DOG_I: an Annotation System for Images of Dog Breeds

Antonios Dimas, Pyrros Koletsis, and Euripides G.M. Petrakis*

Department of Electronic and Computer Engineering,
Technical University of Crete (TUC) Chania, Crete, GR-73100, Greece
{adimas@isc.tuc.gr,{pyrros,euripides}@intelligence.tuc.gr}
http://www.intelligence.tuc.gr

Abstract. DOG_I (Dog OntoloGy Image annotator) is a complete and fully automated semantic annotation system for images of dog breeds. Annotation relies on feature extraction and on associating low-level features with image concepts in an ontology. Because general purpose ontologies for all image types are not yet available, we choose the problem of annotating images of dog breeds as a case study and for the evaluation of our methodology. Nevertheless, DOG_I can be adapted to more image types provided that an ontology for a new image domain becomes available. Therefore, DOG_I offers an ideal test-bed for experimentation and sets the grounds for the annotation and evaluation of virtually any image type. Evaluation results are realized using images collected from the Web. Almost 95% of the test images is correctly annotated (i.e., DOG_I identified their class correctly). $DOG_I$is accessible on the Internet.

Keywords: image annotation, ontology, image similarity, retrieval, machine learning

1 Introduction

Image annotation (or tagging) is the task of assigning a class (i.e., a label) or description to an unknown (query) image [1]. Typically, image annotations are compact consisting of a few meaningful words of phrases summarizing image contents. Annotations can be assigned to images manually or can be extracted automatically by computers. Although humans tend to provide more comprehensive image descriptions than computers can do, the quality of annotations is questionable due to the specificity of image content and subjectivity of image content interpretations. In addition, the process is slow and costly and, does not scale-up easily for the entire range of image types or for large data collections, such as the Web [2]. Overcoming problems of subjectivity, cost and scalability calls for automatic image annotation. Automatic annotation by computers relies on automatic feature extraction from images (e.g., color, texture measurements

* This work was partially supported by project RT3S (FP7-STREP, Project No 248801) of the European Union.

etc.) and on associating image features with high level concepts in an ontology [3], to a set of keywords [4] or, to any combination of concepts, keywords and sentences [5]. A popular approach relates to extracting image annotations from text. This approach is particularly useful in multimedia collections or the Web, where images co-exist with text. For example, images on the Web are described by surrounding text or attributes associated with images in `html` tags (e.g., filename, caption, alternate text etc.). Google Image Search[1] is an example system of this category.

Automatic annotations can be fast and cheap but, compared to annotations by humans, they can be less accurate, as general purpose image analysis algorithms for extracting meaningful and reliable descriptions for all image types are not yet available. An additional problem relates to imprecise mapping of image features to high level concepts, typically referred to as the"semantic gap problem [6]. To handle issues of domain dependence, diversity of image content and achieve high quality results, automatic image annotation methods need to be geared towards specific image types.

Li and Wang [4] show that a computer can learn (from a large set of example images) to annotate images in real-time. They apply statistical modeling and optimization methods for establishing probabilistic relations between images and keywords. Schreiber et.al. [7] introduced an animal ontology providing a description template for annotation. Their solution is not fully automatic, it is in fact a tool for assisting manual annotations to images of animals and aims primarily at alleviating the burden of human annotators. Park et. al. [8] use MPEG-7 visual descriptors in conjunction with domain ontologies. Annotation in this case is achieved using inference rules. Along the same lines, Mezaris et.al. [9] focus on object ontologies (i.e., ontologies defined for image regions or objects). Visual features of segmented regions are mapped to human-readable descriptor values (e.g., "small", "black" etc.). Lacking semantics, the above descriptors can't be easily associated with high-level ontology concepts. Also, the performance of the method is constraint by the performance of image segmentation. SIA [5] follows the ontology approach too: high-level concepts (dog breeds) together with low-level image features are organized in a generalization (IS_A) hierarchy and are associated with concept descriptions, natural language (text) descriptions, and low-level image features.

