Turkish Journal of Computer and Mathematics Education
DOI: https://doi.org/10.17762/turcomat.v9i3.13917 *Vol.09 No.03 (2018), 1245-1252*

Research Article

An Architecture for Retrieval and Annotation of Images from Big Image Datasets

Bhanu Prakash Dube

Asst. Professor, Department of CSE (Computer sc)

GEHU-Dehradun Campus

Abstract:

Rapid advancements in mobile devices and communication technologies have resulted in the daily production of enormous quantities of visual data at both personal and business settings. These developments have stoked a need for improved methods of storing, annotating, and retrieving digital photographs. Images need to be well annotated and indexed for precise retrieval from large-scale picture classes. As a result, there is a growing need for sophisticated strategies for indexing, annotating, and retrieving images. In response to problems discovered during research into existing Image annotation and retrieval systems, the suggested system was developed. The Autoencoder is built with three levels of complexity: an encoder, a decoder, and a single layer of code. Dimensionality reduction is accomplished by extracting the Micro-Structure Descriptor (MSD) of the training picture, which has 72 dimensions. Dimensionality reduction is the focus of a new Autoencoder technique. To organise the wide variety of picture types, a unique Autoencoder Hashing method has been developed.

Keywords: Autoencoder, mobile devices and communication technology, Micro-Structure Descriptor

Introduction

The field of Information Retrieval (IR) focuses on how data may be stored, retrieved, and used. The things are represented and organised in such a way that the user may quickly find the specific pieces of data in which they are interested. An IR system's primary goal, therefore, is to unearth data that the user will find useful or interesting. Books, magazines, papers, pictures, videos, and other media are all examples of information objects. In this research, we investigate how to efficiently retrieve photos from large-scale image classes. The term "image retrieval system" refers to a computer-based system that allows users to browse, search, and retrieve digital photos from a huge database. Annotated pictures are a crucial component of the image retrieval system. Annotation techniques are used by the majority of well-known and widely-used approaches to image retrieval in order to quickly assign information such as keywords, descriptions, and captions to the pictures. Annotating images is therefore fundamental to the image retrieval process. Annotating an image is a kind of multi-label classification that seeks to pair a picture with a collection of descriptive words. It might be used in the future for things like picture indexing and retrieval. Demand for automated picture annotation techniques has increased in response to the current flood of digital photographs on the Internet and in personal collections. As a result, research into Image Annotation and Retrieval has become a vital topic of study.

The rapid development of mobile device technology has resulted in a daily deluge of publicly available digital photographs. This has made it very difficult to find what you're looking for in massive picture databases. In addition, effective picture retrieval from large-scale image classes remains an open challenge and subject of study. As a result, sophisticated techniques for indexing, annotating, and retrieving images are in high demand. In particular, one of the most important research topics in the fields of Pattern recognition and Artificial intelligence is the annotation and retrieval of pictures belonging to large-scale image classes. There are generally three types of image retrieval techniques. The first kind of annotation is the commonplace textual one. In this approach, human beings first annotate the photographs with keywords, and then those images are downloaded from a database. However, manually annotating a large collection of photographs is impractical. Human annotations are also too nebulous and vulnerable to interpretation. The second strategy puts an emphasis on a method called Content Based Image Retrieval (CBIR). Colour, shape, and texture are only few of the low-level properties used in CBIR's automated indexing and retrieval of pictures. However, recent studies have shown that people's interpretations of pictures differ significantly from the low-level properties. Because a visual query is needed for retrieval, the CBIR system is also not very user-friendly for the average person. Automatic picture Annotation (AIA) is used in the third method, which is semantic based picture retrieval. The foundation of the AIA method is the utilisation of a huge collection of training pictures to automatically infer semantic concept models, which are then applied to unlabeled examples. Images may be accessed using keywords once they have been semantically labelled. AIA's main selling point is that it's compatible with both the CBIR and image-based keyword searches. Image annotation and retrieval system efficacy is feature-size and feature-type dependent. Using a high number of image characteristics may boost the system's efficiency. The system's temporal complexity will rise as a result of this, however. Therefore, techniques for reducing the number of dimensions are needed so that the system can run more efficiently and in less time. For efficient indexing and retrieval of pictures belonging to large-scale image classes, this research improves the dimensionality reduction approach, image annotation, and retrieval system. Image annotation, retrieval, and dimensionality reduction techniques are discussed here.

