
Proceedings Paper:
Wang, J.K. orcid.org/0000-0003-0048-3893, Markert, K. and Everingham, M. (2016) Harvesting Training Images for Fine-Grained Object Categories using Visual Descriptions. In: Ferro, N., Crestani, F., Moens, M-F., Mothe, J., Silvestri, F., Di Nunzio, G.M., Hauff, C. and Silvello, G., (eds.) Advances in Information Retrieval. European Conference on Information Retrieval (ECIR2016), 20-23 Mar 2016, Padua, Italy. Lecture Notes in Computer Science, 9626 . Springer International Publishing , pp. 549-560. ISBN 978-3-319-30671-1

https://doi.org/10.1007/978-3-319-30671-1_40

Harvesting Training Images for Fine-Grained Object Categories using Visual Descriptions

Josiah Wang[1], Katja Markert[2,3], and Mark Everingham[3] *

[1] Dept. of Computer Science, University of Sheffield, Sheffield, United Kingdom
j.k.wang@sheffield.ac.uk
[2] L3S Research Center, Leibniz-University Hannover, Hannover, Germany
markert@l3s.de
[3] School of Computing, University of Leeds, Leeds, United Kingdom

Abstract. We harvest training images for visual object recognition by casting it as an IR task. In contrast to previous work, we concentrate on fine-grained object categories, such as the large number of particular animal subspecies, for which manual annotation is expensive. We use 'visual descriptions' from nature guides as a novel augmentation to the well-known use of category names. We use these descriptions in both the query process to find potential category images as well as in image reranking where an image is more highly ranked if web page text surrounding it is similar to the visual descriptions. We show the potential of this method when harvesting images for 10 butterfly categories: when compared to a method that relies on the category name only, using visual descriptions improves precision for many categories.

Keywords: Image retrieval, text retrieval, multi-modal retrieval

1 Introduction

Visual object recognition has advanced greatly in recent years, partly due to the availability of large-scale image datasets such as ImageNet [4]. However, the availability of image datasets for *fine-grained* object categories, such as particular types of flowers and birds [10,16], is still limited. Manual annotation of such training images is a notoriously onerous task and requires domain expertise.

Thus, previous work [2,3,6,7,8,9,12,14] has automatically harvested image datasets by retrieving images from online search engines. These images can then be used as training examples for a visual classifier. Typically the work starts with a keyword search of the desired category, often using the category name e.g. querying Google for "butterfly". As category names are often polysemous and, in addition, a page relevant to the keyword might also contain many pictures not of the required category, images are also filtered and reranked. While some work reranks or filters images using solely visual features [3,6,9,14], others have shown that features from the web pages containing the images, such as the

* Mark Everingham, who died in 2012, is included as a posthumous author of this paper for his intellectual contributions during the course of this work.

This is the first author's self-archived version of the paper (v2), posted on 4th April 2016. This version contains an addendum regarding the use of the term "recall" in the paper. The official publication is available at Springer via
http://dx.doi.org/10.1007/978-3-319-30671-1_40

Fig. 1. A visual description from eNature[1] for the Monarch butterfly *Danaus plexippus*. We explore whether such descriptions can improve harvesting training images for fine-grained object categories.

neighbouring text and metadata information, are useful as well [2,7,8,12] (see Sect. 1.1 for an in-depth discussion). However, prior work has solely focused on basic level categories (such as "butterfly") and not been used for fine-grained categories (such as a butterfly species like "*Danaus plexippus*") where the need to avoid manual annotation is greatest for the reasons mentioned above.

Our work therefore focuses on the automatic harvesting of training images for fine-grained object categories. Although fine-grained categories pose particular challenges for this task (smaller number of overall pictures available, higher risk of wrong picture tags due to needed domain expertise, among others), at least for natural categories they have one advantage: their instances share strong visual characteristics and therefore there exist 'visual descriptions', i.e. textual descriptions of their appearances, in nature guides, providing a resource that goes far beyond the usual use of category names. See Fig. 1 for an example.

