Metric Imitation by manifold transfer for efficient vision applications

Dengxin Dai, Till Kroeger, Radu Timofte, and Luc Van Gool
Computer Vision Lab, ETH Zurich

The CNN features teach LBP and GIST

Problem

1. Cheap features (e.g. LBP, GIST) are **efficient**, but perform **poorly**
2. Sophisticated features (e.g. IFV, CNN) are **expensive,** but perform **greatly**

Metric Imitation

Better metrics for the cheap features = **Better** cheap features in the framework of Mahalanobis distance

Metric Imitation

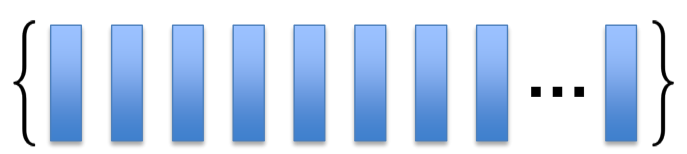
Training

1. Learn the 'intrinsic' manifold W of training data with sophisticated features $\{x_i\}$
2. Learn a (linear) mapping function A for the cheap features $\{y_i\}$ to approximate W

Testing

1. Compute the cheap features $\{y_j\}$ for testing images
2. Obtain the final features $\{\bar{y}_j\}$ by a (linear) mapping: $y \rightarrow \bar{y} = A^T y$

1

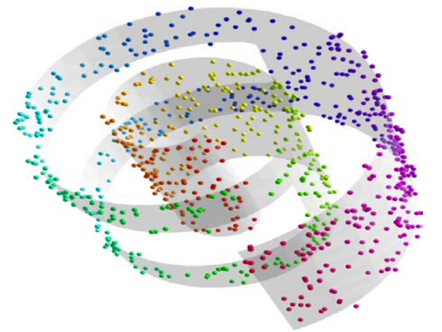


$\{x_i\}$

Sophisticated Features, e.g. CNN

$$\underset{W}{\text{minimize}} \sum_{i=1}^{n} \|x_i - \sum_{j=1}^{n} W_{ij} x_j\|^2 \quad \text{s.t.} \sum_{j=1}^{n} W_{ij} = 1$$

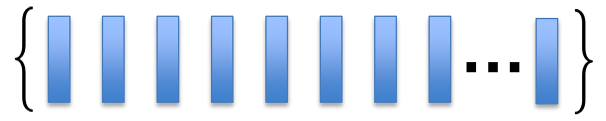
LLE Encoding



'Intrinsic' Manifold *W*

Manifold Transfer

2



$\{y_i\}$

Cheap Features, e.g. LBP

$$\underset{y}{\text{minimize}} \sum_{i=1}^{n} \|\bar{y}_i - \sum_{j=1}^{n} W_{ij} \bar{y}_j\|^2$$

Generalized Eigenvector Problem [1]

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$d(y_i, y_j) = \|\bar{y}_i - \bar{y}_j\|^2 = \|A^T y_i - A^T y_j\|^2$$

Mapping Matrix and the Distance

The **code** and **data** are available at http://people.ee.ethz.ch/~daid/MetricImitation/

Experiments

- Four vision tasks: image clustering, image retrieval, instance-based object retrieval, and super-resolution
- Three sophisticated features: the **CNN features** (4096) [2], **SIFT-LLC** (21504) [3], and **Object Bank** (44604)
- Three cheap features: **GIST** (20), and **PHOG** (40), and **LBP** (59)
- Two types of manifold structures: **LLE** [4] and **LapEigen** [5].

Clustering

TFs = cheap target features; **SFs** = expensive source features

Table 1: Purity of image clustering, where 50% of the images are used for training and the rest for testing.

	TFs	**MI**		SFs	**MI**		SFs	**MI**		SFs
	LBP	MI_LLE	MI_Lap	SIFT-llc	MI_LLE	MI_Lap	CNN	MI_LLE	MI_Lap	OB
Scene-15	0.36	0.40	0.46	0.49	0.47	0.48	0.69	0.42	0.48	0.54
CUReT-61	0.33	0.44	0.46	0.39	0.33	0.41	0.60	0.31	0.37	0.44
Caltech-101	0.32	0.34	0.34	0.51	0.37	0.36	0.68	0.37	0.35	0.52
Event-8	0.39	0.46	0.46	0.57	0.47	0.47	0.82	0.48	0.48	0.46

Table 2: Purity of clustering by Metric Imitation (**MI**) across classes, where half of the classes are used for training and others for testing.