Manual Annotation

When an image's areas are annotated manually, artists draw them and provide descriptions of what they see. When photos are evaluated manually, the user often supplies the most relevant keywords. Good accuracy is achieved with manual annotation since keywords are selected according to how humans understand the pictures' semantic content. However, doing so is a tedious and time-consuming task. Annotations made by hand are not only subjective, but the user may forget them after some time has passed.

Ontology Based Annotation

Ontology specifies the ideas and relationships between them that are utilised to express and represent the topic at hand on the Semantic Web. The purpose of an ontology is to define the potential relationships between concepts, as well as any limitations or requirements placed on their usage, for a given application. Images may have meaningful annotations added using ontology. An picture's semantic relationships are the primary focus of ontology-based image annotation. The first step is to translate the image's low-level features into meaningful concepts. These terms are then linked to ontologies and schemas. The quality of image retrieval is enhanced by the attention paid to providing a comprehensive description of pictures in a semantically abundant ontology. Low-level picture characteristics and high-level textual information work well together to improve ontology's performance. However, developing the semantic web is a time-consuming task.

Semi Automatic Annotation

The user is involved in the annotating process in semi-automatic picture annotation. Combining elements of both automated and human annotation yields the semi-automatic annotation approach. While previewing the retrieved results, the user gives the answer. When a user enters a search term, only those photographs that best match that term will be returned. The user then clicks through the photographs in order of relevancy and gives comments. The technology provides more accurate retrieval results based on user comments on their relevance. In the future, picture retrieval will be improved by having the keywords that had a favourable response placed as a caption to the photographs. When applied to dynamic databases, this method is more accurate and saves time than manual annotation.

Automatic Annotation

The technique of automatically assigning tags to photos that convey the image's high-level semantic information is known as automatic image annotation. The supervised learning model is used in automatic picture annotation. A model is learnt in automated annotation by analysing a large database of annotated training photos. The semantic connections between the textual annotation and the semantic items described in the picture are included in the training set. The query picture's class is predicted by the annotation model, and appropriate textual phrases are added to the image. Annotation time is reduced while more general annotations are added to the picture using this method.

Text Based Image Retrieval

Text-based image retrieval systems, which date back to the 1970s, indexed pictures using keywords. Images are analysed by humans, who then give them the most fitting keywords that best characterise their subject matter and save them as part of the image's metadata. In the query phase, the user inputs search parameters in the form of one or more keywords. The pictures are obtained by performing a keyword matching procedure on the image database and selecting the image whose keywords match the search parameters. There is a challenge in text-based retrieval due to the subjectivity of determining the precise word or phrase that represents the meaning of the picture. Additionally, annotations are language-based, and the resulting variation in annotations will provide a significant obstacle to picture retrieval.

Content Based Image Retrieval

One alternative to text-based image retrieval that emerged in the 1980s is content-based image retrieval (CBIR). Images in CBIR are indexed in a database using the characteristics derived from them. The characteristics of the CBIR may be broken down into two groups: low-level and high-level. Colour, shape, texture, and location are all examples of low-level characteristics. The user initiates the inquiry phase by submitting a picture for consideration. It is assessed how close the query picture is to the photos in the database based on their shared low-level attributes. Similarity is used to rate the photos. The CBIR method does employ these low-level characteristics for picture retrieval, but it still can't compete with human vision. This is because there are not enough primitives to adequately represent the abstract ideas floating about in people's heads.

