Image Annotation with Parametric Mixture Model Based Multi-class Multi-labeling

Zhiyong Wang [1], Wan-Chi Siu [2], Dagan Feng [3]

[1,3] *School of Information Technologies, University of Sydney, Sydney, Australia*
[1] zhiyong@it.usyd.edu.au, [3] feng@it.usyd.edu.au

[2,3] *Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, China*
[2] enwcsiu@polyu.edu.hk

***Abstract*—Image annotation, which labels an image with a set of semantic terms so as to bridge the semantic gap between low level features and high level semantics in visual information retrieval, is generally posed as a classification problem. Recently, multi-label classification has been investigated for image annotation since an image presents rich contents and can be associated with multiple concepts (i.e. labels). In this paper, a parametric mixture model based multi-class multi-labeling approach is proposed to tackle image annotation. Instead of building classifiers to learn individual labels exclusively, we model images with parametric mixture models so that the mixture characteristics of labels can be simultaneously exploited in both training and annotation processes. Our proposed method has been benchmarked with several state-of-the-art methods and achieved promising results.**

I. Introduction

Visual information has been widely used in a wide range of application domains such as digital libraries. Meanwhile, users realize that it has become more and more difficult to find desired visual content such as images. Traditional content-based image retrieval (CBIR) systems allow users to access visual information by providing examples of desired images through low level visual features (e.g. color, shape, and texture) [1]. Though many successful systems have been demonstrated, the semantic gap is widely recognized as a hurdle for practical adoption of CBIR systems[2]. The semantic gap also results in the difficulty in interpreting the retrieval results. Hence, it has been desirable that images can be automatically labeled with linguistic terms so that both computers and human beings can be brought to the same ground of visual perception and the semantic gap can be reduced, or even eliminated.

Image annotation, which labels an image with a set of semantic terms, is generally posed as a multi-class classification problem where each concept corresponds to a class. Some approaches solve this problem by utilizing traditional single-label paradigm where an image is exclusively labelled with one class according to the output of the classifier. However, in the regime of image annotation, these approaches suffer from the following issues: 1) selecting a large set of suitable labels. For example, selecting *Indoor*, *Outdoor*, and *Scenery* as labels will impose difficulties to classifiers, since most *Scenery* images are definitely of *Outdoor* category. Therefore, the classes to be investigated have to be selected either empirically or based on specific application domains. In [3], a hierarchy of a small number of labels (e.g. *City* vs. *Landscape*) was established to organize vacation images. However, how to construct a hierarchy of a large number of labels is another open issue, though ontology can aid this process in some applications [4]; 2) obtaining training data in a large quantity. Since images are of rich contents and belong to multiple classes, it is challenging to manually label images with only one label; 3) capturing correlation among labels, since one label could provide information about the others; and 4) the fact that more and more images are available with multiple accompanying textual labels and it is desirable to utilize such abundant and valuable information.

Recently, approaches considering the multi-label characteristics of images have been investigated. Some approaches choose to fuse the outputs of individual concept classifiers [5] or utilizing a priori knowledge such as concept ontology[6] and others model the concept correlation directly. Ghamrawi and McCallum proposed to exploit dependencies between labels with conditional random field (CRF) models[7]. Qi *et al* proposed to model the correlation among labels with Gibbs random field (GRF) models[8].

Other than model the co-occurrence relationship among concept as aforementioned, we propose to emphasize the contribution from image contents. That is, which word should be assigned to an image depends on what contents (e.g. image regions) are contained in the image. Therefore, a multi-labeled image has a mixture characteristics contributed by different visual features which correspond to labels and can be represented with a mixture model. Hence, we propose to tackle the aforementioned issues with a parametric mixture model based multi-class multi-labeling technique.

II. Related Work

Due to the nature of image annotation, classification approaches have been widely employed for such a task. Most approaches often focus on two issues, extracting representative and discriminative features and designing efficient classifiers. Here we only briefly review several approaches of the extensive literature due to space constraints. Based on visual features (e.g. color and texture), Vailaya *et al.* proposed to classify vacation images with a Bayesian approach[3]. Li and

978-1-4244-2295-1/08/$25.00 © 2008 IEEE

Wang proposed to model the semantic concepts of images with 2D Multi-resolution Hidden Markov Models (MHMMs) and semantic terms were selected from each image category to annotate images [9].