	TFs	**MI**		SFs	**MI**		SFs	**MI**		SFs
	LBP	MI_LLE	MI_Lap	SIFT-llc	MI_LLE	MI_Lap	CNN	MI_LLE	MI_Lap	OB
Scene-15	0.63	0.67	0.70	0.85	0.65	0.66	0.90	0.61	0.59	0.74
CUReT-61	0.62	0.62	0.64	0.65	0.66	0.69	0.77	0.51	0.58	0.68
Caltech-101	0.57	0.62	0.60	0.73	0.59	0.57	0.77	0.64	0.63	0.70
Event8	0.70	0.72	0.74	0.80	0.70	0.72	0.89	0.75	0.73	0.80

Retrieval

TFs = cheap target features; **SFs** = expensive source features

Table 3: MAP of image retrieval with LBP, GIST and PHOG (LGP) as the TFs. 50% images for training and the rest for testing. Recall is set to 0:1.

	TFs	**MI**		SFs	**MI**		SFs	**MI**		SFs
	LGP	MI_LLE	MI_Lap	SIFT-llc	MI_LLE	MI_Lap	CNN	MI_LLE	MI_Lap	OB
Scene-15	0.52	0.60	0.61	0.60	0.64	0.64	0.72	0.62	0.63	0.65
CUReT-61	0.84	0.95	0.93	0.90	0.94	0.96	0.95	0.92	0.90	0.91
Caltech-101	0.42	0.48	0.46	0.57	0.51	0.51	0.79	0.48	0.48	0.59
Event-8	0.52	0.63	0.63	0.70	0.65	0.64	0.88	0.60	0.56	0.58

Table 4: MAP of image retrieval by MI on the Holidays and UKbench datasets, when the recall is set to 1.0.

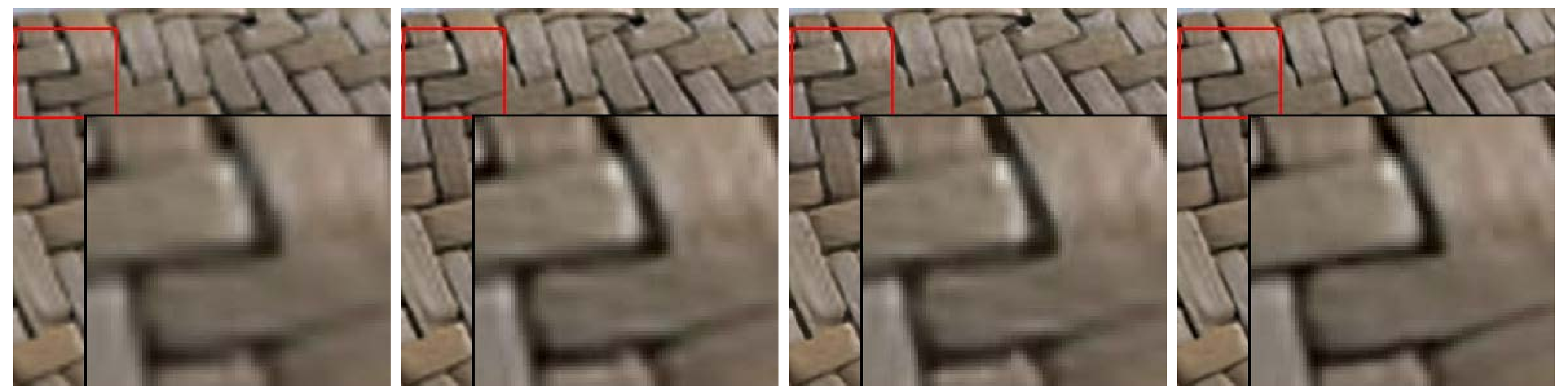
	TFs	MI		SFs	MI		SFs	MI		SFs
	LBP	MI_LLE	MI_Lap	SIFT-llc	MI_LLE	MI_Lap	CNN	MI_LLE	MI_Lap	OB
Holiday	0.38	0.50	0.48	0.66	0.50	0.49	0.72	0.48	0.46	0.48
Ukbench	0.33	0.39	0.38	0.63	0.44	0.39	0.86	0.36	0.38	0.58

Super-resolution

Table 5:Average PSNR on Set5 and Set14.

Benchmark		Bicubic	Zeyde *et al.* [ZEP12]	GR [TDV13]	ANR [TDV13]	NE+LLE [TDV13]	SRCNN [DLHT14]	JOR [DTV15] (5mil)	MI (0.5mil)
Set5	x3	30.39	31.90	31.41	31.92	31.84	32.39	**32.55**	32.53
	x4	28.42	29.69	29.34	29.69	29.61	30.09	**30.19**	30.15
Set14	x3	27.54	28.67	28.31	28.65	28.60	29.00	29.09	**29.10**
	x4	26.00	26.88	26.60	26.85	26.81	27.20	**27.26**	27.25

Bicubic / 25.5 dB ANR / 26.9 dB SRCNN / 27.1 dB MI / 27.7 dB



Reference

[1] X. He, D. Cai, S. Yan, and H.-J. Zhang. Neighborhood preserving embedding. In ICCV, 2005.
[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.
[3] J.Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In CVPR, 2010.
[4] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500):2323–2326, 2000.
[5] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In NIPS, 2001.

European Research Council