We use these visual descriptions for harvesting images for fine-grained object categories to (i) improve search engine querying compared to category name search and (ii) rerank images by comparing their accompanying web page text to the independent visual descriptions from nature guides as an expert source. We show that the use of these visual descriptions can improve precision over name-based search. To the best of our knowledge this is the first work using visual descriptions for harvesting training images for object categorization.[2]

1.1 Related Work

Harvesting training images. Fergus et al. [6] were one of the first to propose training a visual classifier by automatically harvesting (potentially noisy) train-

[1] http://www.enature.com/fieldguides

[2] Previous work [15,1,5] has used visual descriptions for object recognition without any training images but not for the discovery of training images itself.

ing images from the Web, in their case obtained by querying Google Images with the object category name. Topic modelling is performed on the images, and test images are classified by how likely they are to belong to the best topic selected using a validation set. However, using a single best topic results in low data diversity. Li et al. [9] propose a framework where category models are learnt iteratively, and the image dataset simultaneously expanded at each iteration. They overcome the data diversity problem by retaining a small but highly diverse 'cache set' of positive images at each iteration, and using it to incrementally update the model. Other related work includes using multiple-instance learning to automatically de-emphasise false positives [14] and an active learning approach to iteratively label a subset of the images [3].

Harvesting using text and images. The work described so far involves filtering only by images; the sole textual data involved are keyword queries to search engines. Berg and Forsyth [2] model both images *and* their surrounding text from Google *web* search to harvest images for ten animal categories. Topic modelling is applied to the *text*, and images are ranked based on how likely their corresponding text is to belong to each topic. Their work requires human supervision to identify relevant topics. Schroff et al. [12] propose generating training images *without* manual intervention. Class-independent text-based classifiers are trained to rerank images using binary features from web pages, e.g. whether the query term occurs in the website title. They demonstrated superior results to [2] on the same dataset without requiring any human supervision. George et al. [7] build on [12] by retrieving images iteratively, while Krapac et al. [8] add contextual features (words surrounding the image etc.) on top of the binary features of [12].

Like [2,7,8,12], our work ranks images by their surrounding text. However, we tackle fine-grained object categories which will allow the harvesting of training images to scale to a large number of categories. In addition, we do not only use the web text surrounding the image but use the visual descriptions in outside resources to rank accompanying web-text by their similarity to these visual descriptions. In contrast to the manual topic definition in [2], this method does then not require human intervention during harvesting.

1.2 Overview

We illustrate harvesting training images for ten butterfly categories of the Leeds Butterfly Dataset [15], using the provided eNature visual descriptions. Figure 2 shows the pipeline for our method, starting from the butterfly species' name and visual description. We obtain a list of candidate web pages via search engine queries (Sect. 2). These are parsed to produce a collection of images and text blocks for each web page, along with their position and size on the page (Sect. 3). Image-text correspondence aligns the images with text blocks on each web page (Sect. 4). The text blocks are then matched to the butterfly description (Sect. 5), and images ranked based on how similar their corresponding text blocks are to the visual description (Sect. 6). The ranked images are evaluated in Sect. 7, and conclusions offered in Sect. 8.

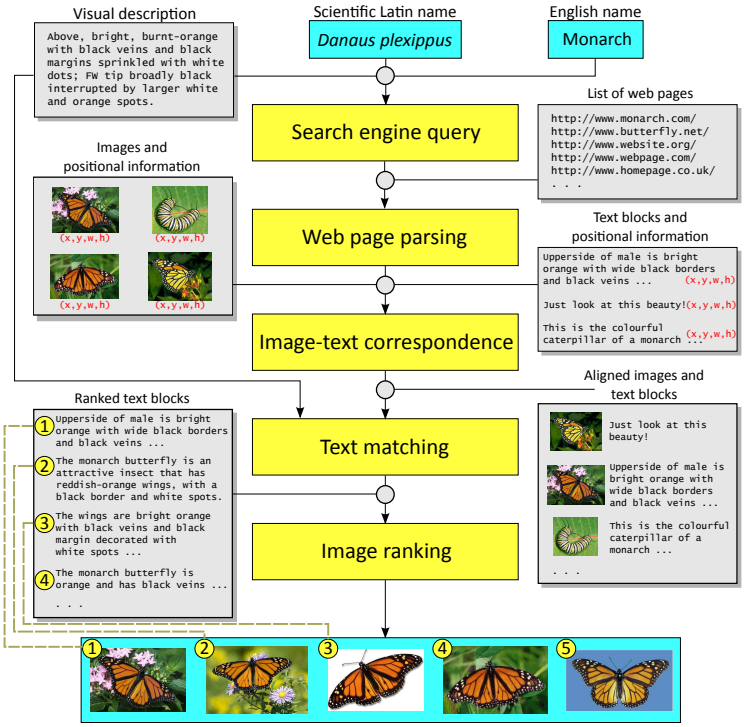

Fig. 2. General overview of the proposed framework, which starts from the butterfly species name (Latin and English) and description, and outputs a ranked list of images.