Multi-label classification methods, which allow a training instance to be associated with multiple labels, have been demanded by many modern applications such as document classification[7][10]. Tsoumakas *et al.* categorized the approaches of multi-label classification into two categories, problem transformation methods and algorithm adaptation methods [11]. The former extend traditional classification approaches by transforming multi-label classification problems either into one or more single-label classification or regression problems [12][13]. For example, Boutell *et al.* investigated multi-label scene classification by incorporating multi-label information for cross-training using Support Vector Machines (SVM) [12]. However, these methods did not take the mixture characteristics into account explicitly. The latter extend specific learning algorithms to handle multi-labeled data directly[10][7][8] by exploiting the co-occurrence of labels, which assumes that training dataset provides sufficient coverage on the co-occurrence of labels.

Recently, image annotation has also been investigated through modeling dual-modality of visual information, visual attributes and textual labels. The co-occurrence of those two modalities was first investigated by Mori *et al.* [14]. In [15], a translation model was proposed to translate a vocabulary of image blobs to a vocabulary of linguistic terms based on the joint probability of image blobs and terms. Based on cross-lingual information retrieval, a cross-media relevance model (CMRM) was proposed to allow for both image annotation and retrieval [16]. In [17], a continuous relevance model (CRM), a continuous version of CMRM, was proposed to handle continuous visual features, which avoided the quantization step adopted in other relevance models. Based on these work, some variant approaches have also been proposed for image annotation and retrieval (e.g. [18]). We will benchmark our proposed approach with the state-of-the-art of this type of methods.

III. Image Representation

Many feature extraction methods have been proposed to characterize image contents[1]. Ideally, objects contained in images can be extracted and described to match human perception, which significantly relies on image segmentation techniques. Currently, two types of image representation schemes are widely employed, segmentation based [19][17] and grid based (i.e. uniform partition) [20]. As indicated in [21], annotation performance varies due to segmentation errors. Sometimes, simple uniform partition based approaches outperformed segmentation based approaches[18][20]. Carneiro *et al* demonstrated that simple uniform partition can achieve best performance[20]. Choosing which scheme depends on both the specific dataset and the annotation approach. In general, each region is represented with high-dimensional visual features. In this paper, each region is represented with 36 features including region color and standard deviation (18-dimension), region average orientation energy (12-dimension with 12 filters), region size, location, convexity, first moment, and ratio of region area to boundary length squared (6-dimension).

In order to mimic the representation of textual documents, a visual vocabulary will be obtained by clustering continuous feature vectors of image regions. Therefore, continuous feature vectors are converted into discrete clusters (i.e. visual term). Note that any better image representation techniques and clustering techniques can be incorporated into our framework.

IV. Parametric Mixture Model

Parametric mixture models were proposed to perform web page classification[10]. In natural language processing domain, documents are generally characterized with attributes derived from a set of words, such as word frequency. In the following discussion, it is assumed that appropriate clustering method has been applied and images have been represented in “visual term” domain; hence we use document(word) and image(visual term) exchangeably.

Given a collection of training documents $\mathcal{D} = \{d^1, ..., d^N\}$, each document d^n is associated with $(\mathbf{x}^n, \mathbf{y}^n)$, where $\mathbf{x}^n$ and $\mathbf{y}^n$ denote the feature vector and the label vector of document d^n, respectively. Let $\mathbf{x}^n = [x_1^n, ..., x_V^n]$ be a feature vector for d^n where x_i^n denotes the frequency of word w_i occurrence in d^n among the vocabulary $\mathcal{V} = \{w_1, ..., w_V\}$ where V is the total number of words in the vocabulary, and $\mathbf{y}^n = [y_1^n, ..., y_L^n]$ be the label vector, where y_l^n takes a value of 1(0) when d^n has (does not have) to the l-th label. Note that L labels represent pre-defined classes or categories and a document always has at least one label.

In the case of multi-class single-label document, it is natural that $\mathbf{x}$ in the l-th category should be generated from a Multinomial distribution

$$P(\mathbf{x}|l) \propto \prod_{i=1}^{V} (\theta_{l,i})^{x_i}, \qquad (1)$$

where $\theta_{l,i}$ is a probability that the i-th word w_i appears in a document belonging to the l-th category and $\sum_{i=1}^{V} \theta_{l,i} = 1$, $\theta_{l,i} \geq 0$.

