Exploring the Long Tail of Social Media Tags

Svetlana Kordumova, Jan van Gemert, and Cees G.M.Snoek

University of Amsterdam
{s.kordumova, j.c.vanGemert, cgmsnoek}@uva.nl,

Abstract. There are millions of users who tag multimedia content, generating a large vocabulary of tags. Some tags are frequent, while other tags are rarely used following a long tail distribution. For frequent tags, most of the multimedia methods that aim to automatically understand audio-visual content, give excellent results. It is not clear, however, how these methods will perform on rare tags. In this paper we investigate what social tags constitute the long tail and how they perform on two multimedia retrieval scenarios, tag relevance and detector learning. We show common valuable tags within the long tail, and by augmenting them with semantic knowledge, the performance of tag relevance and detector learning improves substantially.

1 Introduction

In this paper we focus on the long tail frequency distribution of social tags. It is well known that tag frequencies in social media form a long tail distribution [17, 24]. While some tags are frequent, like *snow, beach, coffee*, there is a large number of tags which are rare, with only few example images per tag, like *mierkat, tank suit, dyippy*, see Fig. 1. Current works note that many tags from the long tail are "misspelled" or "meaningless" words [24], or only useful for "exceptional cases" [17]. It seems this observation has been accepted in the community as such, and no further investigation has been performed so far. We believe that there are also meaningful tags within the long tail which have been overlooked. Since the tags from the long tail make up a significant portion of the data, they deserve more detailed analysis. On that account, we pose the question: *What tags constitute the long tail?*.

We believe that the long tail distribution should not be just accepted as such, but looked at as a challenge. The challenge is to augment the frequencies of rare tags, so that the long tail distribution will change its shape. By augmenting the rare tags, they become a richer source for many multimedia algorithms. Motivated from common approaches in the literature of enriching tags with semantic knowledge [3, 26], we investigate the effect of augmenting the examples of rare tags with semantically related tagged images. We question *What happens when rare tags are augmented?*.

Social tags come for free and are a valuable resource for many multimedia methods that aim to automatically understand visual content. Examples include automatic concept detection [6, 9, 21], user profiling [4, 14], sentiment analysis

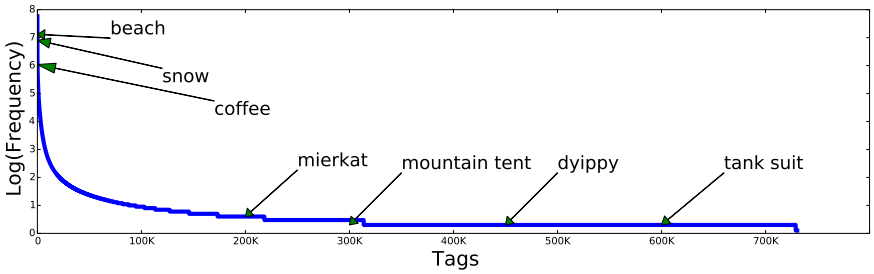

Fig. 1. The log-frequency of Flickr images for 730K tags. Some tags like *beach*, *snow*, *coffee* have millions of example images. We investigate the majority of rare tags such as *miercat*, *mountain tent*, and *tanksuit*.

[1, 23, 25], and assessing tag relevance [10, 20]. These works give excellent results when ample training examples are available per tag. It is not clear, however, how these methods will perform for rare tags with fewer examples. Since there are many tags which fall on the long tail, we argue that it is of great importance for the tags from the long tail to also be successful on these methods. Therefore, we question *What is the effect of rare tags on multimedia retrieval scenarios?*, and evaluate the effect when rare tags are augmented with semantics.

The contributions of this paper are three fold. First, we analyze the type of tags that occur in the long tail distribution of social tags. We base the analysis on a representative snapshot of Flickr containing 1 million photos with 700K diverse tags, and additional 38K tags from three categories *objects, scenes* and *fine-grain animals*. Second, we investigate augmenting the rare tags with semantic knowledge. Third, we exploit two multimedia retrieval scenarios on sampled tags from the long tail distribution, and analyze their performance on both rare and augmented tags.

2 Related Work

Many works in social media are focused on frequent tags with many example images per tag. This is quite understandable, since those tags are popular and evidently important for many users. In Figure 2 we show the trend of social media works over the frequency distribution, which confirms the tendency towards frequent tags. In this paper we analyze the non frequent tags.