2 Search Engine Query

We use Google search to obtain as many candidate pages as possible containing images (along with textual descriptions) of the desired butterfly categories. To later compare our method using visual descriptions to one using category names only, we retrieve candidate pages by several different methods. First, we have four *base queries* mainly based on the category name. Here we use both the butterfly's (i) Scientific (Latin) name; (ii) common (English) name. As English names may be polysemous, the term "butterfly" is appended to these for better precision. To increase the recall of visual descriptions, additional queries are produced by appending "description OR identification" to the butterfly name. Our four *base queries* are: (i) "Latin name"; (ii) "English name" + butterfly; (iii) "Latin name" + (description OR identification); (iv) "English name" + butterfly + (description OR identification).

Besides the base queries, we aim to raise precision by also using phrases from the eNature textual descriptions themselves as *seed* terms for the queries; this returns web pages with similar phrases which could potentially include visual descriptions for the butterfly category. The seed phrases are restricted to *noun*

phrases and *adjective phrases*, obtained via phrase chunking as in [15]. The number of seed phrases per category ranges from 5 to 17 depending on the length of the description; an example list is shown in Fig. 3. We query Google with the butterfly name augmented with each seed phrase individually, and with all possible combinations of seed phrase pairs and triplets (e.g. *'Vanessa atalanta' bright blue patch pink bar white spots*).

Two sets of seeded queries are used: one with the Latin and one with the English butterfly name. For each category, all candidate pages from the base and the seeded queries (54 to 1670 queries per category, mean 592) are pooled. For de-duplication, only one copy of pages with the same web address is retained.

Description: FW tip extended, clipped. Above, black with orange-red to vermilion bars across FW and on HW border. Below, mottled black, brown, and blue with pink bar on FW. White spots at FW tip above and below, bright blue patch on lower HW angle above and below.

Seed phrases: black brown and blue; bright blue patch; fw tip; hw border; lower hw angle; orange red to vermilion bars; pink bar; white spot

Fig. 3. Seed phrases for *Vanessa atalanta* extracted from its visual description.

3 Web Page Parsing

Previous work [2,7,8,12] performs image-text correspondence by parsing the HTML source code of a web page, and extracting any non-HTML text surrounding an image link, assuming that such text is positioned close to the image. However, this assumption is not always correct as the HTML source does not always dictate how a web page is *displayed*. The presentation of a web page is most often controlled by style sheets or scripts that dynamically change the web page's layout. As such, web page elements may be freely positioned independent of their sequence in the HTML source. Another example is the use of tables, where cells are defined from left-to-right and then top-to-bottom. Thus, text in a table cell might not be aligned to an image in the cell above since they may be positioned far apart from each other in the HTML source. These issues could be alleviated by using DOM trees, e.g. [17], but they still encode mainly structural and semantic information of web page elements and not positional information.

To address this issue, we match text and images by *where* they are located on the page as rendered to the user. Such positional information is not available from the HTML source or DOM tree, but is dependent on a browser layout engine which generates this information. We use QtWebKit[3], an implementation of the WebKit web browser engine in the Qt Framework. It provides details of all elements in a web page, including the tag name, content, horizontal and vertical positions, width, and height. The nature of the elements themselves also provide additional information, for example whether they are displayed at 'block level'

[3] http://trac.webkit.org/wiki/QtWebKit

(e.g. a paragraph) or 'inline level' (e.g. <*span*>, <*a*>, <*i*>). For our work, we consider as *text blocks* all text within block-level elements (including tables and table cells) and those delimited by any images or the <*br*> element. All images and text blocks are extracted from web pages, along with their height, width, and (x, y) coordinates as would be rendered by a browser. The renderer viewport size is set as 1280×1024 across all experiments.