Therefore, a multi-class and multi-label document can be generalized as

$$P(\mathbf{x}|\mathbf{y}) \propto \prod_{i=1}^{V} (\varphi_i(\mathbf{y}))^{x_i}, \qquad (2)$$

where $\sum_{i=1}^{V} \varphi_i(\mathbf{y}) = 1$ and $\varphi_i(\mathbf{y}) \geq 0$. $\varphi_i(\mathbf{y})$ is a label-dependent probability that the i-th word appears in a document having label vector $\mathbf{y}$. Obviously, it is impractical to independently set a Multinomial parameter vector to each of a distinct $\mathbf{y}$ since there are $2^L - 1$ possible combinations and efficient parameterization is required.

In general, words in a document having multiple labels can be thought as a mixture of characteristic words related to each of the categories. Let $\boldsymbol{\theta}_l = (\theta_{l,1}, ..., \theta_{l,V})$ and $\boldsymbol{\varphi}(\mathbf{y}) =$

$(\varphi_1(\mathbf{y}), ..., \varphi_V(\mathbf{y}))$. Therefore, we can have the following parametric mixture:

$$\boldsymbol{\varphi}(\mathbf{y}) = \sum_{l=1}^{L} h_l(\mathbf{y})\boldsymbol{\theta}_l, \qquad (3)$$

where $h_l(\mathbf{y})$ is a mixing proportion satisfying $h_l(\mathbf{y}) > 0$ and $\sum_{l=1}^{L} h_l(\mathbf{y}) = 1$, and can be interpreted as a degree that $\mathbf{x}$ has the l-th label. Theoretically, any suitable function can be deployed for $h_l(\mathbf{y})$. Here, we follow the linear mixture model proposed by Ueda *et al.*[10],

$$h_l(\mathbf{y}) = \frac{y_l}{\sum_{l'}^{L} y_{l'}}. \qquad (4)$$

Substituting $\boldsymbol{\varphi}(\mathbf{y})$ (Equation 3) and $h_l(\mathbf{y})$ (Equation 4) into $P(\mathbf{x}|\mathbf{y})$ (Equation 2), we can have

$$P(\mathbf{x}|\mathbf{y}, \boldsymbol{\Theta}) \propto \prod_{i=1}^{V} (\frac{\sum_{l=1}^{L} y_l \theta_{l,i}}{\sum_{l'=1}^{L} y_{l'}})^{x_i}, \qquad (5)$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_l\}_{l=1}^{L}$ is a set of unknown model parameters.

A. Parameter Estimation

The unknown parameter $\boldsymbol{\Theta}$ is estimated by maximizing posterior $P(\boldsymbol{\Theta}|D)$. Assume that $P(\mathbf{y})$ is independent of $\boldsymbol{\Theta}$ and using Bayes' Law we have

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} \{\log P(d_n|\mathbf{y}_n, \boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta})\} \qquad (6)$$

In general, a generative model for K non-negative u_1, ..., u_k that satisfy $\sum_{k=1}^{K} u_k = 1$ is known as a Dirichlet distribution

$$p(u_1, ..., u_k) \propto \prod_{k=1}^{K} u_k^{\xi-1} \qquad (7)$$

Thus, the objective function to find $\hat{\boldsymbol{\Theta}}$ is given by

$$J(\boldsymbol{\Theta}; D) = L(\boldsymbol{\Theta}; D) + (\xi - 1) \sum_{l=1}^{L} \sum_{i=1}^{V} \log \theta_{l,i} \qquad (8)$$

where $L(\boldsymbol{\Theta}; D)$ is the likelihood function given by

$$L(\boldsymbol{\Theta}; D) = \sum_{n=1}^{N} \sum_{i=1}^{V} x_{n,i} \log \sum_{l=1}^{L} h_l^n(\mathbf{y}_n)\theta_{l,i} \qquad (9)$$

The optimization problem given by Equation 8 can be solved in a similar manner to the EM algorithm.

$$\begin{aligned} g_{l,i}^n(\boldsymbol{\Theta}) &= \frac{h_l^n \theta_{l,i}}{\sum_{l=1}^{L} h_l^n \theta_{l,i}} \\ \theta_{l,i}^{(t+1)} &= \frac{\sum_{n=1}^{N} x_{n,i} g_{l,i}^n(\boldsymbol{\Theta}^{(t)}) + \xi - 1}{\sum_{i=1}^{V} \sum_{n=1}^{N} x_{n,i} g_{l,i}^n(\boldsymbol{\Theta}^{(t)}) + V(\xi - 1)} \end{aligned} \qquad (10)$$

