Attributes and Categories for Generic Instance Search from One Example

Ran Tao[1], Arnold W.M. Smeulders[1], Shih-Fu Chang[2]
[1]ISLA, Informatics Institute, University of Amsterdam. [2]Department of Electrical Engineering, Columbia University.

This paper aims for generic instance search from 1 query example. *Generic* implies we consider one algorithm for arbitrary instances ranging from near-planar and one-sided objects with limited viewpoint variations (*e.g.*, buildings and logos) to 3D objects recorded from all possible imaging angles of the viewing sphere (*e.g.*, shoes and cars). A very hard case in this respect is a query specified in the front view while the relevant images in the search set show a view from the back which has never been seen before. Humans solve the search task by employing two types of general knowledge. First, when the query instance is a certain class, say *female*, answers should be restricted to be from that class. And, queries in the frontal view showing one attribute, say *brown hair*, will limit answers to show the same or no such attribute, even when the viewpoint is from the back. In this paper, we employ these two types of general knowledge to address generic instance search.

Good performance has been achieved in instance search of near-planar and one-sided objects like buildings and logos [1, 2, 5]. Yet, it is unclear how the state-of-the-art approaches perform on generic instance search: *Can we search for other objects like shoes with the same method for buildings?* To answer that, as the first contribution of the paper, we evaluate four methods, *ExpVLAD* [5], *Triemb* [1], *Fisher* [2] and *Deep* [3], on both buildings (*Oxford5k* [4]) and shoes (*CleanShoes*, with 6624 images of 1000 different shoes). Figure 1 shows examples. We observe that what works best for buildings loses its generality for shoes and reverse. None of the methods work well on both buildings (*i.e.*, near-planar instance search) and shoes (*i.e.*, 3D full-view instance search).

Figure 1: Examples of buildings in *Oxford5k* and shoes in *CleanShoes*. Shoes show a much wider range of viewpoint variability.

As the second contribution, we propose to use automatically learned category-specific attributes to address the large appearance variations in generic instance search. Provided with a set of training instances from a certain category, *e.g.*, *shoes*, we learn a set of attribute classifiers using the method of [6] and perform instance search on new instances from the same category using attribute representation. In this way, the representation is robust against the appearance variations of an instance, more than low-level features such as Fisher vector, while being discriminative among instances from the same category.

We evaluate the effectiveness of the learned category-specific attributes on the problem of searching among instances from the same category as the query. Three datasets are considered, *CleanShoes* for *shoes*, *Cars* for *cars* and *OxfordPure* for *buildings*. *Cars* contains 1100 images of 270 cars. *OxfordPure* is composed by gathering the 567 images of the 55 Oxford landmarks from *Oxford5k*. The attributes are learned on 2100 images of 300 shoes, 1520 images of 300 cars and 8756 images of 300 buildings respectively. The instances in the evaluation sets are not present in the training sets.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

method↓	*dim*	CleanShoes	Cars	OxfordPure
ExpVLAD	—	16.14	23.70	**87.01**
Triemb	8064	25.06	18.56	75.33
Fisher	16384	20.94	18.37	70.81
Deep	4096	36.73	22.36	59.48
Attributes	1000	**56.56**	**51.11**	77.36

Table 1: Performance of learned attributes and existing methods for generic instance search from one example, *ExpVLAD* [5], *Triemb* [1], *Fisher* [2] and *Deep* [3]. The learned category-specific attributes achieve much better performance than others on *shoes* and *cars*, and are on par with others on *buildings*.

Table 1 shows the results. The attribute representation works significantly better than the others on the *shoe* and *car* datasets. Attributes are superior in addressing the large appearance variations caused mainly by the large imaging angle difference present in the *shoe* and *car* images, even though the attributes are learned from other instances. The attribute representation also works well on *buildings* while having a much lower dimensionality. The proposed method using automatically learned category-specific attributes is more generic than other approaches.

As the third contribution, we extend our method to search objects without restricting to the known category. The category-specific attributes are optimized to make distinctions among instances of the same category where the distinction from the instances of other categories is another hurdle. In order to address the confusion of the query instance with instances from other categories, we first use the category-level information. We consider two types of category-level information, the deep learning features learned from large-scale image categorization and the category-level classification scores. We show that the proposed method of combining the category-level information with the category-specific attributes is effective, outperforming the combination of the category-level information with Fisher vector.

Acknowledgments This research is supported by the Dutch national program COMMIT/.

[1] Hervé Jégou and Andrew Zisserman. Triangulation embedding and democratic aggregation for image search. In *CVPR*, 2014.

[2] Hervé Jégou, Florent Perronnin, Matthijs Douze, Jorge Sánchez, Patrick Pérez, and Cordelia Schmid. Aggregating local image descriptors into compact codes. *TPAMI*, 34(9):1704–1716, 2012.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[4] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*, 2007.

[5] Ran Tao, Efstratios Gavves, Cees G M Snoek, and Arnold W M Smeulders. Locality in generic instance search from one example. In *CVPR*, 2014.

[6] Felix X Yu, Liangliang Cao, Rogerio S Feris, John R Smith, and Shih-Fu Chang. Designing category-level attributes for discriminative visual recognition. In *CVPR*, 2013.