
# Automatic Image Annotation via Label Transfer in the Semantic Space

Tiberio Uricchio, Lamberto Ballan, Lorenzo Seidenari, Alberto Del Bimbo

PII: S0031-3203(17)30206-6
DOI: 10.1016/j.patcog.2017.05.019
Reference: PR 6158

To appear in: *Pattern Recognition*

Received date: 15 August 2016
Revised date: 15 April 2017
Accepted date: 20 May 2017

Please cite this article as: Tiberio Uricchio, Lamberto Ballan, Lorenzo Seidenari, Alberto Del Bimbo, Automatic Image Annotation via Label Transfer in the Semantic Space, *Pattern Recognition* (2017), doi: 10.1016/j.patcog.2017.05.019

# Automatic Image Annotation via Label Transfer in the Semantic Space

Tiberio Uricchio[a], Lamberto Ballan[b,*], Lorenzo Seidenari[a], Alberto Del Bimbo[a]

[a]*Media Integration and Communication Center (MICC), Università degli Studi di Firenze, Viale Morgagni 65, 50134 Firenze, Italy*
[b]*Department of Mathematics "Tullio Levi-Civita", Università degli Studi di Padova, Via Trieste 63, 35121 Padova, Italy*

**Abstract**

Automatic image annotation is among the fundamental problems in computer vision and pattern recognition, and it is becoming increasingly important in order to develop algorithms that are able to search and browse large-scale image collections. In this paper, we propose a label propagation framework based on Kernel Canonical Correlation Analysis (KCCA), which builds a latent *semantic space* where correlation of visual and textual features are well preserved into a semantic embedding. The proposed approach is robust and can work either when the training set is well annotated by experts, as well as when it is noisy such as in the case of user-generated tags in social media. We report extensive results on four popular datasets. Our results show that our KCCA-based framework can be applied to several state-of-the-art label transfer methods to obtain significant improvements. Our approach works even with the noisy tags of social users, provided that appropriate denoising is performed. Experiments on a large scale setting show that our method can provide some benefits even when the semantic space is estimated on a subset of training images.

*Keywords:* Automatic image annotation, Image tagging, Label transfer, Canonical correlation, Semantic space

## 1. Introduction

A lot of modern applications require image annotation to search, access and navigate the huge amount of visual data stored in personal collections or shared online. Whenever you want to retrieve photos from a particular concert, recall that pleasant summer day in which you napped on your comfortable hammock or look up a person, it is automatic image annotation that enables a plethora of useful applications. The exponential growth of media on sharing platforms, such as Flickr or Facebook, has led to the availability of a huge quantity of images that are enjoyed by millions of people. In such a huge sea of data, it is indispensable to teach computers to correctly label the visual content and help us search and browse image collections.

In this paper, we tackle the challenging task of automatic image annotation. Given an image, we want to assign a set of relevant labels by taking into account image appearance and eventually some prior knowledge on the joint distribution of visual features and labels. Due to its importance, this is a very active subject of research [1, 2, 3, 4, 5, 6, 7, 8]. Previous work typically use images and associated labels to build classifiers and then assign relevant labels to novel images. The early works usually rely on images labeled by domain experts [9, 2, 3, 10, 11], while recently several approaches use weak labels such as user-generated tags in social networks [12, 13, 14] or query terms in search engines [15, 16].

Despite the source of the labeling, non-parametric models which rely on a nearest-neighbor based voting scheme have received a lot of attention for automatic image annotation [17, 10, 18, 19, 20]. The main reason is that these methods have the ability to adapt to complex patterns as more training data become available. To annotate a new image, they apply a common strategy: first, they retrieve similar images in the training set, and second, they rank labels according to their frequency in the retrieval set. Automatic image annotation is thus achieved by transferring the most frequent labels in the neighborhood to the test image. This is essentially a lazy learning paradigm in which the image-to-label association is delayed at test time. In contrast, discriminative models such as support vector machines [21, 22, 23, 24] or fully supervised end-to-end deep networks [8], require to define in advance the vocabulary of labels. This is particularly problematic in a large-scale scenario, such as images on social networks, in which you may have thousands of labels that may also change or increase over time.

Several issues may arise in a nearest-neighbor approach. The set of retrieved images may contain many incorrect labels, mostly because of the so-called *semantic gap* [25]. This happens because visual features may not be powerful enough in abstracting the visual content of the image. Thus the proposed algorithms tend to retrieve just the images whose features are very close in the visual space, but the semantic content is not well preserved. Researchers tried to cope with this issue by improving visual features. To this end, the most significant improvement has been the shift from handcrafting features to end-to-end feature learning, leading to current state-of-the art convolutional neural network representations [26, 27, 28]. Nearest neighbors methods may also suffer when images are not paired with enough label information, leading to a poor statistical quality of the retrieved neighborhood. This is mostly due to the fact

*Corresponding author. A major part of this work has been done while the author was on an EU Marie Curie Fellowship at Stanford University and Univ. of Florence.

*Email addresses:* tiberio.uricchio@unifi.it (Tiberio Uricchio), lamberto.ballan@unipd.it (Lamberto Ballan), lorenzo.seidenari@unifi.it (Lorenzo Seidenari), alberto.delbimbo@unifi.it (Alberto Del Bimbo)