Comparison of Various Web Image Re – Ranking Techniques

Ms. Pooja P. Dutta[1] Prof. Anand Chauhan[2]
[1,2]Department of Computer Engineering
[1,2]Narnarayan Shastri Institute of Technology, Jetalpur, Ahmedabad, GTU-Gujarat, India

Abstract— Image re-ranking, is an quite an efficient way to improve the results that are fetched from the web-based image search query. Given a query keyword for the image, a pool of images are first retrieved based on textual information, then the images are re-ranked based on their visual similarities with the query image according to the user input. But when, the images' visual features do not match with the semantic meanings of the users' entered query or keyword, it becomes a major challenge to make available the actual searched image. Hence, in this paper, the various Web image Re- ranking techniques are studied, on how it approaches towards the Web Image search that the user has input in query.

Key words: Web Image, Web Image Re – Ranking, Image Query, Content Based Retrieval, Semantic Signatures, Visual Similarities, Keywords Expansion

I. INTRODUCTION

Web - scale image search engines mostly use keywords, surrounding text to search images, and content based image retrieval with relevance feedback. But, since the web is a vast database of images in comparison to web-scale systems, the image retrieval accurately, according to the user's intention becomes a major problem. Major web image search engines have adopted 'Online image re-ranking' strategy, which limits users' effort to just one-click feedback.

Web image search engines use keywords as queries and search images supporting the text related to them. It's a little difficult for users to accurately describe the visual content of target pictures required, hence solely mistreatment keywords and therefore text-based image search suffers from the anomaly of question keywords. For instance, mistreatment apple as a question keyword, the retrieved pictures belong to completely different classes, like apple laptop computer, apple logo, apple fruit. To capture users' search intention, extra data needs to be utilized in order to resolve the anomaly. Text-based keyword expansion is a technique to create the textual description of the query in more details. Existing strategies search either synonyms or different linguistic-related words from wordbook. However, the intention of users may be extremely different and can't be accurately captured by these expansions, even with an equivalent question keywords.

Content-based image retrieval with relative feedback is widely utilized in order to resolve this ambiguity. Users are required to pick out the multiple relevant and irrelevant image examples and also the visual similarity metrics are learned through on-line training from them. Images are re-ranked according to the learned visual similarities. However, for web-scale business systems, users' feedback needs to be restricted to the minimum while not on-line training.[2]

The recently proposed method is a one in which when the user inputs a query, a pool of images are retrieved that are having similar visualities with that to the query input. The user is asked to select a pool that is having image similar to his search query image, and then the images are re ranked according to the semantic space of the images in the pool with the semantic space of the query input[1]. But, characterizing the highly diverse images from the web is difficult because it is impossible to learn a universal visual semantic space.[2].

II. TECHNIQUES FOR WEB IMAGE RE - RANKING

A huge development has been made in the area regarding search image from the web according to the user input. The various conventional methods like content based retrieval, attribute dominance[3], pseudo relevance[15], mapped visual features to a universal concept dictionary for image retrieval[5] that did not use query images.

The key part of image re-ranking is to figure visual similarities presenting linguistics relevance of pictures. Many visual options [8] are developed in recent years. However, for various query images, the effective low-level visual options are completely different.

Cui et al. [9] classified query pictures into eight predefined intention classes and gave completely different feature weighting schemes to differing kinds of question pictures. Still, it's tough for the eight coefficient schemes to hide the large diversity of all the net pictures. It's also possible for a question image to be classified to a wrong class. In order to cut back the linguistics gap, query-specific linguistics signature was initially planned in [11]. Kuo et al. [12] recently augmented every image with relevant linguistics options through propagation over a visible graph and a matter graph that were correlate.

Later the conventional framework was found out that used query images so as to make the actual search of image according to the user's intention. Then this method too had an improvement where the semantic space of the query input of the user was compared with the semantic signatures of the images in the search pool retrieved[1].

The various techniques for the web image re ranking are as follows.