DOG_I builds-upon our previous work for SIA [5] where the core ideas are discussed. It is a fully functional image annotation system which, in addition, extends SIA methodology in the following ways: DOG_I incorporates a more elaborate ontology of 40 dog breeds (30 in SIA) which is enhanced with more image information and textual descriptors. All possible associations between classes and associations between image classes and class properties (e.g., part-of, functional associations) are depicted. To deal with the diversity of image content on the Web and handle uncertainty due to variations in lighting, noise, scaling and posing, each class is represented by a sufficient number (i.e., 9) of image instances (6 in SIA).

[1] http://www.google.com/imghp

Class descriptions are complemented by an enhanced set of low-level features extracted from the image instances of the class. Images of dog breeds are mainly characterized by the spatial distribution of color intensities. These representations are computed in terms of descriptors of color and texture (such as those implemented in LIRE[2]) and may be augmented with more descriptors (e.g., features of shape and image structure will be added in future implementations). Feature selection for computing the relative importance of features is implemented by machine learning. Nevertheless, the focus of this work is not on novel image feature extraction but on showing how to enhance the accuracy of automatic annotation for a given and well established set of image features.

Annotation is modeled as an image retrieval process: the unknown (query) image is compared with images in the ontology (whose class descriptions are known). DOG_I incorporates more elaborate criteria for computing the likelihood of the query image to belong to each one of the candidate classes yielding an improvement of the overall annotation accuracy from 89% in SIA, to 95%. Evaluation results of the method are realized on images of 40 dog breeds (30 in SIA) collected from the Web. Finally, DOG_I annotations are stored in the `exif`[3] meta-data tag of the annotated image. Overall, DOG_I is a fully automatic image annotation system for images of dog breeds and is available on the Web[4].

DOG_I is discussed in detail in Sec. 2. The discussion includes DOG_I resources and processes in detail, namely the ontology, image analysis, image similarity and image annotation. Evaluation results are presented in Sec. 3 and the work is concluded in Sec. 4.

2 DOG_I System

DOG_I provides a graphical user interface for loading and displaying images and their properties, for loading and viewing ontologies and annotation results. The user can also select annotation method (a feature useful for evaluating the performance of the 4 competing methods implemented and discussed in Sec. 2.3) and subsequently, save the result in MPEG-7 format in the images' `exif` meta-data header. Fig. 1 illustrates DOG_I interface.

Because an image may contain several regions, a user may select a region of interest (e.g., by manually dragging a rectangle around the region surrounding the dog's head) or, let the system apply feature extraction and annotation on the entire image.

2.1 DOG_I Ontology

Information about 40 classes of dog breeds is organized in an ontology [5,10]. The ontology consists of an IS_A (generalization) hierarchy of dog breeds together with the "traits hierarchy" holding properties of dog breeds (e.g., coat color, fur,

[2] http://www.semanticmetadata.net/lire/
[3] http://www.exif.org
[4] http://www.intelligence.tuc.gr/prototypes.php

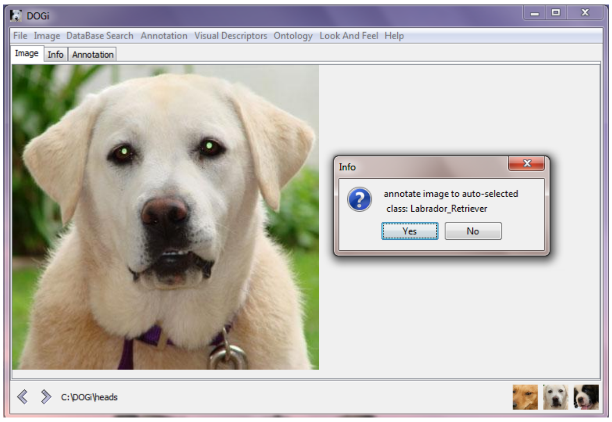


Fig. 1: DOG_I Graphical User Interface.

size, personality, country of origin). The traits hierarchy is also enriched with textual information (from Wordnet[5] and Wikipedia[6]). The leaf classes in the hierarchy correspond to the different semantic categories of the ontology (i.e., 40 dog breeds). To deal with the diversity of image content on the Web and handle uncertainty due to variations in lighting, noise, scaling and posing, each class is represented by 9 image instances (i.e., 6 in SIA). Therefore, the ontology stores 360 images total (i.e., 40 classes with 9 instances each) together with their representations.