Frequent tags. One widely recognized problem is that tags are noisy, ambiguous and often not directly related to the visual content [10, 26]. For that reason, Li *et al.* [10] defined a *tag relevance* metric, calculated by counting neighborhood votes from visually similar images, with the intuition that visually similar images should share the same tags. Their method has shown impressive performance when evaluated on frequent tags. For rare tags we believe that this method could be problematic. For example if a tag is relevant to an image but occurs very rarely or not at all in the corpus, it will not get enough votes from

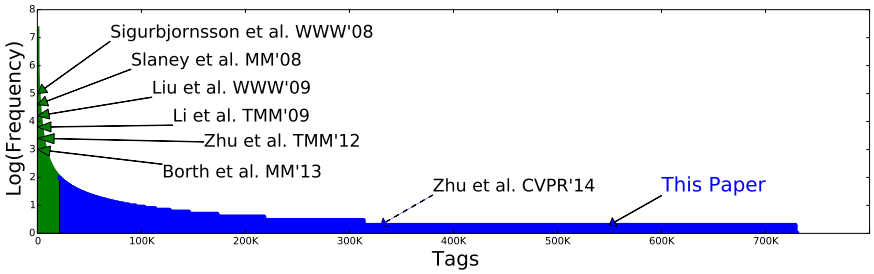

Fig. 2. The trend of related works over tag frequency distribution, confirming the emphasis in the literature towards frequent tags.

its neighbors and will be falsely considered as irrelevant. Zhu *et al.* [26] use the related tags to handle imprecise and incomplete image tags. They calculate the relevance of an image tag by collectively analyzing statistics from Flickr and the WordNet hierarchy, so that a tag receives examples from its child nodes. It might be troublesome if a rare tag is a leaf node in the WordNet hierarchy, without child nodes to be enriched with. Having this in mind, we investigate how calculating tag relevance on rare tags performs, and if including semantic knowledge would help improving the relevance of a tag.

Many other diverse topics exist when using frequent tags. The authors of [1] have created a visual sentiment ontology by sending queries of adjective noun pairs (ANP) to Flickr. In their work ANP candidates (about 320k) are ranked by occurrence frequency, with the goal to remove extremely rare constructions and to keep only frequent ones. While the frequent ANPs show valuable performance when learning detectors, it is unclear what will be the performance of the rare ANPs. The tag ranking method of [12] automatically ranks tags associated with an image using the neighborhood of images containing a specific tag. Finding tagged neighbors of rare tags is problematic, which might lead to assigning low scores for relevant tags. In other works, like tag ambiguity [18], learning detectors to tag [24] or tag recommendation [17], the authors use co-occurrence statistics. If a tag is rare, its co-occurrence statistics will be unreliable, which might result in erroneous scores. In [6,9,21] social tags are used to collect training data and learn concept detectors. For rare tags there are less images to be used as training data. Therefore, we believe it is interesting to also investigate the learning performance of rare tags. We choose to evaluate detectors learning, since it is a popular topic, and the performance can be evaluated on a standard benchmark dataset [2].

Rare tags. To the best of our knowledge there is no related work investigating rare tags in social media. The problem of rare tags is somewhat similar to few- or zero-shot learning [8, 19], designed for problems where there is a lack of training samples. Most of these works use textual or attribute descriptions of known concepts and compare with the textual description or attribute scores of the zero-shot concept. These textual or attribute detectors are learned with manual annotations, which limits the type of concepts that can be detected. The

manual annotations are also quite precise compared to the noisy tags from social media. Therefore, applying these techniques to multimedia methods is difficult, and also requires modifications in the existing algorithms. We investigate a more generic approach that takes the noise of social tags into account, and still deals with the low number of examples.

Closest to our work is a recent study on objects [27], where it is shown that object categories follow a long tail distribution. Most of the object subcategories are rare, which makes it difficult for learning detectors. The authors address the lack of training data by allowing the rare subcategories to share training examples with dominant ones in their learning model. This is possible since the sematic relationships of object categories-subcategories are known, and the images have precise manual annotations. However, social media does not have known semantic relationships between tags. Tagged images are easy to obtain, but they come with noise as additional complication. The method of [27] does not address these challenges, thus, it can not be directly applied on multimedia applications that use social media tags.

