Beyond Photo-Domain Object Recognition: Benchmarks for the Cross-Depiction Problem

Hongping Cai, Qi Wu and Peter Hall

Department of Computer Science, University of Bath, UK

Introduction

Humans are able to recognise objects in an astonishing variety of forms. The same is not true of computers. We address an under-researched area – **Cross-Depiction Problem** – recognising visual objects regardless of whether they are photographed, painted, drawn, etc.

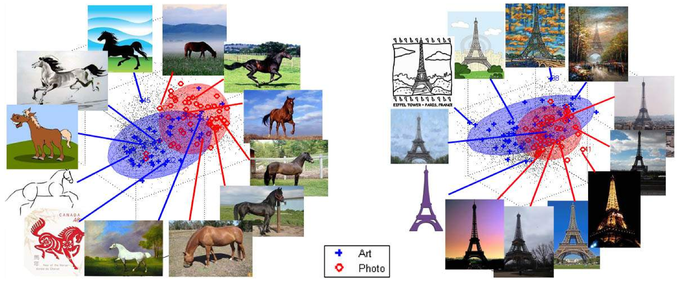
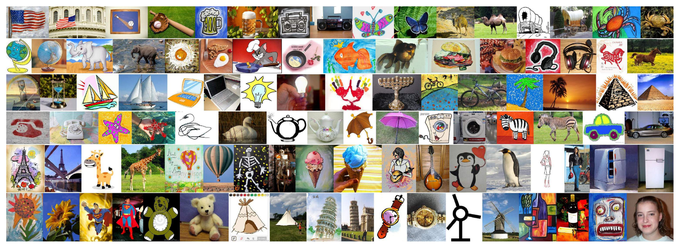


Figure: Example images and their distribution in the BoW-SIFT feature space. The features are generated by projecting the 5000-d BoW-SIFT features to 3-d space using PCA. Gray clouds represent all categories in the *Photo-Art-50* dataset.

Dataset: *Photo-Art-50* [1]

- 50 objects
- 90 to 138 images for each object with approximately half photos and half art images.



Divergence of Cross-depiction Data

Cross-domain datasets [2,4]					Photo-Art-50 [1]
C-A	C-D	A-W	D-A	D-W	Photo-Art
0.079	0.271	0.239	0.292	0.047	**0.466**

Table: Comparison of K-L divergence between domain pairs. C - Caltech-256, A - Amazon, W - WebCam, D - DSLR.

- The intuition is confirmed that the variance across photo and art domains is much larger than that in conventional domain adaptation research, which introduce more challenges.

Classification Benchmarks

model		BoW					FV	CNN
train	test	SIFT	GB	SSD	HOG	edgeHOG	SIFT	Pre-trained
Photo	Photo	83.69 ± 0.6	76.83±1.4	66.48±1.3	72.40±0.8	70.04±1.0	**87.42±0.5**	96.95±0.3
A+P	Photo	80.38 ± 1.1	71.94±1.1	57.85±0.9	64.67±1.4	63.25±1.3	**83.53±0.7**	96.23±0.5
Art	Photo	63.93 ± 1.1	59.90±0.8	38.89±1.6	42.45±1.1	50.13±1.4	**65.67±0.5**	90.50±0.7
Art	Art	74.25 ± 1.1	72.05±1.4	49.03±1.4	55.13±0.6	59.55±0.6	**76.74±0.5**	89.24±0.5
A+P	Art	69.47 ± 1.1	67.08±0.6	45.27±2.1	49.87±1.0	56.07±2.0	**72.82±1.0**	87.13±1.2
Photo	Art	43.78 ± 0.6	**50.42±1.4**	31.16±1.0	28.99±1.4	39.91±1.6	47.35±1.2	72.54±1.3

Table: Categorisation performance on the Photo-Art-50, with 30 images per category for training. Average correct rates are reported by running 5 rounds with random training-test split. 'A+P' stands for a mixture training set of 15 photo images and 15 art images.

- Both shallow and deep representations share the same trend: All methods show a significant drop when trained on one depiction style and tested on another.
- The most difficult one is the 'train-on-photo-test-on-art' setting. It can be explained by the degree of variation in the features as evidenced by the K-L divergence.