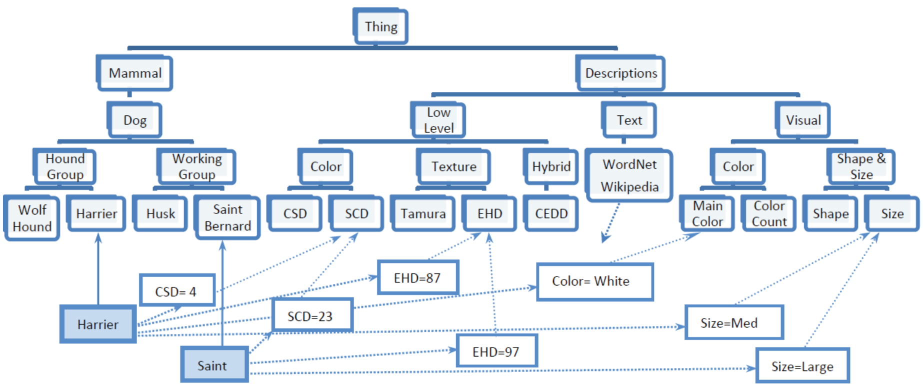


Fig. 2: DOG_I Ontology.

The image content representation of a class comprises of the representation of each of its 9 instances. In turn, this is computed as a vector of 12 image descriptors (7 in SIA). DOG_I incorporates the following set of 10 LIRE[7] features with the addition of Color Structure Descriptor (CSD) and Fuzzy-Color Tex-

[5] http://wordnet.princeton.edu/
[6] http://www.wikipedia.org/
[7] http://www.semanticmetadata.net/lire/

ture Histogram Descriptor (FCTHD): Color Layout Descriptor (CLD), Scalable Color Descriptor (SCD), Auto-Corellogram Descriptor (ACD), Fuzzy-Color Histogram Descriptor (FCHD), Simple-Color Histogram Descriptor (SCHD), Gabor Descriptor (GD), Tamura Descriptor (TD), Edge Histogram Descriptor (EHD), Color-Edge Directivity Descriptor (CEDD) and Joint CEDD-FCTH Descriptor(JD).

Object properties connect instances of the leaf classes with features in the traits hierarchy. Each class is characterized by it's own property values (e.g., a border collie has country of origin "Scotland", main color "black and white") and, in addition, each one of its 9 instances is associated with the values of its 12 image features in the traits hierarchy. The ontology is constructed using Protégé [8] in OWL. Protégé OWL-API was used for implementing the links between OWL properties and the image instances of a class. Fig. 2 illustrates a snapshot of the ontology with two instances (i.e., "saint-bernard" and "harrier"). Not all classes and image properties are shown.

2.2 Image Annotation Methodology

Image annotation is modeled as a retrieval process. The query image is compared with the 360 images in the ontology. The output consists of the same 360 images ordered by similarity. Image similarity between a query (Q) and an image (I) is computed as a weighted sum of similarities between their corresponding feature vectors

$$D(Q, I) = \sum_{i=1}^{12} w_i (1 - d_i(Q, I)), \tag{1}$$

where i indexes features from 1 through 12, $d_i(Q, I)$ is the distance between the two images for feature i and w_i represents the relative importance of feature i. All distances $d_i(A, B)$ are normalized in $[0, 1]$ by Gaussian normalization. The advantage of Gaussian normalization is that the presence of a few large or small values does not bias the importance of a feature in computing the similarity.