Semantic relationships. Relationships between tags have been calculated using co-occurrence statistics [17, 18, 24], or semantic knowledge from external sources [3, 17, 26]. Co-occurrence statistics can be unreliable for calculating tag relationships of rare tags, since they occur rarely or not at all. In the work of Fergus *et al.* [3] an ontology from an external source is adopted to expand query sets for label propagation. In [26] Zhu *et al.* use an ontology to expand the training data. For example, a training set of a non-leaf concept (e.g., building) is enriched by including representative examples from its child nodes (e.g., church). In this paper we investigate external semantic knowledge for a different purpose, to augment the frequency of rare tags.

Although there are many works investigating the possibilities of socially tagged images, none has so far looked into the tags from the long tail of the frequency distribution. In this paper we do a first attempt to analyze rare tags.

3 What tags constitute the long tail?

Tag vocabulary. We analyze tags from Flickr as one of the most popular social media platform. We consider the *Flickr 1M* dataset from [10] as a representative sample. This dataset has 1 million images, downloaded by randomly generating photo ids as queries, making it unbiased towards any tags. The images come with about 700K diverse tags. Since its unpractical to analyze all these tags manually, we consider tags from three categories of existing image recognition datasets: *objects, scenes* and *fine-grained* animals. Many object tags are available in ImageNet [2], 22K classes resulting in 40K tags, since some classes come with multiple tags, like *sea cow, sirenian mammal, sirenian* represent one object class. The SUN Attribute dataset [15] has the highest number of scene tags so far, 717 classes like *amusement park, coast, squash court*. Tags from fine-grained animal categories are available in the 120 dog tag from Stanford Dogs [5] like *pekinese, irish terrier, chihuahua*, and 200 bird tags from Caltech Birds [22] like *shiny*

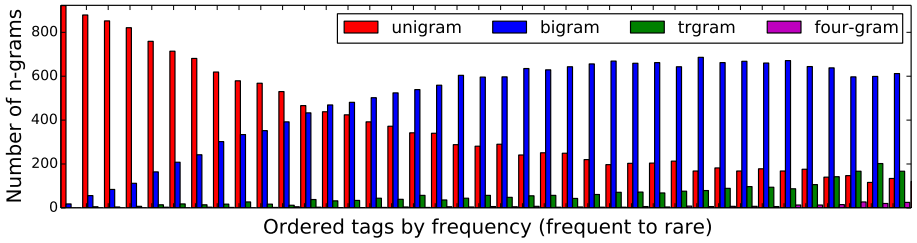

Fig. 3. Occurrence of n-grams, following the long tail frequency distribution.

cowbird, bobolink, blue jay. All tags together form a vocabulary of 730K unique tags. We call this tag set the *Tag vocabulary*, and we analyze it in this experiment.

Tag composition. It is interesting to investigate the occurrence of n-grams, following the long tail frequency distribution. We aim to find if there is a correlation between the number of words a tag is composed of, and its occurrence frequency. It has been noted that tags from the long tail are mostly complex phrases [17], we investigate if this is indeed the case. We merge the tags of objects, scenes and fine-grained animal categories, in total 38K. We send queries with the tag names to Flickr with the Flickr api, and count the occurrence of each tag. We order the tags by their frequency and we count the unigrams, bigrams, trigrams and four-grams in 40 steps. We visualize the histograms in Figure 3. It can be observed that the frequent tags are mostly unigrams. As the frequency goes down, more and more bigrams appear, and towards the end trigrams and four-grams start occurring. Some frequent bigrams are *christmas tree* with frequency 250K, or *polar bear* with frequency 160K. The bigrams *zebrawood tree* or *kaffir cat* have frequency zero. The histograms shows that most of the rare tags are bigrams, and less uni-/tri-/four-grams. We believe this is so because users often use general one word tags to describe an image, and do not try to be precise. We also manually analyzed sampled Flickr tags from the Tag vocabulary, not necessarily only coming from the preselected categories. We looked into 50 tags around 10 steps over the long tail distribution. Interestingly, among the frequent tags we found tags like *iphoneography* and *instagramapp*, which are added from popular mobile applications. We expect that with thousands of images being uploaded in Flickr daily, the frequency of the tags changes, new tags appear, and old ones become more frequent. However, at any given time, when a snapshot of Flickr is taken, like at 2008 [17], at 2009 [24] and ours, the distribution stays heavy tailed. Among the sampled tags from the long tail there were attribute and noun pairs like *scratching post, showing work*, named entities like *saint petersburg russia, arnold aragon*, phrases like *what color is your time, i enjoyed all types of outdoor adventures as a child*, number and word compositions like *photo domino 357, hp850* which represent models of products, and also tags in witch we could not find meaning like *wo0, ell:mcc=222*. Moreover, we summarize, the long tail contains attribute noun pairs, entities of less famous

