[1] Radu Timofte[1] Luc Van Gool[1,2]
[1]Computer Vision Lab, ETH Zurich [2] VISICS, ESAT/PSI, KU Leuven
{dai, kroegert, timofter, vangool}@vision.ee.ethz.ch

1. Image Clustering

Below we show the performance of Metric Imitation (MI) on the task of Image Clustering, when the GIST feature [12] and the PHOG feature [1] are used as the target features (TFs). Source features (SFs), remain the same as used in the paper; they are SIFT-llc [14], object-bank(OB) [10], and the CNN feature [2]. The four datasets Scene-15 [8], CUReT-61 [5], Caltech-101 [6], and Event-8 [9] are used.

We follow [3, 4, 13] and used Purity (bigger=better) for evaluation. Table 1 and Table 2 show the results of the GIST feature and the PHOG feature, when 50% images are used as training and the rest for testing. It can be observed from tables that MI yields better results for the task image clustering (relative to the original TFs), by transferring manifold from domain of the source features.

2. Category-based Image Retrieval

Below we show the performance of Metric Imitation (MI) on the task Category-based image retrieval, when the GIST feature [12] and the PHOG feature [1] are used as the target features (TFs). Source features (SFs), remain the same as used in the paper; they are SIFT-llc [14], object-bank(OB) [10], and the CNN feature [2]. The four datasets Scene-15 [8], CUReT-61 [5], Caltech-101 [6], and Event-8 [9] are used. Mean of average precision (MAP) is used as the evaluation criterion, when recall is set to 0.1 as precision of top retrieved images is often more important than recall.

MAP is used as the evaluation criterion. Table 3 and Table 4 show the results of GIST PHOG, when 50% images are used as training and the rest for testing. It can be observed from tables that MI also yields better performance for the task of category-based image retrieval (relative to the original TFs), by transferring manifold from the domain of source features.

3. Instance-based Object Retrieval

Below we show the performance of Metric Imitation (MI) on the task of instance-based object retrieval, when the GIST feature [12] and the PHOG feature [1] are used as the target features (TFs), features in target domain. Source features (SFs), features in source domain, remain the same as used in the paper; they are SIFT-llc [14], object-bank(OB) [10], and the CNN feature [2]. The two popular datasets INRIA Holidays datase [7] and the UKbench dataset [11] are used. Mean of average precision (MAP) is used as the evaluation criterion, when recall is set to 1. MI is trained on different datasets: trained on Scene-15 for Holidays as both contains images of scenes, and trained on Caltech-101 for UKbench as both contain images of objects.

MAP is used as the evaluation criterion. Table 5 and Table 6 show the results of the GIST feature and the PHOG feature. It can be seen from tables that MI yields better performance for instance-based object retrieval (relative to the original TFs), by transferring manifold from the domain of source features, even learned from different datasets and without using any labels. The improvement for PHOG is not as significant as what has been shown for LBP and GIST. This is due to the fact that PHOG is inherently not as suitable for the task of instance-based object retrieval when invariance to rotations and views is desired.

References

[1] A. Bosch, A. Zisserman, and X. Muoz. Image classification using random forests and ferns. In *ICCV*, 2007. 1

[2] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *BMVC*, 2014. 1

[3] D. Dai, M. Prasad, C. Leistner, and L. V. Gool. Ensemble partitioning for unsupervised image categorization. In *ECCV*, 2012. 1

[4] D. Dai, T. Wu, and S. C. Zhu. Discovering scene categories by information projection and cluster sampling. 2010. 1

[5] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Trans. Graph.*, 18(1):1–34, 1999. 1

978-1-4673-6964-0/15/$31.00 ©2015 IEEE

	TFs	MI		SFs	MI		SFs	MI		SFs
	GIST	MI_LLE	MI_Lap	SIFT-llc	MI_LLE	MI_Lap	CNN	MI_LLE	MI_Lap	OB
Scene-15	0.39	0.40	0.42	0.50	0.45	0.45	0.71	0.42	0.41	0.58
CUReT-61	0.32	0.38	0.37	0.41	0.31	0.34	0.63	0.33	0.32	0.40
Caltech-101	0.31	0.33	0.33	0.51	0.34	0.33	0.68	0.32	0.33	0.49
Event-8	0.34	0.46	0.45	0.55	0.44	0.45	0.78	0.45	0.45	0.44

Table 1. Purity of clustering by Metric Imitation (MI) when GIST is used as the TFs, where 50% of the images are used for training and the rest for testing, with recall set to 0.1.