A. Intent Search

In this technique, a completely unique net image search approach is presented. It solely needs the user to click on one question image with minimum effort and pictures from a pool retrieved by text-based search square measure re ranked supported each visual and TEXT content. The key contribution is to capture the users' search intention from this one-click question image in four steps.

1) The question image is categorized into one in every of the predefined adaptational weight classes that replicate users' search intention at a rough level. within every class, a specific weight schema is employed to mix visual options adaptational to the present reasonably image to higher re rank the text-based search result.

2) Supported the visual content of the question image designated by the user and thru image clump, question keywords area unit distended to capture user intention.
3) Distended keywords area unit accustomed enlarge the image pool to contain additional relevant pictures.
4) distended keywords also are accustomed expand the question image to multiple positive visual examples from that new question specific visual and textual similarity metrics area unit learned to more improve content-based image reranking. of these steps area unit automatic, while not additional effort from the user.

B. Attribute Dominance

When we look into a picture, some properties or attributes of the image stand out quite others. Once describing an image, people are probably to explain these dominant attributes first. Attribute dominance could be a results of a fancy interplay between the varied properties present or absent in the image, which attributes in a picture are additionally dominant than others and reveals rich information concerning the content of the image. This technique, emphasis on information, by modeling attribute dominance. It tends to show that this helps improve the performance of vision systems on a range of human-centric applications like zero-shot learning, image search and generating matter descriptions of pictures.
The technique consists of four steps as follows.

1) Annotating Attribute Dominance

Here images are annotated in attribute dominance train the attribute dominance predictor. The dominance annotations are collected at the category level, although approach trivially generalizes to image level dominance annotations as well.

2) Modeling Attribute Dominance

Given a novel image x_t, we predict the dominance $d^m{}_t$ of attribute m in that image using

$$d^m{}_t = w^T{}_m \varphi(x_t)$$

We represent image xt via an image descriptor. We use the output scores of binary attribute classifiers to describe the image. This exposes the complex interplay among attributes discussed in the introduction that leads to the dominance of certain attributes in an image and not others. The relevant aspects of the interplay are learnt by our model. φ(xt) can be just xt or an implicit high- (potentially infinite-) dimensional feature map implied by a kernel. For training, we project the category-level attribute dominance annotations to each training image.

3) Zero-shot Learning

In zero-shot learning , the supervisor describes novel N' previously unseen categories in terms of their attribute signatures $\{g^m{}_{n'}\}$, n' ∈ {1, . . . , N'}. With a pre-trained set of M binary classifiers for each attribute Direct Attribute Prediction (DAP) model, the probability that an image *x* belongs to each of the novel categories $C_{n'}$ is

$$pa_{n'}(x) \; \alpha \; {}^M\pi_m{=}1pa^m(x)$$

where, $pa^m(x)$ is the probability that attribute a^m takes the value $g^m{}_{n'} \in \{0, 1\}$ in image x as computed using the binary classifier for attribute a^m. The image is assigned to the category with the highest probability $pa_{n'}$ (x). This approach forms our baseline. It relies on an interface where a supervisor goes through every attribute in a pre-defined arbitrary order and indicates its presence or absence in a test category.[3]

4) Image Search

Here, the image search situation wherever a user has a target class in mind is considered, and provides as question an inventory of attributes that describe that class. it's unlikely that the user can offer the values of all M attributes once describing the query. (S)he is probably going to use the attributes dominant in the target construct, naming the foremost dominant attributes first.

C. Automatic Query Expansion[4]

This technique explores the ways to derive better object models given the query area, so as to enhance retrieval performance. It tends to keep the form of the model fixed: it's still a configuration of visual words. However, instead of merely extracting the model from the one input question region, it tends to enrich it with extra information from the corpus; and tend to check with latent model described in this technique.
An outline of the approach is as follows[4]:

1) Given a query area, search the corpus and retrieve a set of image regions that match the query area. We use bag-of-visual-words retrieval along with special verification, but the approach would apply to retrieval systems that use completely different object models.
2) Mix the retrieved regions, together with the first query, to create a richer latent model of the item of interest.
3) Re-query the corpus exploitation this swollen model to retrieve a swollen set of matching regions.
4) Repeat the method as necessary, alternating between
5) model refinement and re-querying.