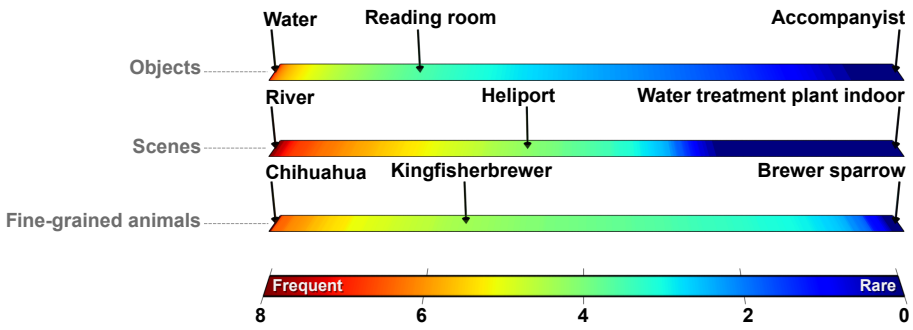

Fig. 4. Heat bars of tag occurrence frequency in Flickr for objects, scenes and fine-grained animal tags. We observe there are many object and scene tags that appear rarely in Flickr, whereas most of the fine-grained animal tags occur reasonably often.

people or geographic places, as well as phrases. The bigrams occur most often in the long tail, and there are less unigrams, trigrams and fourgrams.

Tag categories. We visualize in heat bars the tag occurrence of object, scenes, and fine-grained animal categories in Figure 4. Many **object** tags are colored in blue and appear rarely in Flickr. In numbers, 67% of the object tags appear in less then 1K images. The rare objects are mostly specific tags like *lingberry* which so far occurs in only 3 Flickr images, *marsh tea* in 28 images or *space vehicle* in 67 images. From the **scene** category, 46% of the tags occur in less then 1K images. The strong blue response in the heat bar shows that a great portion of scene tags have low frequency. These rare tags mostly represent fine-grained scenes like *artists loft, canal urban, bakery kitchen*. From the **fine-grained** animal categories, green is the dominant color in the hear bar, and red and blue take small proportions at the ends of the bar. This shows that images of fine grained animal categories occur reasonably often in Flickr. In numbers, 41% of the fine-grained animal tags appear in 1K-10K images, and 36% appear in less then 1K images. This hints that fine-grained classification can be made even more challenging than the classes suggested in the existing fine-grained datasets. Within the fine-grained dogs and birds classes, some unpopular ones are *brabancon griffon* with frequency 0, *brewer blackbird* with frequency 7 and *blenheim spaniel* with frequency 36. Overall, our analysis go against the prevailing norm in the literature where tags from the long tail have been considered unimportant. We show that there are meaningful tags of objects, scenes and fine-grained categories that occur rarely in Flickr and should not be overlooked.

4 Utilizing the long tail

4.1 Augmenting rare tags

We investigate augmenting rare tags from the long tail with semantically similar tags. We believe that augmenting the rare tags is important for good performance

of multimedia retrieval methods. Therefore we analyze the performance of both rare and augmented tags on two multimedia retrieval scenarios in sections 4.2 and 4.3. Motivated from common approaches in the literature of using semantics from external sources like Wikipedia or WordNet [3,26], we investigate if rare tags can be augmented with semantics. From WordNet we consider synonyms and child nodes of a tag in the hierarchy. From Wikipedia we consider titles of redirect pages, as commonly used for semantic linking in information retrieval [13].