Semantic Based Image Retrieval

Due to the semantic gap between the low-level characteristics and the high-level picture ideas, the performance of image retrieval based on low-level features is not satisfactory. The semantic gap may be reduced by the careful selection of low-level characteristics that differentiate the pictures and the acquisition of fresh features that accurately depict the ideas' semantic meaning. Semantic concept retrieval employs a cognitive model of the human brain to translate between concrete visual characteristics and abstract ideas. Low-level picture characteristics are extracted first.

Research Motivation

After written documents, visuals are currently the primary way information is gathered and processed. This is because of the proliferation of mobile devices, which has led to a meteoric increase in the amount of visual content available online across private, commercial, and public domains. As a result, there has been a huge uptick in interest in digital photographs and the development of more efficient methods for storing, organising, and retrieving image data. Extremely difficult and of growing interest in both academia and business is the challenge of annotating, indexing, and retrieving enormous numbers of photos from large-scale image classes. Improving retrieval performance by precise labelling and indexing of photographs is a primary objective for this study. Low-level visual characteristics should be retrieved from pictures of large-scale image classes and utilised to train the system to annotate fresh photos in Automatic annotation based image retrieval systems. High-dimensional image feature vectors enhance the computational complexity of the annotation system for large-scale picture classes. As a result, there is a demand for effective dimensionality reduction techniques that may speed up processes and cut down on wait times. In order to quickly and accurately retrieve photos in response to a query, the annotated images must be effectively indexed. Therefore, effective indexing and retrieval techniques are required for processing the pictures of massive image classes. The author was inspired to go on with this study because of this reason.

Dimensionality Reduction

This component streamlines the picture characteristics by decreasing their dimensionality. Since the picture is often represented by a high dimensionality of characteristics prior to any processing, dimensionality reduction is essential. As a result, more space and processing time is required for the images. Therefore, it is crucial to reduce the sizes of the picture characteristics without losing the system's accuracy. In order to improve pattern recognition, the Autoencoder is used to compress the high-dimensional picture feature set.