D. Query by Semantic Example[5]

A combination of query-by-visual-example (QBVE) and linguistics retrieval (SR), denoted as query-by-semantic-example (QBSE), is used in this technique. Pictures are tagged with regard to a vocabulary of visual ideas, as is common in SR. Every image is then represented by a vector, cited as a linguistics multinomial, of posterior concept probabilities. Retrieval relies on the query-by-example paradigm: the user provides a query image, for which 1) a linguistics multinomial is computed and 2) matched to those within the information. QBSE is shown to own tow main properties of interest, one mostly practical and also the other philosophical.

From a practical point of view, because it inherits the generalization ability of SR within the area of familiar visual ideas (referred to because the linguistics space) however performs far better outside of it, QBSE produces retrieval systems that are additional correct than what was antecedently potential. Philosophically, as a result of it permits a direct comparison of visual and linguistics representations beneath a common question paradigm, QBSE permits the planning of experiments that expressly check the worth of linguistics representations for image retrieval. An implementation of QBSE beneath the minimum probability of error (MPE) retrieval framework, antecedently applied with success to each QBVE and SR, is planned, and used to demonstrate the two properties. Specifically, an in depth objective comparison of QBSE with QBVE is given, showing that the previous considerably outperforms the latter each within and outside the linguistics area. By rigorously dominant the structure of the linguistics area, it's additionally shown that this improvement will only be attributed to the linguistics nature of the illustration on which QBSE relies.[5]

E. Conventional Image Re – Ranking Framework[1]

Online image re-ranking that limits users' effort to simply one-click feedback, is an efficient means to improve search results and its interaction is easy enough. Major internet image search engines have adopted this strategy. Its diagram is shown in Fig. 1. Given a query keyword input by a user, a pool of pictures relevant to the question keyword area unit retrieved by the computer program according to a stored word-image index file. Typically the size of the came back image pool is mounted, e.g., containing 1,000 pictures. By asking the user to pick out a question image, which reflects the user's search intention, from the pool, the remaining pictures within the pool area unit re-ranked based on their visual similarities with the question image. The word image index file and visual options of pictures area unit pre-computed offline and stored.1 the most on-line process cost is on comparison visual features. To achieve high potency, the visual feature vectors got to be short and their matching has to be quick. Some standard visual features area unit in high dimensions and potency isn't satisfactory if they're directly matched.[1]