Datasets. Since it is not possible to evaluate all 730K tags from the Tag vocabulary, we obtain a representative sample by uniformly sampling each 2,000th tag from the distribution shown in Figure 1. We make sure that the sampled tags have ground truth annotations in some dataset, so that we can evaluate their performance on tag relevance and learning detectors. We consider ground truth classes from ImageNet, as one of the largest available image dataset, with 50 validation images per tag for 1000 classes. If the 2000th tag does not appear in one of the 1000 ImageNet classes, we move to the 2001th tag, and so on. In this manner we select 81 representative tags. These 81 tags contain frequent tags like *light, marmot, blue jean* and rare tags like *whiskey jug, bottle screw, rock snake*. For each tag we download up to 2,000 images from Flickr if available, otherwise as much as we can. This resulted in a new *LongTail* dataset with 13K Flickr images. Additionally, we also downloaded images tagged with semantically similar tags found from WordNet and Wikipedia, forming a new *LongTailAugmented* dataset with 160K images.

Analysis. In Figure 6 (a) we show the frequency of the sampled 81 tags, as well as their frequency when augmented with images of semantically related tags. For most of the rare tags we could find images tagged with their synonym tags, magnifying the frequency when joined. Some rare tags have synonyms from WordNet which are quite frequent, like for example tag 18:*patrol wagon* (id:tag, where id is the position of the tag on the horizontal axis in Figure 6) appears in only 41 images, whereas its WordNet synonyms appear in much more: *police van* in 2310, *paddy wagon* in 3736, *wagon* in 200K images. For tag 56:*Chlamydosaurus kingi* which so far appears in one image, its WordNet synonym *frilled lizard* has 250 images, and Wikipedia finds more synonyms, *Chlamydosaurus* in 278, *frilled dragon* in 150 and *frillnecked* in 150 images. For some tags like 17:*english foxhound* and 67:*mountain tent*, we did not find semantically similar tags, and for tag 39:*plumbers helper* its synonym *plunger* has 0 tagged images. We expect that with more sophisticated language processing more semantically related tags can be found, and the rare tags will be even better augmented. Overall we conclude, that rare tags from the long tail can be augmented with simple synonyms.

4.2 Tag Relevance

In this experiment we investigate the performance of calculating tag relevance for tags which fall on the long tail of the frequency distribution. We investigate the performance of the most popular tag relevance method proposed be Li *et al.* [10], due to its good performance [20].

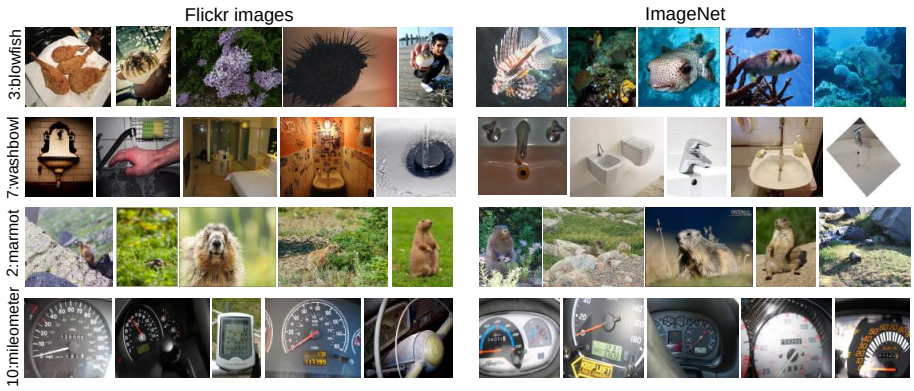

Fig. 5. Example images from Flickr and ImageNet. Tags 3:*blowfish* and 7:*washbowl* have visually very diverse appearance. Tags 2:*marmot* and 10:*mileometer* have more consistent visual features, thus show better performance.

Augmented Tag Relevance. We simply extend the tag-relevance method of [10] to take semantically related tags into account. Instead of just counting votes of the same tag, we also count votes from its synonyms. The augmented tag relevance of a rare tag is computed as follows.

We denote a rare tag with r and $S_r = \{s_1, s_2, ..., s_n\}$ is its set of n synonyms. For an image I_r with tags $\{t|t \in \text{tags}(I)\}$, the set of images tagged with semantically similar tags within its k visual nearest neighbors is

$$N(I_r, S_r, k) = \{I|I \in \text{NN}(I_r, k) \wedge \exists t(t \in \text{tags}(I) \wedge t \in S_r)\}. \quad (1)$$

We calculate the rare tag relevance R for an image I_r as

$$R(r, I_r, k) = |N(I_r, S_r, k)| - P(S_r, k), \quad (2)$$

where $P(S_r, k)$ is the prior tag distribution, which in our case denotes the average prior of the synonym set

$$P(S_r, k) = \frac{1}{|S_r|} \sum_{s \in S_r} k \frac{|L_s|}{|L|}, \quad (3)$$

where k is the number of visual neighbors, $|L_s|$ the number of all images labeled with s, and $|L|$ the size of the entire collection. The difference in this formulation from [10], is adding the set of synonyms S_r, on places where only one tag t was used in the voting.