4 Image-text Correspondence

The list of images and text blocks with their positional information is then used to align text blocks to images (see Fig. 4 for an illustration). An image can correspond to multiple text blocks since we do not want to discard any good candidate visual descriptions by limiting ourselves to only one nearest neighbouring text. On the other hand, each text block may only be aligned to its closest image; multiple images are allowed only if they both share the same distance from the text block. This relies on the assumption that the closest image is more likely to correspond to the text blocks than those further away.

An image is a candidate for alignment with a text block only if all or part of the image is located directly above, below or either side of the text block. All candidate images must have a minimum size of 120×120. For each text block, we compute the perpendicular distance between the closest edges of the text block and each image, and select the image with the minimum distance subject to the constraint that the distance is smaller than a fixed threshold (100 pixels in our experiments). Text blocks without a corresponding image are discarded.

5 Text Matching

The text matching component computes how similar a text block is to the visual description from our outside resource, using IR methods. We treat the butterfly's visual description as a *query*, and the set of text blocks as a collection of *documents*. The goal is to search for documents which are similar to the query and assign each document a similarity *score*.

There are many different ways of computing text similarity, and we only explore one of the simplest in this paper, namely a bag of words, frequency-based vector model. It is a matter of future research to establish whether more sophisticated methods (such as compositional methods) will improve performance further. We represent each document as a vector of term frequencies (tf). Separate vocabularies are used per query, with the vocabulary size varying between 1649 to 9445. The vocabulary consists of all words from the *document* collection, except common stopwords and *Hapax legomena* (words occurring only once). Terms are case-normalised, tokenised by punctuation and Porter-stemmed [11]. We use the *lnc.ltc* weighting scheme of the SMART system [13], where the *query* vector uses the log-weighted term frequency with idf-weighting, while the *document* vector uses the log-weighted term frequency without idf-weighting. The relevance score between a query and a document vector is computed using the cosine similarity measure.

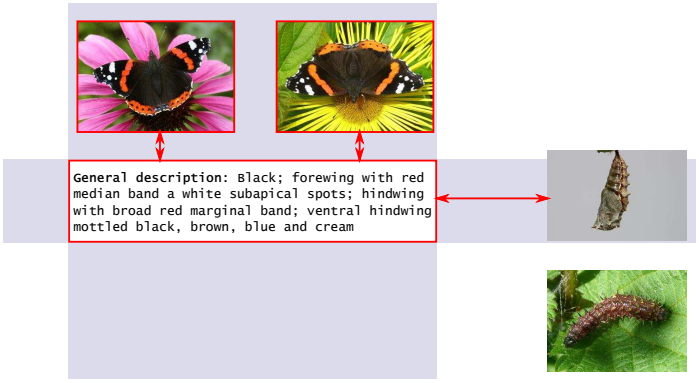

Fig. 4. An illustration of the proposed image-text correspondence algorithm. The text block is matched to the two top images as they are both of the same distance from the text block. The image on the right is not matched as it is further away from the text block than the top two images. The caterpillar image on the bottom right is not considered as it is outside the 'candidate region' (shaded region in figure), i.e. it is not directly above or below, or directly to the left or right of the text block.

6 Image Ranking and Filtering

Each text block from Sect. 5 is treated as a candidate butterfly description, and assigned a similarity score with regard to the category visual description. Images are ranked by the *maximum* score among an image's neighbouring text blocks. Intuitively, for each image we choose the text block most likely to be a visual description and use this score to rerank the image collection. As many images from a web page may be irrelevant (e.g. page headers, icons, advertisements), we filter by retaining only images where their metadata (image file name, *alt* or *title* attribute) contains the butterfly name (Latin or English) and excludes a predefined list of 'negative' terms (e.g. caterpillar, pupa).

7 Experimental Results

We evaluate the image rankings via precision at selected recall levels. We compare our reranked images using visual descriptions to the Google ranking produced by name search only.

Annotation. For each category, we annotated the retrieved images as 'positive' (belonging to the category), 'negative' or 'borderline'. Borderline cases include non-photorealistic images, poor quality images, images with the butterfly being too small, images with major occlusions or extreme viewpoints, etc. Only positive and negative cases are considered during evaluation. For a fair evaluation we ignore borderline cases as they are not exactly 'incorrect' but are just poor examples; it would have been acceptable to have them classified either way.