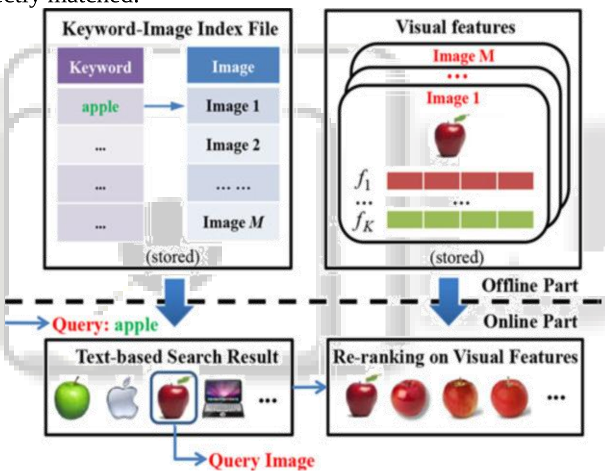

Fig. 1: Conventional Image Re – Ranking Framework

F. Web Image Re – Ranking using Semantic Signatures

Another major challenge while not on-line training, is that, the similarities of low-level visual features might not well correlate with images' high-level linguistics meanings which interpret users' search intention. Moreover, low-level options are sometimes inconsistent with perception. As an example, if pictures of a similar object are captured from completely different viewpoints, beneath completely different lightings or perhaps with different compression artifacts, their low-level options may change considerably, though humans suppose the visual content doesn't change a lot of, to reduce this semantic gap and inconsistency with perception, there are variety of studies to map visual options to a collection of predefined ideas or attributes as semantic signatures [7], as an example, Kovashka et al. [7] planned a system that refined image search with relative attribute feedback. Users described their search intention with reference pictures and a collection of pre-defined attributes. These ideas and attributes are pre-trained offline and have tolerance with variation of visual content. However, these approaches are solely applicable to closed image sets of comparatively tiny sizes, however not appropriate for on-line web-scale image re-ranking.

Since the topics of net pictures amendment dynamically, it is fascinating that the ideas and attributes are often automatically found rather than being manually outlined.

In this technique, a completely unique framework is planned for net image re-ranking. Rather than manually shaping a universal idea dictionary, it learns totally different linguistics areas for various query keywords severally and mechanically. The semantic space associated with the pictures to be re-ranked is significantly narrowed down by the question keyword provided by the user, for instance, if the question keyword is "apple," the ideas of "mountain" and "Paris" area unit irrelevant and should be excluded. Instead, the ideas of "computer" and "fruit" are going to be used as dimensions to find out the linguistics house associated with "apple." The query-specific semantic spaces can more accurately model the images to be re-ranked, since they have excluded different potentially unlimited variety of irrelevant ideas, that serve solely as noise and deteriorate the re-ranking performance on each accuracy and computing price. The visual and matter features of pictures area unit then projected into their connected semantic areas to urge semantic signatures. At the online stage, pictures area unit re-ranked by scrutiny their linguistics signatures obtained from the semantic space of the question keyword. The linguistics correlation between ideas is explored and incorporated once computing the similarity of linguistics signatures.

Experiments show that the semantic space of a question keyword is delineate by simply 20-30 ideas (also referred as "reference classes"). Thus, the semantic spaces are terribly short and on-line image re-ranking becomes extremely economical. As a result of the big variety of keywords and the dynamic variations of the net, the linguistics spaces of question keywords area unit mechanically learned through keyword growth.[1]

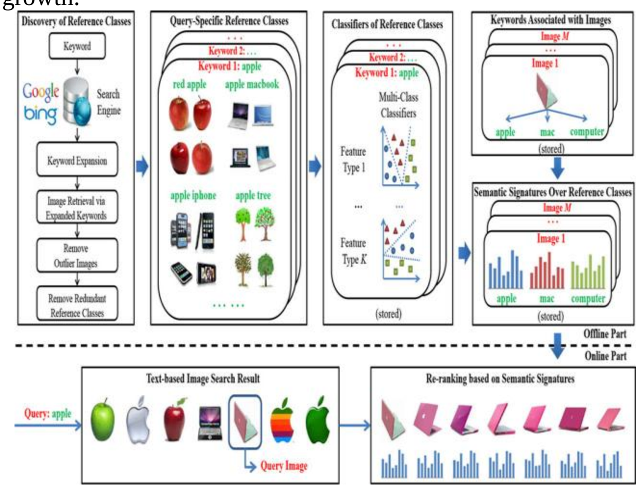

Fig. 2: Image Re – Ranking Framework using Semantic Signatures

III. COMPARATIVE STUDY OF THE VARIOUS TECHNIQUES

The various techniques described above have their advantages and disadvantages. They tabular representation of the comparison is described in Table I.