B. Automatic Annotation

Let $\hat{\boldsymbol{\Theta}}$ denote the estimated parameter. Then, applying Bayes' rule, the optimum label vector $\mathbf{y}^*$ for $\mathbf{x}^*$ of a new document is defined as: $\mathbf{y}^* = \arg\max_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}^*; \hat{\boldsymbol{\Theta}})$ under a uniform class prior assumption. Since this maximization problem is an NP-hard problem, an exhaustive search is prohibitive for a large L. Therefore, a greedy-search algorithm is applied. That is, first , only one y_{l_1} value is set to 1 so that $P(\mathbf{y}|\mathbf{x}^*; \hat{\boldsymbol{\Theta}})$ is maximized. Then, for the remaining elements, only one y_{l_2} value that mostly increase $P(\mathbf{y}|\mathbf{x}^*; \hat{\boldsymbol{\Theta}})$ is set to 1 under a fixed y_{l_1} value. This procedure is repeated until $P(\mathbf{y}|\mathbf{x}^*; \hat{\boldsymbol{\Theta}})$ cannot increase any further or a certain number of labels have been obtained.

V. Experimental Results

Our experiments were conducted on the dataset provided by Duygulu *et al.*[19] which has been widely utilized for evaluating image annotation. This allows us to benchmark the performance of the proposed approach in a strictly controlled manner. The dataset consists of 5,000 images from 50 Corel Stock Photo CDs of which 4500 images for training and 500 images are used for testing. Each CD includes 100 images on the same topic and each image is labeled with 1 to 5 keywords. Overall there are 374 unique keywords of which 263 keywords appear in the testing set. Each image is segmented into regions using Normalized Cuts[22], and there are typically 5 to 10 regions for each image since only the regions larger than a threshold are utilized. Each region is represented with 36 features including region color and standard deviation (18-dimension), region average orientation energy (12-dimension with 12 filters), region size, location, convexity, first moment, and ratio of region area to boundary length squared (6-dimension). Regions of all the images are then clustered into 500 clusters(blobs) using k-means algorithm (Refer to [19] for details about this dataset). Hereby, blobs of an image are equivalent to words of a document.

We compared the annotation performance of our proposed approach (PMM) with other four models: the Co-occurrence Model[14], the Translation Model[19], CMRM[16], and CRM[17]. Note that the first three methods use discrete features obtained through clustering and CRM directly uses continuous features. We followed the experimental methodology used by [19][16][17] to automatically annotate each given image with top 5 words and compute annotation recall and precision for every word in the testing set. Let A be the number of images automatically annotated with a given word, B the number of images correctly annotated with that word, and C the number of images having that word in ground-truth annotation. Then Recall is the ratio between B and C, and Precision is the ratio between B and A. Recall (Precision) values were averaged over the set of testing words, named Mean per-word Recall (Mean per-word Precision).

Table I shows that our proposed approach achieves the best performance in terms of recall rate and the number of words predicted (*i.e.* words having recall rate greater than

TABLE I
COMPARISON WITH OTHER FOUR MODELS

Models	Co-occurrence	Translation	CMRM	CRM	**PMM**
#words with Recall > 0	19	49	66	107	**74**
Results on words with recall ≥ 0					
Mean per-word Recall	0.27	0.21	0.35	0.46	**0.43**
Mean per-word Precision	0.41	0.32	0.39	0.39	**0.24**
Results on words appearing in the testing set					
Mean per-word Recall	0.02	0.04	0.09	0.19	**0.12**
Mean per-word Precision	0.03	0.06	0.10	0.16	**0.07**

0) among the approaches based on discrete models (i.e. Co-occurrence, Translation Model, and CMRM) which perform quantization on visual feature space, and obtains comparable results with CRM based on continuous models which utilizing visual features directly. Note that the figures of those four models are cited from [17]. As discussed in [17][18], direct utilization of continuous features contributed to the performance improvement of CRM over its discrete version CMRM, since there is information loss while continuous feature vectors are quantized and clustering errors may affect the quality of discrete models. Similarly, it could be expected that better performance can be achieved while the parametric mixture model is extended to continuous feature space.

More detailed experimental results of 74 predicted words are shown in Table II. It is observed that our proposed method is not biased towards the labels which have dominant training samples, being different with the approaches modeling the co-occurrence of dual-modality. As shown in the Translation Model [19], the words having more training samples (such as *sky* and *tree* were better recalled than others, which led to quite low precision of these words, since Translation model attempted to annotate images with such words as much as possible. On the contrary, the parametric mixture model performs better for words which do not have dominant training samples, since it is based on the mixture characteristics of the representative ”document words” (i.e. blobs) It is difficult to have the same analysis for other methods since annotation performance on individual words were not provided in the literature.