	TFs	MI		SFs	MI		SFs	MI		SFs
	GIST	MI_LLE	MI_Lap	SIFT-llc	MI_LLE	MI_Lap	CNN	MI_LLE	MI_Lap	OB
Scene-15	0.28	0.33	0.33	0.47	0.31	0.33	0.72	0.31	0.33	0.59
CUReT-61	0.21	0.38	0.38	0.40	0.26	0.37	0.61	0.29	0.33	0.41
Caltech-101	0.32	0.33	0.33	0.50	0.33	0.33	0.66	0.32	0.34	0.49
Event-8	0.38	0.45	0.39	0.57	0.39	0.41	0.82	0.39	0.42	0.45

Table 2. Purity of clustering by Metric Imitation (MI) when PHOG is used as the TFs, where 50% of the images are used for training and the rest for testing, with recall set to 0.1.

	TFs	MI		SFs	MI		SFs	MI		SFs
	GIST	MI_LLE	MI_Lap	SIFT-llc	MI_LLE	MI_Lap	CNN	MI_LLE	MI_Lap	OB
Scene-15	0.45	0.47	0.47	0.59	0.48	0.48	0.72	0.47	0.47	0.64
CUReT-61	0.72	0.86	0.86	0.90	0.78	0.81	0.95	0.79	0.78	0.90
Caltech-101	0.27	0.34	0.35	0.57	0.33	0.35	0.79	0.33	0.34	0.59
Event-8	0.44	0.53	0.52	0.71	0.51	0.53	0.87	0.52	0.53	0.60

Table 3. MAP of category-based image retrieval by Metric Imitation (MI) when GIST is used as the TFs, where 50% of the images are used for training and the rest for testing, with recall set to 0.1.

	TFs	MI		SFs	MI		SFs	MI		SFs
	PHOG	MI_LLE	MI_Lap	SIFT-llc	MI_LLE	MI_Lap	CNN	MI_LLE	MI_Lap	OB
Scene-15	0.32	0.34	0.35	0.60	0.35	0.36	0.72	0.35	0.36	0.65
CUReT-61	0.66	0.76	0.71	0.90	0.71	0.72	0.95	0.67	0.67	0.90
Caltech-101	0.32	0.35	0.34	0.58	0.34	0.35	0.79	0.35	0.35	0.60
Event-8	0.45	0.45	0.46	0.71	0.44	0.45	0.88	0.43	0.43	0.61

Table 4. MAP of category-based image retrieval by Metric Imitation (MI) when PHOG is used as the TFs, where 50% of the images are used for training and the rest for testing, with recall set to 0.1.

	TFs	MI		SFs	MI		SFs	MI		SFs
	LBP	MI_LLE	MI_Lap	SIFT-llc	MI_LLE	MI_Lap	CNN	MI_LLE	MI_Lap	OB
Holiday	0.30	0.37	0.36	0.67	0.36	0.38	0.74	0.36	0.35	0.51
Ukbench	0.19	0.22	0.22	0.61	0.22	0.23	0.84	0.36	0.38	0.58

Table 5. MAP of instance-based object retrieval by Metric Imitation (MI) on the Holidays and UKbench datasets when GIST is used as the TFs, when the recall is set to 1.

	TFs	MI		SFs	MI		SFs	MI		SFs
	LBP	MI_LLE	MI_Lap	SIFT-llc	MI_LLE	MI_Lap	CNN	MI_LLE	MI_Lap	OB
Holiday	0.34	0.37	0.37	0.68	0.38	0.38	0.73	0.34	0.34	0.48
Ukbench	0.18	0.22	0.20	0.61	0.23	0.21	0.83	0.22	0.22	0.58

Table 6. MAP of instance-based object retrieval by Metric Imitation (MI) on the Holidays and UKbench datasets when PHOG is used as the TFs, when the recall is set to 1.

[6] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. In *Workshop on Generative-Model Based Vision*, 2004. 1

[7] H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*, 2008. 1

[8] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, 2006. 1

[9] L.-J. Li and L. Fei-Fei. What, where and who? classifying event by scene and object recognition. 2007. 1

[10] L.-J. Li, H. Su, E. P. Xing, and F.-F. Li. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In *NIPS*, 2010. 1

[11] D. Nistér and H. Stewénius. Scalable recognition with a vocabulary tree. In *CVPR*, 2006. 1

[12] A. Oliva and A. Torralba. Modeling the shape of the scene:

A holistic representation of the spatial envelope. *IJCV*, 42(3):145–175, 2001. 1

[13] T. Tuytelaars, C. H. Lampert, M. B. Blaschko, and W. Buntine. Unsupervised object discovery: A comparison. *IJCV*, 2009. 1

[14] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *CVPR*, 2010. 1