Domain Adaptive Benchmarks

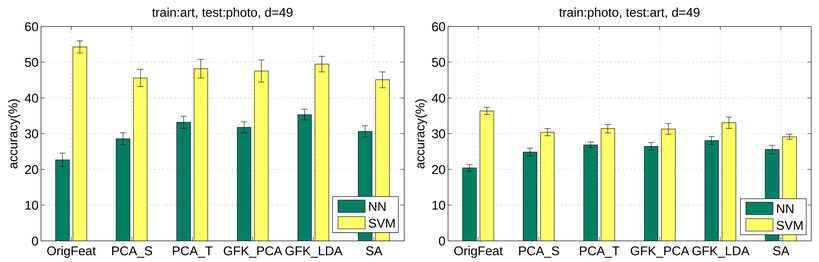


Figure: Classification accuracies without (OrigFeat, PCA S and PCA T) and with (GFK PCA, GFK LDA, SA) domain adaptive methods on Photo-Art-50. 'OrigFeat' means classifying with the original 5000-bin BoW-SIFT histograms. 'pca s' and 'pca t' denote PCA on the source domain and target domain, respectively. Except OrigFeat, the rest methods are with 49-d projected features.

- Different from the effectiveness in conventional domain adaptive problem, GFK [2] and SA [3] fail in the cross-depiction problem.

Reference

[1] Q. Wu, H. Cai, and P. Hall. Learning graphs to model visual objects across different depictive styles. In ECCV, 2014.

[2] B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In CVPR, pages 2066-2073, 2012.

[3] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In ICCV, 2013.

[4] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In ECCV, pages 213-226, 2010.

DPM with Cross-Depiction Expansion (DPM-CDE)

We borrow the idea of query expansion for cross-depiction detection.

- **STEP1:** Train the DPM model for each object class in the source domain $\mathcal{S}$.
- **STEP2:** Apply the models on the target domain $\mathcal{T}$. A confidence set $\mathcal{T}_{ex} \subset \mathcal{T}$ is picked from the target domain for training expansion.
- **STEP3:** Re-learn the DPM model on the expanded training set $\mathcal{S} \cup \mathcal{T}_{ex}$. Then this adapted DPM model, named **DPM-CDE**, is used in the detection task.

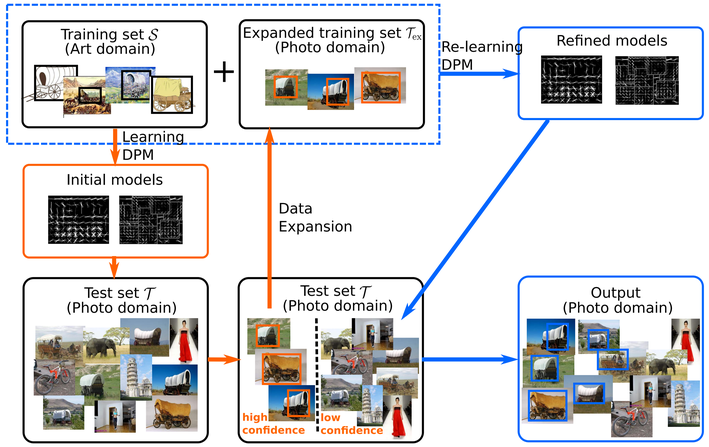


Figure: The pipeline of DPM with cross-depiction expansion.

train	test	DPM	DPM-CDE
Photo	Photo	0.957	–
Art	Photo	0.798	**0.843**
Art	Art	0.839	–
Photo	Art	0.727	**0.783**

Table: Mean average precision (mAP) on Photo-Art-50, 30 images per object for training.

- DPM-CDE provide clear performance benefits, which demonstrated that the expanded set does help to refine the models according to the target domain.

Summary

- We confirmed the intuition by experiment that the variance across photo and art domains is much larger than the conventional cross domain problem.
- We benchmarked leading recognition (both shallow and deep representations) and detection methods, state-of-the-art domain adaptive methods for cross-depiction task, showing none perform well.
- Inspired by query expansion, we adapted DPM on an expanded training set. The performance gain implies that the cross-depiction expansion is a simple but effective way of bridging the gap between photo and art domains.

Email: h.cai@bath.ac.uk

Homepage: http://people.bath.ac.uk/hc551