Though the precision of our proposed method is lower than that of others, it does not mean that the our recall rate is achieved by sacrificing precision, since our experiments were conducted in the same way as others (e.g. 5 words are auto-annotated). It is due to that some words were over-predicted as shown in Table II. For example, word *sun* (the second of 74 words shown in Table II) appears 10 times in the ground truth annotation data and is annotated 34 times with our proposed method, because *sun* did not appear in the ground truth annotation data of some images that should have been annotated with it.

VI. CONCLUSIONS AND FUTURE WORK

A parametric mixture model based multi-class multi-labeling is presented to tackle image annotation in this paper. Each image is modeled with a mixture model which takes into account the characteristics of each label (or category). Very promising experimental results have been demonstrated with even a simple parametric mixture model on the widely used dataset for image annotation. It is expected that a suitable mixture model which can better characterize the generative nature will definitely further improve the performance of image annotation. It is also noticed that image features have been clustered, which may result in information loss and affect the performance, compared with the continuous relevance model (CRM). An immediate extension of this parametric mixture model is to investigate better mixture models directly utilizing continuous visual features. In addition, different large image datasets will be utilized for future experiments, since the manually annotated Corel dataset has been criticized for its suitability in image annotation, which is also confirmed with our experimental results.

ACKNOWLEDGMENTS

The work presented in this paper is partially supported by ARC and UGC grants.

REFERENCES

[1] D. Feng, W. C. Siu, and H. J. Zhang, *Multimedia Information Retrieval and Management*. Berlin Heideberg, Germany: Springer-Verlag, 2003.

[2] M. S. Lew, N. Sebe, and R. Jain, “Content-based multimedia information retrieval: State of the art and challenges,” *ACM Trans. on Multimedia Computing, Communications and Applications*, vol. 2, no. 1, pp. 1–19, February 2006.

[3] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang, “Image classification for content-based indexing,” *IEEE Trans. on Image Processing*, vol. 10, no. 1, pp. 117–130, 2001.

[4] L. Kennedy and A. Hauptmann, “LSCOM lexicon definition and annotation Version 1.0,” in *DTO Challenge Workshop on Large Scale Concept Ontology for Multimedia*. ADVENT Technical Report # 217-2006-3, Columbia University, March 2006.

[5] E. Chang, K. Goh, G. Sychay, and G. Wu, “CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 1, pp. 26–38, January 2003.

[6] J. Fan, H. Luo, Y. Gao, and R. Jain, “Incorporating concept ontology for hierarchical video classification, annotation, and visualization,” *IEEE Trans. on Multimedia*, vol. 9, no. 5, pp. 939–957, August 2007.

[7] N. Ghamrawi and A. McCallum, “Collective multi-label classification,” in *The 14th ACM International Conference on Information and Knowledge Management*. Bremen, Germany, Novemeber 2005, pp. 195–200.

[8] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, T. Mei, and H.-J. Zhang, “Correlative multi-label video annotation,” in *The ACM International Conference on Multimedia*. Augsburg, Germany, September 2007.

[9] J. Li and J. Z. Wang, “Automatic linguistic indexing of pictures by a statistical modeling approach,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, 2003.

[10] N. Ueda and K. Saito, “Parametric mixture models for multi-labeled text,” in *Advances in Neural Information Processing Systems*, vol. 15. MIT Press, Cambridge, MA, 2003, pp. 721–728.

[11] G. Tsoumakas and I. Katakis, “Multi-label classification: an overview,” *International Journal of Data Warehousing and Mining*, vol. 3, no. 3, pp. 1–3, 2007.

[12] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, “Learning multi-label scene classification,” *Pattern Recognition*, vol. 37, pp. 1757–1771, 2004.

[13] J. Li and J. Z. Wang, “Real-time computerized annotation of pictures,” in *The ACM SIGMM International Conference on Multimedia*. Santa Barbara, USA, Oct 2006.

[14] Y. Mori, H. Takahashi, and R. Oka, “Image-to-word transformation based on dividing and vector quantizing images with words,” in *The First International Workshop on Multimedia Intelligent Storage and Retrieval Management (MISRM99)*. Florida, USA, 1999.

[15] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. Jordan, “Matching words and pictures,” *Journal of Machine Learning Research*, vol. 3, pp. 1107–1135, March 2003.

[16] J. Jeon, V. Lavrenko, and R. Manmatha, “Automatic image annotation and retrieval using cross-media relevance models,” in *The ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, USA, 2003.