The tag relevance method is developed for tags which are composed of only one word. If a tag is an n-gram, composed of two or more words, we follow the recommendation from [11], and average the tag relevances or the augmented tag relevances for each word.

Features. As visual features for tag relevance and augmented tag relevance we use the same settings as the multi-feature color and texture variant of [11].

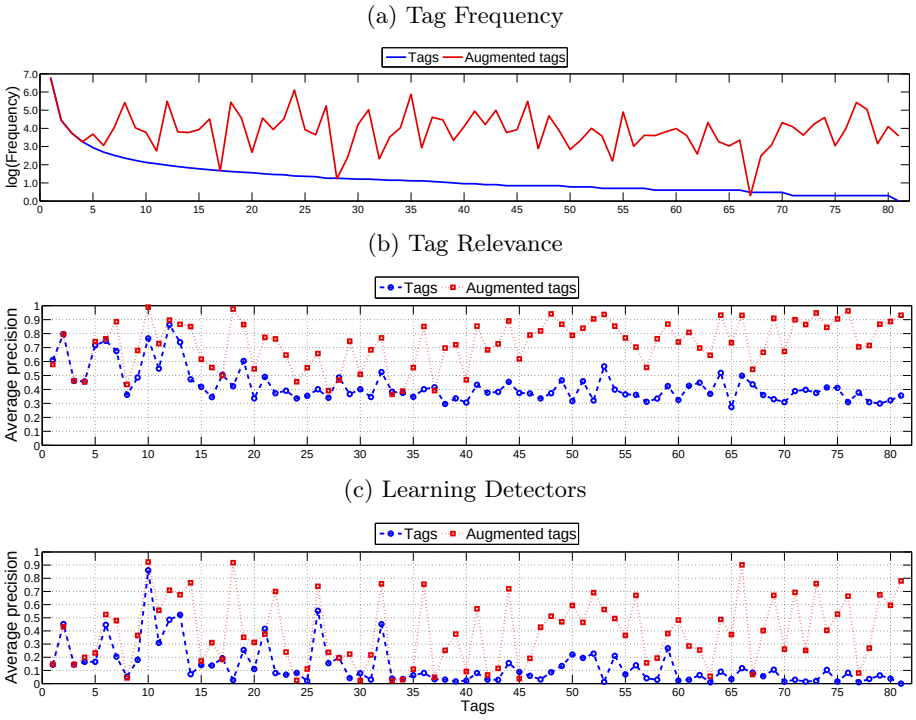

Fig. 6. Frequency and AP of 81 tags, without and with synonym augmentation.

Results and analysis. To evaluate, we use the ImageNet validation images of the same 81 classes as the selected tags in the LongTail dataset. For each tag, there are 50 positive images in the ImageNet validation set. To evaluate tag relevance we also need images with noisy tags. Thus, for each tag we additionally sample 100 random images from the other classes of ImageNet, resulting in 150 images for evaluation of each tag. We evaluate with average precision (AP) per tag and mean average precision (MAP) for overall score.

We plot the AP per tag in Figure 6 (b), for both regular and augmented tag relevance. A negative effect of the long tail can be clearly seen for tag relevance, where most of the the rare tags obtain low average precision. One reason for this is that when there are few or even zero images in Flickr which have the tag, it is unlikely for the visual neighbors to contain the tag resulting in no votes and failing to learn the relevance of the tag. We also notice some outliers, like for 3:*blowfish*, 4:*bluejean* and id 7:*washbowl* which have lower tag relevance performance compared to their neighbors in the frequency distribution. In Figure 5 we show how visually diverse and ambiguous these concepts are, compared to for example 2:*marmot* and 5:*easternfox squirrel*, and 10:*mileometer*.