Table 1. Statistics of annotated images, before and after pre-filtering.

Category	Number of retrieved images				Number of images after pre-filtering			
	Positive	Negative	Borderline	Total	Positive	Negative	Borderline	Total
Danaus plexippus	23.1%	59.2%	17.7%	**12470**	42.7%	34.4%	23.0%	**5240**
Heliconius charitonius	45.9%	39.4%	14.7%	**2053**	70.8%	8.2%	21.0%	**1025**
Heliconius erato	31.5%	61.2%	7.2%	**1132**	37.9%	55.2%	6.8%	**701**
Junonia coenia	45.5%	39.5%	15.0%	**3055**	66.8%	9.4%	23.8%	**1507**
Lycaena phlaeas	52.5%	39.0%	8.4%	**1947**	73.5%	15.8%	10.7%	**945**
Nymphalis antiopa	36.7%	50.6%	12.7%	**3078**	60.7%	18.2%	21.1%	**1297**
Papilio cresphontes	29.0%	52.5%	18.5%	**3815**	48.9%	18.0%	33.0%	**1571**
Pieris rapae	34.6%	56.6%	8.8%	**2742**	59.7%	27.9%	12.4%	**1112**
Vanessa atalanta	26.6%	63.3%	10.0%	**6822**	63.8%	16.2%	20.0%	**2150**
Vanessa cardui	19.4%	72.6%	8.0%	**10301**	47.2%	37.3%	15.4%	**3158**

Statistics and Filtering Evaluation. Table 1 provides the statistics for our annotations. The table shows the level of noise, where many images on the web pages are unrelated to the butterfly category. Filtering via metadata dramatically reduces the number of negative images without too strongly reducing the number of positive ones. The cases where the number of negative images is high after filtering are due to the categories being visually similar to other butterflies, which often have been confused by the web page authors.

Baselines. We use the four base queries (using predominantly category names) as independent baselines for evaluation. For each base query, we rank each image according to the rank of its web page returned by Google followed by its order of appearance on the web page. Images are filtered via category name appearance in metadata just as in our method. We also compare the results with two additional baselines, querying Google Images with (i) "Latin name"; (ii) "English name" + butterfly. These are ranked using the ranks returned by Google Images.

Results. We concentrate on the *precision* of images at early stages of recall, i.e. obtaining as many correct images as possible for top-ranked images. Figure 5 shows the precision-recall curves for our method against the baselines, up to a recall of 50 images. The precision for *Junonia coenia*, *Lycaena phlaeas*, *Pieris rapae* and *Vanessa atalanta* is consistently higher than all baselines across different recall levels. The precision of most remaining categories is relatively high, although not better than all baselines. There were some misclassifications at very early stages of recall for *Danaus plexippus* and *Papilio cresphontes*; however, the overall precision for these is high, especially at later stages of recall. The performance of *Heliconius charitonius* and *Nymphalis antiopa* is comparable to their best baselines. *Vanessa cardui* also gave higher precision than its baselines up to a recall of about 20 images. The only poor performance came from *Heliconius erato*: many subspecies of this butterfly exist which are visually different from the nature guide description, making ranking by similarity to description unsatisfactory. Our method needs categories with strong shared visual characteristics to work fully.

The main mistakes made by our method can be attributed to (i) the web pages themselves; (ii) our algorithm.

In the first case, the ambiguity of some web page layouts causes a misalignment between text blocks and images. In addition, errors arise from mistakes made by the page authors, for example confusing the Monarch (*Danaus plexippus*) with the Viceroy butterfly (*Limenitis archippus*).

For mistakes caused by our algorithm, the first involves the text similarity component. Apart from similar butterflies having similar visual descriptions, some keywords in the text can also be used to describe non-butterflies, e.g. "pale yellow" can be used to describe a caterpillar or butterfly wings. The second mistake arises from text-image misalignment as a side-effect of the filtering step: there were cases where a butterfly image does not contain the butterfly name in its metadata while a caterpillar image on the same page does. Since the butterfly image is discarded, the algorithm matches a text block with its next nearest image – the caterpillar. This could have been rectified by not matching text blocks associated with a previously discarded image, but it can be argued that such text blocks might still be useful in certain cases, e.g. when the discarded image is an advertisement and the next closest image is a valid image.