[17] V. Lavrenko, R. Manmatha, and J. Jeon, “A model for learning the semantics of pictures,” in *The 16th Annual Conference on Neural Information Processing Systems*, 2003.

[18] S. L. Feng, R. Manmatha, and V. Lavrenko, “Multiple Bernoulli relevance models for image and video annotation,” in *The IEEE International Conference on Computer Vision and Pattern Recognition(CVPR)*, vol. 2, 2004.

[19] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth, “Object recognition as machine translation: learning a lexicon for a fixed image vocabulary,” in *The Seventh European Conference on Computer Vision*, 2002, pp. 97–112.

[20] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos, “Supervised learning of semantic classes for image annotation and retrieval,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 394–410, March 2007.

[21] K. Barnard, P. Duygulu, R. Guru, P. Gabbur, and D. Forsyth, “The effects of segmentation and feature choice in a translation model of object recognition,” in *The IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2. Wisconsin, USA, June 2003, pp. 675–682.

[22] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

TABLE II
ANNOTATION DETAILS FOR EACH WORD WHOSE RECALL RATE IS GREATER THAN 0.

Words	# in Groundtruth	# of Annotations	# of Correct Annotations	Recall	Precision
sphinx	1	4	1	1.00	0.25
sun	10	34	9	0.90	0.26
tiger	10	44	9	0.90	0.20
pillar	10	25	9	0.90	0.36
mare	9	23	8	0.89	0.35
foals	9	26	8	0.89	0.31
sunset	7	35	6	0.86	0.17
jet	19	80	15	0.79	0.19
horses	12	32	9	0.75	0.28
swimmers	8	25	6	0.75	0.24
cat	11	43	8	0.73	0.19
garden	10	40	7	0.70	0.18
polar	13	37	9	0.69	0.24
leaf	12	32	8	0.67	0.25
pool	11	13	7	0.64	0.54
tracks	11	44	7	0.64	0.16
flowers	27	47	16	0.59	0.34
cars	17	43	10	0.59	0.23
coral	9	20	5	0.56	0.25
train	11	21	6	0.55	0.29
scotland	11	34	6	0.55	0.18
face	2	5	1	0.50	0.20
light	6	15	3	0.50	0.20
ruins	12	80	6	0.50	0.08
fruit	2	3	1	0.50	0.33
petals	4	7	2	0.50	0.29
railroad	8	11	4	0.50	0.36
snow	31	52	15	0.48	0.29
plane	25	72	12	0.48	0.17
ocean	9	20	4	0.44	0.20
stone	21	54	9	0.43	0.17
frost	7	16	3	0.43	0.19
nest	7	20	3	0.43	0.15
people	74	74	31	0.42	0.42
bear	22	35	9	0.41	0.26
bridge	15	37	6	0.40	0.16
plants	15	27	6	0.40	0.22
reefs	5	10	2	0.40	0.20
forest	11	22	4	0.36	0.18
sky	105	82	38	0.36	0.46
field	17	26	6	0.35	0.23
street	26	50	9	0.35	0.18
mountain	38	46	13	0.34	0.28
ice	12	22	4	0.33	0.18
arctic	3	7	1	0.33	0.14
tulip	3	3	1	0.33	0.33
buildings	54	58	15	0.28	0.26
rocks	22	78	6	0.27	0.08
sculpture	11	20	3	0.27	0.15
clouds	26	34	7	0.27	0.21
shops	4	14	1	0.25	0.07
arch	4	5	1	0.25	0.20
deer	4	23	1	0.25	0.04
antlers	4	3	1	0.25	0.33
herd	4	1	1	0.25	1.00
zebra	4	14	1	0.25	0.07
rodent	4	12	1	0.25	0.08
formula	4	3	1	0.25	0.33
grass	51	61	12	0.24	0.20
fox	9	16	2	0.22	0.13
water	116	65	25	0.22	0.38
smoke	10	3	2	0.20	0.67
valley	11	22	2	0.18	0.09
birds	17	41	3	0.18	0.07
beach	18	18	3	0.17	0.17
bengal	6	2	1	0.17	0.50
sand	19	22	3	0.16	0.14
wall	13	8	2	0.15	0.25
tree	94	36	11	0.12	0.31
locomotive	9	5	1	0.11	0.20
statue	11	16	1	0.09	0.06
desert	11	7	1	0.09	0.14
boats	15	14	1	0.07	0.07
house	19	10	1	0.05	0.10
Average				0.43	0.24