Not all features are equally important. Appropriate weights for all features are computed by a decision tree: the training set consists of 3,474 image pairs collected from the Web (1584 pairs of similar images and 1890 pairs of dissimilar images). For each image pair, a 12-dimensional feature vector is formed. The attributes of this vector are computed as the Gaussian normalized feature distances. The decision tree accepts pairs of images and classifies them into similar or not. The decision tree was pruned with confidence value 0.25 and achieved 85.15% classification accuracy. Notice that, weights can take 0 values (i.e., their respective features are ignored). The evaluation method is stratified cross validation. Appropriate weights are computed from the decision tree as follows:

$$w_i = \sum_{nodes\ of\ feature\ i} \frac{maxdepth + 1 - depth(feature_i)}{\sum_{j=1}^{all\ nodes} maxdepth + 1 - depth(node_j)}, \tag{2}$$

[8] http://protege.stanford.edu/

where i indexes features from 1 through 12, j indexes tree nodes ($node_j$ is the j-th node of the decision tree), $depth(feature_i)$ is the depth of feature i and $maxdepth$ is the maximum depth of the decision tree. The summation is taken over all nodes of feature i (there may exist more than one node with feature i in the tree). This formula suggests that the higher a feature is in the decision tree and the more frequently it appears, the higher its weight will be.

2.3 Image Annotation Criteria

The input image is compared with the 360 ontology images (40 classes with 9 instances each) by applying Eq. 1. The answer is sorted by decreasing similarity. The class description of the input image can be computed by any of the following methods (the first 4 of them are also from [5]). The query image inherits the properties of its classes higher in the hierarchy.

Best Match: Selects the class of the most similar instance.

Max Occurrence: Selects the class that has the maximum number of instances in the first n answers (in this work n is set to 20). If more than one classes have the same number of instances within the first n answers, then Best Match is applied to choose between them.

Average Retrieval Rank (AVR): Selects the description of the class with the highest AVR. Assuming that there are $NG(q)$ images having the same class with the input image q (i.e., 9 in this work) and $rank(i)$ is the rank of the i-th ground truth image in the results list, AVR is computed as:

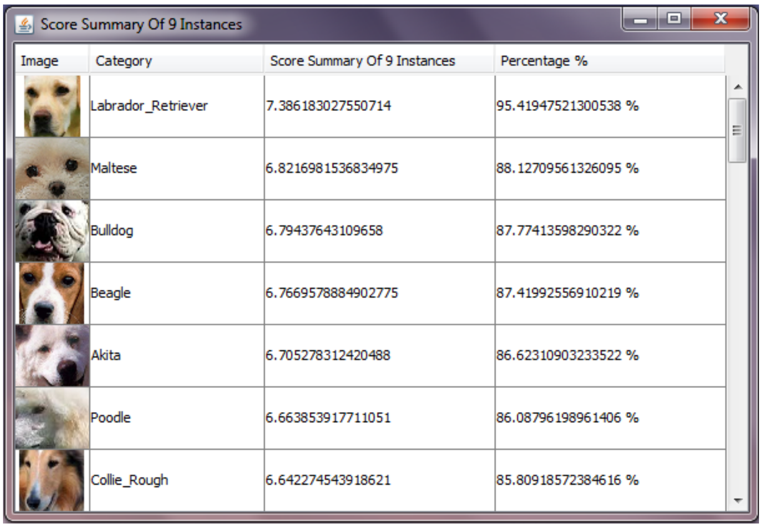
$$AVR(q) = \sum_{i=1}^{NG(q)} \frac{rank(i)}{NG(q)} \tag{3}$$

Max Similarity: The query is compared with all the $NG(q)$ (i.e., 9) image instances of each category and their similarities are added. The method selects the category with the maximum similarity score.