Figure 6 shows the top ranked images for *Danaus plexippus*, along with the retrieved textual descriptions. All descriptions at early stages of recall are indeed of *Danaus plexippus*. This shows that our proposed method performs exceptionally well given sufficient textual descriptions. The two image misclassifications that still are present are from image-text misalignment, as described above.

8 Conclusion

We have proposed methods for automatically harvesting training images for fine-grained object categories from the Web, using the category name and visual descriptions. Our main contribution is the use of visual descriptions for querying candidate web pages and reranking the collected images. We show that this method often outperforms the frequently used method of just using the category name on its own with regards to precision at early stages of recall. In addition, it retrieves further textual descriptions of the category.

Possible future work could explore different aspects: (i) exploring better language models and similarity measures for comparing visual descriptions and web page text; (ii) training generic butterfly/non-butterfly visual classifiers to further filter or rerank the images; (iii) investigating whether the reranked training set can actually induce better visual classifiers.

Acknowledgements

The authors thank Paul Clough and the anonymous reviewers for their feedback on an earlier draft of this paper. This work was supported by the EU CHIST-ERA D2K 2011 Visual Sense project (EPSRC grant EP/K019082/1) and the Overseas Research Students Awards Scheme (ORSAS) for Josiah Wang.

References

1. Ba, J.L., Swersky, K., Fidler, S., Salakhutdinov, R.: Predicting deep zero-shot convolutional neural networks using textual descriptions. In: Proceedings of the IEEE International Conference on Computer Vision (2015) 2
2. Berg, T.L., Forsyth, D.A.: Animals on the web. In: Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition. vol. 2, pp. 1463–1470 (2006) 1, 2, 3, 5
3. Collins, B., Deng, J., Li, K., Fei-Fei, L.: Towards scalable dataset construction: An active learning approach. In: Proceedings of the European Conference on Computer Vision. pp. 86–98 (2008) 1, 3
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database. In: Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition. pp. 248–255 (2009) 1
5. Elhoseiny, M., Saleh, B., Elgammal, A.: Write a classifier: Zero-shot learning using purely textual descriptions. In: Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition (2013) 2
6. Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A.: Learning object categories from Google's image search. In: Proceedings of the IEEE International Conference on Computer Vision. vol. 2, pp. 1816–1823 (2005) 1, 2
7. George, M., Ghanem, N., Ismail, M.A.: Learning-based incremental creation of web image databases. In: Proceedings of the 12th IEEE International Conference on Machine Learning and Applications (ICMLA 2013). pp. 424–429 (2013) 1, 2, 3, 5
8. Krapac, J., Allan, M., Verbeek, J., Jurie, F.: Improving web-image search results using query-relative classifiers. In: Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition. pp. 1094–1101 (2010) 1, 2, 3, 5
9. Li, L.J., Wang, G., Fei-Fei, L.: OPTIMOL: Automatic Object Picture collecTion via Incremental MOdel Learning. In: Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition. pp. 1–8 (2007) 1, 3
10. Nilsback, M.E., Zisserman, A.: Automated flower classification over a large number of classes. In: Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing. pp. 722–729 (2008) 1
11. Porter, M.F.: An algorithm for suffix stripping. Program 14(3), 130–137 (1980) 6
12. Schroff, F., Criminisi, A., Zisserman, A.: Harvesting image databases from the Web. IEEE Transactions on Pattern Analysis & Machine Intelligence 33(4), 754–766 (2011) 1, 2, 3, 5
13. Singhal, A., Salton, G., Buckley, C.: Length normalization in degraded text collections. In: Proceedings of Fifth Annual Symposium on Document Analysis and Information Retrieval. pp. 149–162 (1996) 6
14. Vijayanarasimhan, S., Grauman, K.: Keywords to visual categories: Multiple-instance learning for weakly supervised object categorization. In: Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition (2008) 1, 3
15. Wang, J., Markert, K., Everingham, M.: Learning models for object recognition from natural language descriptions. In: Proceedings of the British Machine Vision Conference. pp. 2.1–2.11. BMVA Press (2009) 2, 3, 5
16. Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S., Perona, P.: Caltech-UCSD Birds 200. Tech. Rep. CNS-TR-2010-001, California Institute of Technology (2010) 1
17. Zhou, N., Fan, J.: Automatic image-text alignment for large-scale web image indexing and retrieval. Pattern Recognition 48(1), 205–219 (2015) 5

Addendum

Added by Josiah Wang on 4th April 2016.