Techniques	Description	Features	Limitations
content-based retrieval[13]	data only offers linguistic similarity by processing	searches on tradition of pattern matching	semantic gap, retrieval content of image
attribute dominance[3]	Focuses data modelling by attribute dominance.	models attribute dominance	notion of dominance for relative attributes
pseudo relevance[15]	derives better object models for query area for better retrieval performance.	shows great effect for multimedia retrieval in very noisy data.	retrieval focuses unit of data and not the video document to be retrieved.
Query by Semantic Example (QBSE)[5]	combination of query-by-visual-example (QBVE) and linguistics retrieval (SR)	semantic representations have an intrinsic benefit for image retrieval	feed forward processing with few neural layers, the classification results not great for human standards, time consuming
Conventional Image Re – ranking Framework[1]	pool of images relevant to query keyword, given by user is retrieved based on a stored word-image index file.	Online image re-ranking that limits users' effort to simply one-click feedback	Correlation of the images' high level semantics with low level visual features
Web Image Re – Ranking by Semantic Signatures[1]	offline stage - new semantic signatures learned, online stage - images are retrieved on it.	Offline stage of learning increases efficiency and images searched on semantic signatures	cannot directly increase the diversity of search result.

Table 1: Comparison Of The Various Techniques

IV. CONCLUSIONS AND FUTURE WORK

Web Image Re ranking is one of the popular research areas, where a large number of techniques have been proposed for it. Starting with the Content Based Retrieval to Semantic signatures a huge work has been done, where the images are retrieved and re ranked based on attributes, keywords, visual features, their dominant attributes, relevance feedback of the user, pseudo relevance and many more. But the re ranking with semantic signatures has been quite effective in this area, since it learns the semantic signatures offline and then uses these semantic signatures for image retrieval at the online stage thereby taking the time only for the online stage. In future work, a new technique using semantic signatures will be proposed to improve execution and efficiency.

REFERENCES

[1] Xiaogang Wang, Shi Qiu, Ke Liu, and Xiaoou Tang. "Web Image Re-Ranking UsingQuery-Specific Semantic Signatures." IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 36, No. 4, April 2014

[2] Harshil Jain, Riya Lodha and Khushali Deulkar. "Web-based Image Search using Image Re-Ranking Technique: A Review Paper" International Journal of Current Engineering and Technology, Vol.4, No.5, Oct 2014

[3] Naman Turakhia and Devi Parikh. "Attribute Dominance: What Pops Out?"IEEE International Conference on Computer Vision 2013

[4] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2007.

[5] N. Rasiwasia, P.J. Moreno, and N. Vasconcelos, "Bridging the Gap: Query by Semantic Example," IEEE Trans. Multimedia, vol. 9, no. 5, pp. 923-938, Aug. 2007.

[6] Xiaoou Tang, Ke Liu, Jingyu Cui, Fang Wen and Xiaogang Wang. "IntentSearch: Capturing User Intention for One-Click Internet Image Search". IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 34, NO. 7, JULY 2012

[7] A. Kovashka, D. Parikh, and K. Grauman, "WhittleSearch: Image Search with Relative Attribute Feedback," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2012.

[8] J. Philbin, M. Isard, J. Sivic, and A. Zisserman, "Descriptor Learning for Efficient Retrieval," Proc. 11th European Conf. Computer Vision ECCV), 2010.

[9] J. Cui, F. Wen, and X. Tang, "Real Time Google and Live Image Search Re-Ranking," Proc. 16th ACM Int'l Conf. Multimedia, 2008.

[10] J. Cui, F. Wen, and X. Tang, "Intent Search: Interactive on-Line Image Search Re-Ranking," Proc. 16th ACM Int'l Conf. Multimedia, 2008.

[11] X. Wang, K. Liu, and X. Tang, "Query-Specific Visual Semantic Spaces for Web Image Re-Ranking," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2010.

[12] Y. Kuo, W. Cheng, H. Lin, and W. Hsu, "Unsupervised Semantic Feature Discovery for Image Object Retrieval and Tag Refinement," IEEE Trans. Multimedia, vol. 14, no. 4, pp. 1079-1090, Aug. 2012.

[13] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval," IEEE

Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349-1380, Dec. 2000.

[14] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-Based Image Retrieval,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349-1380, Dec. 2000.

[15] R. Yan, E. Hauptmann, and R. Jin, “Multimedia Search with Pseudo-Relevance Feedback,” Proc. Int’l Conf. Image and Video Retrieval, 2003.