3 Experimental Evaluation

The purpose of the following experiment is to demonstrate the annotation efficiency of DOG_I. For the evaluation, 40 test (unknown) images are used as queries. Fig. 3 illustrates that the Max Similarity method has recognized the class of the query image of Fig. 1 correctly with confidence 95.41%. Fig. 4 illustrates the annotation assigned to the query image.

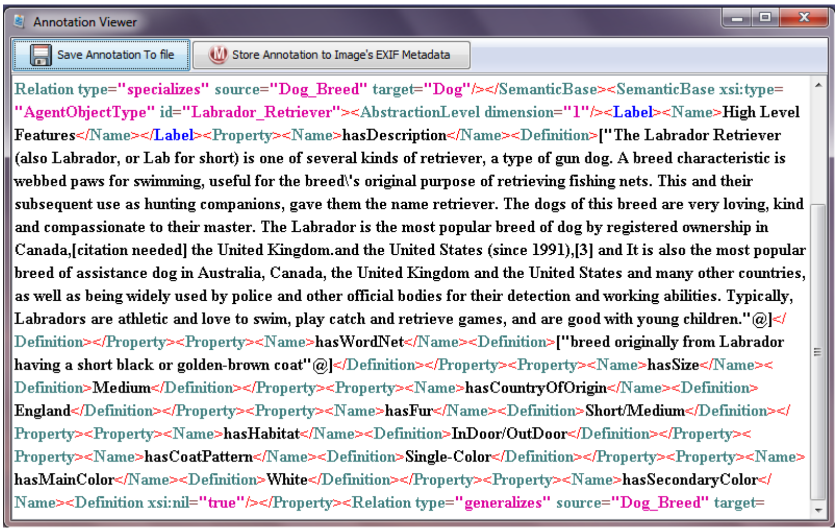
Table 1 illustrates the accuracy of the Best Match, Max. Occurrence, AVR and Maximum Similarity methods of Sec. 2.3. All measurements are average over 40 test images. The image ranked first has always higher probability of being annotated correct. However, there are cases where the correct annotation is provided by the image ranked second or third. Maximum Similarity outperforms all other methods followed by AVR (the best method in [5]): the image ranked



Score Summary Of 9 Instances

Image	Category	Score Summary Of 9 Instances	Percentage %
	Labrador_Retriever	7.386183027550714	95.41947521300538 %
	Maltese	6.8216981536834975	88.12709561326095 %
	Bulldog	6.79437643109658	87.77413598290322 %
	Beagle	6.7669578884902775	87.41992556910219 %
	Akita	6.705278312420488	86.62310903233522 %
	Poodle	6.663853917711051	86.08796198961406 %
	Collie_Rough	6.642274543918621	85.80918572384616 %

Fig. 3: Annotation results for the query image of Fig. 1.



Relation type="specializes" source="Dog_Breed" target="Dog"/></SemanticBase><SemanticBase xsi:type="AgentObjectType" id="Labrador_Retriever"><AbstractionLevel dimension="1"/><Label><Name>High Level Features</Name></Label><Property><Name>hasDescription</Name><Definition>["The Labrador Retriever (also Labrador, or Lab for short) is one of several kinds of retriever, a type of gun dog. A breed characteristic is webbed paws for swimming, useful for the breed\'s original purpose of retrieving fishing nets. This and their subsequent use as hunting companions, gave them the name retriever. The dogs of this breed are very loving, kind and compassionate to their master. The Labrador is the most popular breed of dog by registered ownership in Canada,[citation needed] the United Kingdom.and the United States (since 1991),[3] and It is also the most popular breed of assistance dog in Australia, Canada, the United Kingdom and the United States and many other countries, as well as being widely used by police and other official bodies for their detection and working abilities. Typically, Labradors are athletic and love to swim, play catch and retrieve games, and are good with young children."@]</Definition></Property><Property><Name>hasWordNet</Name><Definition>["breed originally from Labrador having a short black or golden-brown coat"@]</Definition></Property><Property><Name>hasSize</Name><Definition>Medium</Definition></Property><Property><Name>hasCountryOfOrigin</Name><Definition>England</Definition></Property><Property><Name>hasFur</Name><Definition>Short/Medium</Definition></Property><Property><Name>hasHabitat</Name><Definition>InDoor/OutDoor</Definition></Property><Property><Name>hasCoatPattern</Name><Definition>Single-Color</Definition></Property><Property><Name>hasMainColor</Name><Definition>White</Definition></Property><Property><Name>hasSecondaryColor</Name><Definition xsi:nil="true"/></Property><Relation type="generalizes" source="Dog_Breed" target=