In the paper, we used the term "**recall**" in terms of *number* of images rather than to be a real number between 0.0 and 1.0. We later realised after submitting the camera-ready manuscript that the term "**rank**" would have been more accurate and succinct. Thus, to be consistent with IR terminology, we clarify that the evaluation measure "**Precision@**K" was used, and we concentrated on achieving high precision at small values of K. The x-axis in Figure 5 actually refers to the *rank*, not *recall*. The graphs in Figure 5 are plots of the precision at selected ranks of up to $K = 50$.

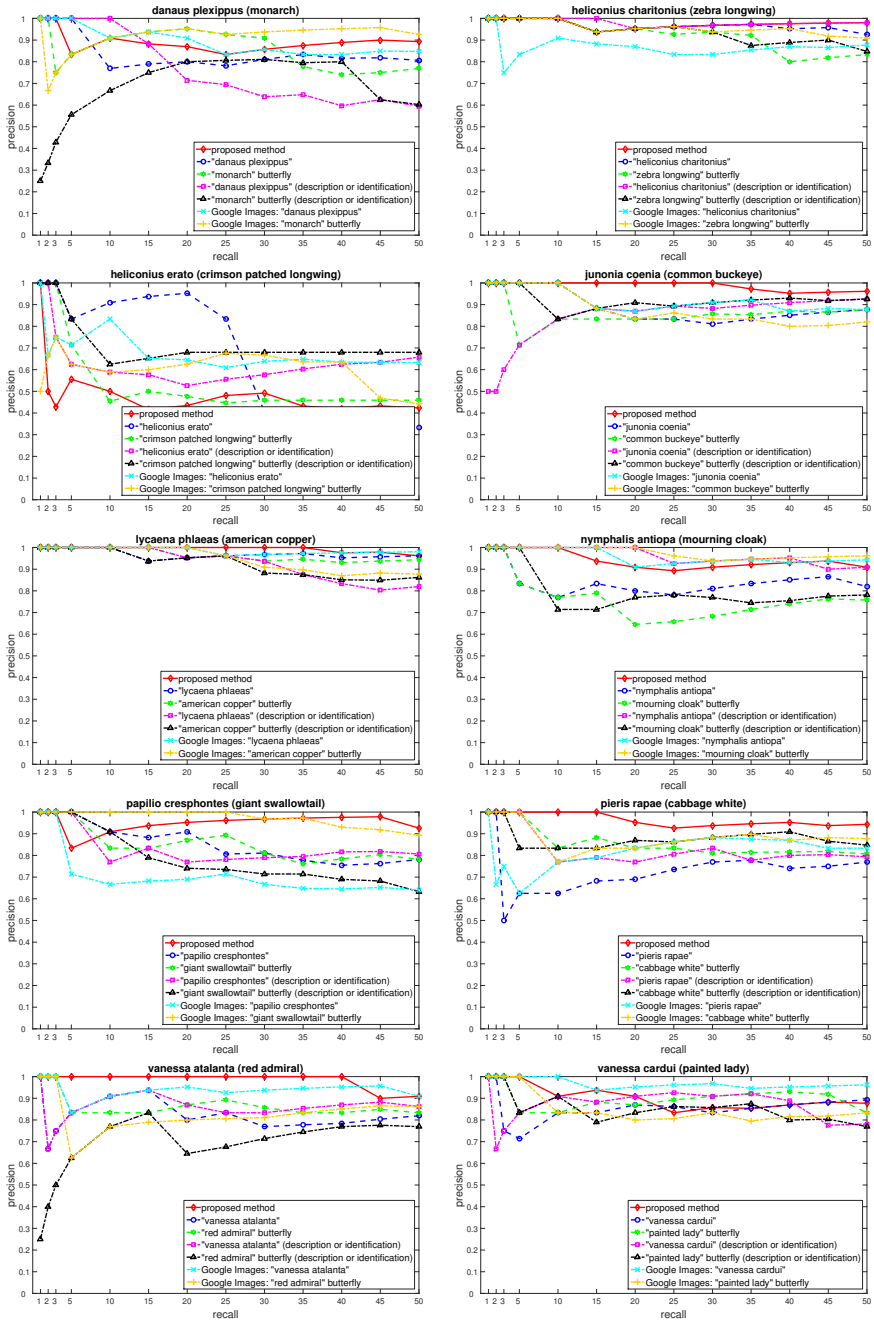

Fig. 5. Precision at selected levels of recall for the proposed method for ten butterfly categories, compared to baseline queries. The recall (x-axis) is in terms of *number* of images. For clarity we only show the precisions at selected recalls of up to 50 images.

1		Monarch Butterfly (Danaus plexippus) Description 3 1/2-4" (89-102 mm). Very large, with FW long and drawn out. Above, bright, burnt-orange with black veins and black margins sprinkled with white dots; FW tip broadly black interrupted by larger white and orange spots. Below, paler, duskier orange. 1 black spot appears between HW cell and margin on male above and below. Female darker with black veins smudged.
2		Description : Family: Nymphalidae, Brush-footed Butterflies view all from this family Description 3 1/2-4" (89-102 mm). Very large, with FW long and drawn out. Above, bright, burnt-orange with black veins and black margins sprinkled with white dots; FW tip broadly black interrupted by larger white and orange spots. Below, paler, duskier orange. 1 black spot appears between HW cell and margin on male above and below. Female darker with black veins smudged. Similar Species Viceroy smaller, has shorter wings and black line across HW. Queen and Tropic Queen are browner and smaller. Female Mimic has large white patch across black FW tips. . . .
3		The wings are bright orange with black veins and black margin decorated with white spots. Female's veins are thicker.
4		Diagnosis: The Monarch is one of the largest Canadian butterflies (wingspan: 93 to 105 mm). The upperside is bright orange with heavy black veins, and a wide black border containing a double row of white spots. There is a large black area near the wing tip containing several pale orange or white spots. The underside is similar except that the hindwing is much paler orange. Males have a sex patch, a wider area of black scales on a vein just below the centre of the hindwing.