The augmented tag relevance has a positive effect on the rare tags, since for most of the rare tags the average precision improves. For example, the AP of

73:*lycaenid butterfly*, grows from 37% to 64%, and for 18:*patrol wagon* improves from 49% to 86%. For some tags, there is none or a small improvement. For example 17:*english foxhound* and 67:*mountain tent* have no improvement since no semantically related tags were found. In some cases although the synonym tags are frequent, the results do not improve. For example tag 24:*woodworking plane* has a synonym *plane* which is not as specific and contains many diverse and visually different images, see Figure 7. In cases like this, we expect a more sophisticated semantic method for augmenting the rare tags would help.

Overall, when only few tagged images are present, the tag relevance is determined to fail upfront. When we use augmented tag relevance with synonyms, the mean average precision grows from 43% to 73%. We conclude the tag relevance of the rare tags can be better calculated with augmented semantics.

4.3 Learning Detectors

We investigate learning concept detectors from tagged images, and analyze their performance in correlation with the tag frequency occupance. As a training set we use the LongTail and LongTailAugmented datasets. From the LongTail dataset we select the top ranked images based on their tag relevance score, and from the LongTailAugmented dataset we select the top ranked images based on their augmented tag relevance score. We select the top 1,300 ranked images per tag, as the settings of ImageNet, or less if not as many available. We evaluate on validation images from ImageNet, with AP per tag, or MAP overall.

Features. We recognize that deep learning has shown a great improvement in image classification. The features used from the last layer of a Convolutional Neural Network (CNN), or one layer before the last have become popular and widely used [7]. The CNN is mostly trained on images from ImageNet. Thus, these features have already seen all the classes of ImageNet. Since we evaluate on ImageNet classes, and we want to see the performance of using only few images from a rare tag to learn a concept detector, we do not use the CNN features to keep the evaluation fair. Instead as features we employ the once popular Fisher vector encoding [16] with a GMM of size 1,024 and a spatial pyramid of 1x1 and 1x3. We extract SIFT descriptors with dense sampling at every 6 pixels at two scales, PCA reduced to 80D. As a classifier we use a one-vs-all linear SVM.

Results and analysis. We visualize the results in Figure 6 (c). As expected, the less frequent the tags are, the lower the average precision is. Similarly as in tag relevance, some tags have low AP, even though they are frequent. For example tag 1:*light* has low score since its meaning is ambiguous, as well as tags 3:*blowfish*, 4:*bluejean* and 7:*washbowl*, see Figure 5. For most rare tags, the results improve when we augment the training data with images tagged with their synonyms. We show augmented examples of few tags in Figure 7. For tag 26:*chrysanthemum dog* and 53:*galeocerdo cuvieri* the augmented images are quite relevant, improving the tag annotation result from 55% to 74%, and 2% to 56% respectfully. For tags 33:*barracouta* and 34:*sleuthhound* the results do not improve. For example *snoek* is a synonym of *sleuthhound*, and also a name of a car model, which adds noise to the training data of *sleuthhound*.

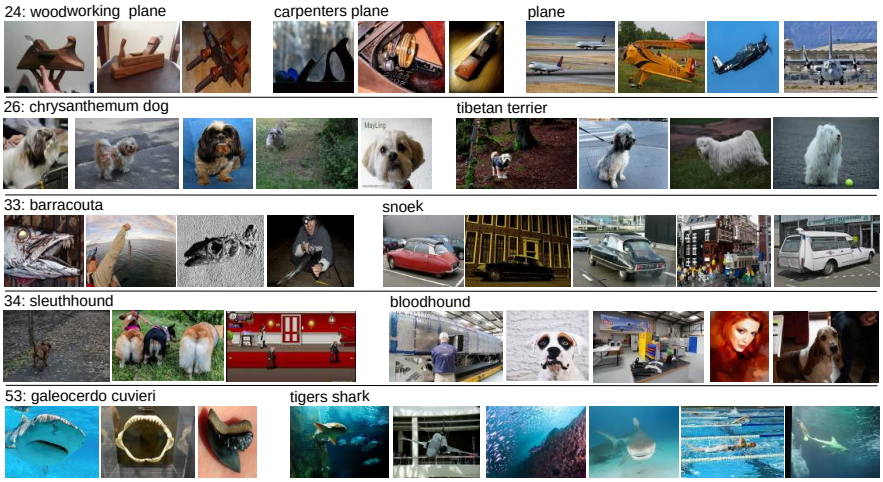

Fig. 7. Example images of tags with the images of their synonyms. Some synonyms augment the rare tags with good images, while others are more ambiguous.