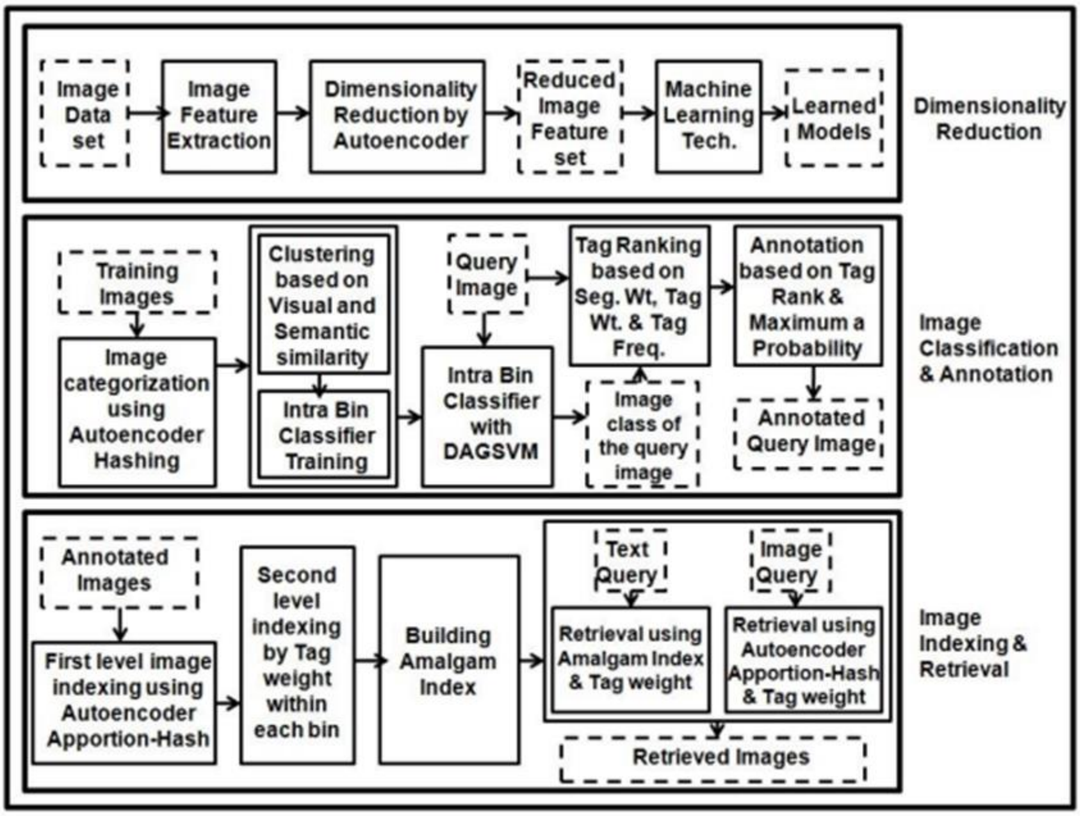

Figure 1 Image Annotation and Retrieval System

Image Indexing and Retrieval System

This section introduces an innovative system for cataloguing and locating photographs that fall under broad categories. For fast and accurate picture retrieval from huge image collections, image indexing is essential. To index the photos, we come up with the Autoencoder Apportion-Hash technique using tag weight. To accommodate both textual and visual queries, an Amalgam index is constructed to connect all of the picture keywords. In order to find the right pictures, a ranking system based on tag weights has been developed.

Image indexing using Autoencoder Apportion-Hash

Images from many different picture types may be efficiently indexed using the hashing technique. It is learnt which hash function and hash-code have a good distribution. Coordinate Descent is used to discover the hash-code for each picture category. Using an Autoencoder Apportion-Hash method, the hash function is taught to the system. In order to index images, their Micro-Structure Descriptors (MSDs) are converted into hash codes.

DOI: https://doi.org/10.17762/turcomat.v9i3.13917 Vol.09 No.03 (2018), 1245-1252

There are two steps involved in the process of converting an image's characteristics into a hash-code. The first step is utilising the Autoencoder to reduce the image's 72-dimensional Micro-Structure Descriptor to a 28-dimensional feature. In the second step, the Apportion-Hash technique is used to encode the 28-dimensional feature into a 7-bit hashcode. The Autoencoder is built using two encoder layers, two decoder levels, and a single code layer. The Autoencoder has 72 units total, 72 in the input layer and 72 in the output layer. Both the encoder and the decoder have one secret layer. The first two secret layers total 54 cells. There are 28 modules in the code layer. The coding layer's 28-dimensional MSD feature is broken down into 7 sections, each of which has a 4-dimensional feature. To train the hash function, we utilise the picture class's learnt hash-code as an endpoint. The hash-code stores the results of each division as a separate bit. The resulting 7-bit hashcode is then used to index the provided picture. Similarly, all photos belonging to large-scale image classes are catalogued into appropriate hash containers.

Experiments and Analysis

It is a computationally demanding and time-consuming process to train classifiers for large-scale picture classes, as well as to classify a new image instance and annotate it. An original method of picture annotation and categorization is presented in this paper. Both the training of the classifiers and the classification of the picture examples are expected to take less time using the suggested method. The performance of picture instance categorization, classification accuracy, and annotation are also intended to be enhanced.
Several tests have been carried out to prove the viability of our suggested method. For this purpose, we utilise the NUS-WIDE dataset and the MS-COCO 2017 dataset to test our methods. Classification performance is measured using Wang's dataset, PASCAL VOC2012, and MS-COCO 2017, whereas annotation performance is measured using NUSWIDE. Images from the NUS-WIDE dataset, the MS-COCO dataset, and Wang's dataset are shown as examples in Figure 2.