5		male bright orange w/oval black scent patch (for courtship) on HW vein above, and abdominal "hair-pencil;" female dull orange, more thickly scaled black veins
6		Description: This is a very large butterfly with a wingspan between 3 3/8 and 4 7/8 inches. The upperside of the male is bright orange with wide black borders and black veins. The hindwing has a patch of scent scales. The female is orange-brown with wide black borders and blurred black veins. Both sexes have white spots on the borders and the apex. There are a few orange spots on the tip of the forewings. The underside is similar to the upperside except that the tips of the forewing and hindwing are yellow-brown and the white spots are larger. The male is slightly larger than the female.
7		General description: Wings orange with black-bordered veins and black borders enclosing small white spots. Male with small black scent patch along inner margin. Ventral hindwing as above but paler yellow-orange and with more prominent white spots in black border. Female duller orange with wider black veins; lacks black scent patch on dorsal hindwing.
8		A large butterfly, mainly orange with black wing veins and margins, with two rows of white spots in the black margins. The Monarch is much lighter below on the hindwing, and males have a scent patch - a dark spot along the vein - in the center of the hindwing.
9		Wingspan: 3 1/2 to 4 inches Wings Open: Bright orange with black veins and black borders with white spots in the male. The male also has a small oval scent patch along a vein on each hind wing. The female is brownish-orange with darker veins Wings Closed: Forewings are bright orange, but hind wings are paler
		...
16		The Monarch's wingspan ranges from 3–4 inches. The upper side of the wings is tawny-orange, the veins and margins are black, and in the margins are two series of small white spots. The fore wings also have a few orange spots near the tip. The underside is similar but the tip of the fore wing and hind wing are yellow-brown instead of tawny-orange and the white spots are larger. The male has a black patch of androconial scales responsible for dispersing pheromones on the hind wings, and the black veins on its wing are narrower than the female's. The male is also slightly larger.

Fig. 6. Top ranked images for *Danaus plexippus*, along with their corresponding descriptions. A red border indicates that the image was misclassified. The first description is almost identical to the eNature description.