Overall, learning detectors from tagged images with rare tags gives poor performance since there are not enough images to learn reliable detectors. When simply augmenting the training data with images tagged with synonyms, the MAP improves from 53% to 79%. We conclude, learning detectors for rare tags from the long tail can be improved by augmenting the training data with images tagged with their synonyms.

5 Conclusions

We have looked into the long tail of social tags, and analyzed three questions: *What tags constitute the long tail?, What happens when rare tags are augmented?* and *What is the effect of rare tags on multimedia retrieval scenarios?*. We uncover that the long tail has valuable tags of objects, scenes and fine-grained animal categories. We show that by augmenting the rare tags with simple semantics, the performance of tag relevance and detector learning improves considerably. Thus, we conclude the rare tags from the long tail are valuable and perform better when augmented with semantic knowledge.

Acknowledgments. This research is supported by the STW STORY project and the Dutch national program COMMIT.

References

1. D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *MM*, 2013.

2. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009.
3. R. Fergus, H. Bernal, Y. Weiss, and A. Torralba. Semantic label sharing for learning with many categories. In *ECCV*, 2010.
4. A. L. Ginsca, A. Popescu, B. Ionescu, A. Armagan, and I. Kanellos. Toward an estimation of user tagging credibility for social image retrieval. In *MM*, 2014.
5. A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei. Novel dataset for fine-grained image categorization. In *CVPR*, 2011.
6. S. Kordumova, X. Li, and C. Snoek. Best practices for learning video concept detectors from social media examples. *MTAP*, 2014.
7. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*. 2012.
8. C. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot learning of object categories. *TPAMI*, 2013.
9. G. Li, M. Wang, Y.-T. Zheng, H. Li, Z.-J. Zha, and T.-S. Chua. Shottagger: tag location for internet videos. In *ICMR*, 2011.
10. X. Li, C. G. M. Snoek, and M. Worring. Learning social tag relevance by neighbor voting. *TMM*, 2009.
11. X. Li, C. G. M. Snoek, M. Worring, and A. W. M. Smeulders. Harvesting social images for bi-concept search. *TMM*, 2012.
12. D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang. Tag ranking. In *WWW*, 2009.
13. E. Meij, W. Weerkamp, and M. de Rijke. Adding semantics to microblog posts. In *WSDM*, 2012.
14. Y. Ni, M. Zheng, J. Bu, C. Chen, and D. Wang. Personalized automatic image annotation based on reinforcement learning. In *ICME*, 2013.
15. G. Patterson, C. Xu, H. Su, and J. Hays. The sun attribute database: Beyond categories for deeper scene understanding. *IJCV*, 2014.
16. J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek. Image classification with the fisher vector: Theory and practice. *IJCV*, 2013.
17. B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *WWW*, 2008.
18. M. Slaney, K. Weinberger, and R. van Zwol. Resolving tag ambiguity. In *MM*, 2008.
19. R. Socher, M. Ganjoo, C. D. Manning, and A. Y. Ng. Zero-shot learning through cross-modal transfer. In *NIPS*, 2013.
20. B. Truong, A. Sun, and S. Bhowmick. Content is still king: the effect of neighbor voting schemes on tag relevance for social image retrieval. In *ICMR*, 2012.
21. A. Ulges, M. Koch, and D. Borth. Linking visual concept detection with viewer demographics. In *ICMR*, 2012.
22. C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical report, 2011.
23. H. Wang, F. Wu, X. Li, S. Tang, J. Shao, and Y. Zhuang. Jointly discovering fine-grained and coarse-grained sentiments via topic modeling. In *MM*, 2014.
24. L. Wu, L. Yang, N. Yu, and X.-S. Hua. Learning to tag. In *WWW*, 2009.
25. Y. Yang, P. Cui, W. Zhu, H. V. Zhao, Y. Shi, and S. Yang. Emotionally representative image discovery for social events. In *ICMR*, 2014.
26. S. Zhu, C.-W. Ngo, and Y.-G. Jiang. Sampling and ontologically pooling web images for visual concept learning. *TMM*, 2012.
27. X. Zhu, D. Anguelov, and D. Ramanan. Capturing long-tail distributions of object subcategories. In *CVPR*, 2014.