Fig. 4: Annotation assigned to the example image of Fig. 1.

first is correctly annotated in 72.5% of the images tested. Overall, the correct annotation is provided by any of the top 3 ranked images in 95% of the images tested.

4 Conclusion

We introduce the *DOG_I* annotation framework for images of dog breeds. It is easily extendible to images of more categories provided that an ontology of such

Annotation Result	Max Similarity	AVR	Ma. Occurrence	Best Match
Ranked 1^{st}	72.5%	62.5 %	65.0%	50.0%
Ranked 2^{nd}	17.5%	22.5 %	15.0%	10%
Ranked 3^{rd}	5.0 %	10.0 %	10.0%	10%
Overall: 1-3 answers	95.0 %	92.5 %	90.0 %	90%

Table 1: Performance of Annotation.

images is given. This is an interesting subject for future research. Enhancing the ontology with features of shape and image structure are also issues for future work. DOG_I demonstrated certain performance improvements over our previous work [5] for more (i.e., 40) image categories (30 in SIA). Additional experiments [10] demonstrate that the performance of DOG_I and the number of image categories are clearly traded-off, implying that the performance of annotation can be improved further by reducing the number of image categories from 40 (in this work) to 20 or 10.

References

1. Hanbury, A.: A Survey of Methods for Image Annotation. Journal of Visual Languages and Computing **19**(5) (Oct. 2008) 617–627
2. Yan, R., Natsev, A., Campbell, M.: Hybrid Tagging and Browsing Approaches for Efficient Manual Image Annotation. IEEE MultiMedia **16** (April 2009) 26–41
3. A.M. Tousch and S. Herbin and J.-Y. Audibert: Semantic Hierarchies for Image Annotation: A Survey. Pattern Recognition **45** (Jan. 2012) 333–345
4. Li, J., Wang, J.Z.: Real-Time Computerized Annotation of Pictures. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(6) (June 2008) 985–1002
5. Koletsis, P., E.G.M.Petrakis: SIA: Semantic Image Annotation Using Ontologies and Image Content Analysis. In: Intern. Conf. on Image Analysis and Recognition (ICIAR). Volume 6111 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2010) 374–383
6. Smeulders, A.W., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-Based Image Retrieval at the End of the Early Years. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(11) (Dec. 2000) 1349–1380
7. Schreiber, A., Dubbeldam, B., Wielemaker, J., Wielinga, B.: Ontology-Based Photo Annotation. IEEE Intelligent Systems **16**(3) (May 2001) 66–74
8. Park, K.W., Jeong, J.W., Lee, D.H.: OLYBIA: Ontology-Based Automatic Image Annotation System Using Semantic Inference Rules. In: Advances in Databases: Concepts, Systems and Applications. Volume 4443. Springer Berlin / Heidelberg (2007) 485–496
9. Mezaris, V., Kompatsiaris, J., MStrintzis: Region-Based Image Retrieval using an Object Ontology and Relevance Feedback. EURASIP Journal on Applied Signal Processing **2004**(1) (Jan. 2004) 886–901
10. Dimas, A.: DOG_I: An Automatic Image Annotation Tool for Images of Dog Breeds. diploma thesis, Dept. of Electronic and Computer Engineering, TUC, Chania, Crete, Greece (July 2011)