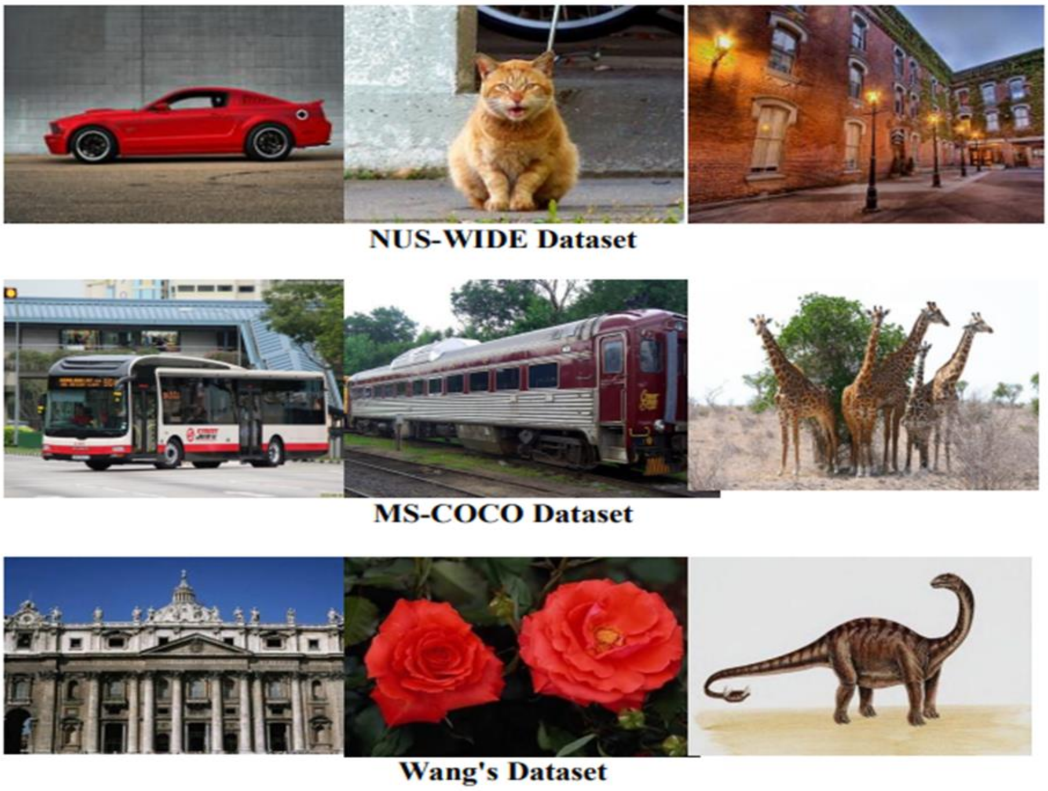

Figure 2. The Sample Images of NUS-Wide, MS-COCO and Wang's Dataset

DOI: https://doi.org/10.17762/turcomat.v9i3.13917 Vol.09 No.03 (2018), 1245-1252

Table 1. Comparison of MAP and Precision for Image categorization using the 8-bit hash-code

Method	MS-COCO 2017 dataset		NUS-WIDE dataset	
	MAP in %	Precision in %	MAP in %	Precision in %
SPLH	33.23	35.27	32.12	34.8
SDH	41.12	60.10	40.78	61.12
SVH	53.11	53.47	53.64	53.78
SPDH	54.17	57.24	59.43	60.21
Autoencoder Hashing	53.27	72.59	56.65	73.27

The results of a comparison between the MAP and the Precision of an 8-bit hash-code are shown in Table 1. When compared to other hashing algorithms, Autoencoder's accuracy stands out as particularly noteworthy. When compared to other techniques, the MAP performs well and is on level with SPDH. The suggested hashing approach outperforms the existing hashing methods because Autoencoder hashing has a lower time complexity compared to SPDH.

Comparison of the Retrieval Performance of the Proposed System with that of the Statistical Concept Model

Image Concepts	Proposed Method		Statistical Concept model	
	Precision%	Recall %	Precision %	Recall %
Beach	87.7	85.2	85.3	83.5
Water	93.2	88.5	92.6	93.2
Red flower	93.4	87.1	95.2	93.6
Sunset	94.6	87.2	95.8	86.4
Surfing	89.7	84.8	88.7	83.4
Waterfall	93.4	86.7	78.8	96.6
Tiger	88.2	84.3	76.7	82.5
Elephant	87.4	83.6	79.2	81.4
Penguin	87.7	86.7	82.6	84.5
Tree	91.4	89.2	88.4	89.3
Car	88.6	85.4	86.5	84.3
Grass	92.1	88.7	93.2	94.6
Waterway	92.3	85.8	86.3	85.7
Sky	91.6	86.8	94.6	95.3

Conclusion

More and more picture collections are being amassed as a result of the proliferation of social media and smartphones. These pictures need to be catalogued and stored away for later use. The present picture classification and annotation approaches are effective only for certain types and sizes of image sets. Therefore, there is room for development of methods to manage massive amounts of photos utilising the Big data idea, high speed computing, and hardware accelerators. After constructing the system, thorough testing is performed on its output. This study demonstrates that a large-scale picture class's classification, annotation, and retrieval accuracy has dramatically increased while the system's time complexity has decreased. Also, compared to the current system, the indexing system performs better when indexing photos from large-scale image classes.

References

1. Baopeng, Z, Hangzai, L & Jianping, F 2016, ‘Statistical modelling for automatic image indexing and retrieval’, Neurocomputing, vol. 207, pp.105-119.
2. Belkin, M & Niyogi, P 2003, 'Laplacian eigenmaps for dimensionality reduction and data representation', Neural Computing, vol.15, no. 6, pp. S1373-S1396.
3. Bo, Y, Xuequn, S & Shanmin, P 2017, ‘Isometric hashing for image retrieval’, Signal Processing: Image Communication, Elsevier, vol. 59, pp. 117–130.
4. Bolun, C, Ling, C & Yixin, C 2013, ‘Efficient ant colony optimization for image feature selection’, Signal Processing, vol. 93, no. 6, pp. 1566-1576.
5. Cabral, R, De-la-Torre, F, Costeira, JP & Bernardino, A 2015, 'Matrix completion for weakly-supervised multi-label image classification', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 1, pp. 121-135.
6. Chen, T, Lu, S & Fan, J 2017, ‘S-CNN: Subcategory-aware convolutional networks for object detection’, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 10, pp. 2522 - 2528.
7. Cheng, B, Zhuo, L, Zhang, P & Zhang, J 2014, 'Large-scale image retrieval based on the vocabulary tree', Proceedings of the VISAPP 2014, pp. 299-304.
8. Chih-Chin, L & Ying-Chuan, C 2011, ‘A user oriented image Retrieval System based on interactive Genetic Algorithm’, IEEE Transaction on Instrumentation and Measurement, vol. 60, no. 10, pp. 3318-3325.
9. Christos, D, George, S, Panagiotis, P, Christos, P, Nikos, D & Anastasios, D 2010, 'Large scale concept detection in multimedia data using small training sets and cross domain concept fusion', IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 12, pp. 1808-1822
10. Cilibrasi, R & Vitanyi, P 2007, ‘The Google similarity distance’, IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 3, pp. 370–383.
11. Dong, P, Mei, K, Zheng, N, Lei, H & Fan, J 2012, ‘Training interrelated classifiers for automatic image classification and annotation’, Pattern Recognition, Elsevier vol. 46, no. 5, pp. 1382-1395
12. Fan, J, Gao, Y & Luo, H 2008a, ‘Integrating concept ontology and multitask learning to achieve more effective classifier training for multilevel image annotation’, IEEE Transactions on Image Processing, vol. 17, no. 3, pp. 407-426
13. Fan, J, Gao, Y, Luo, H & Jain, R 2008b, ‘Mining multilevel image semantics via hierarchical classification’, IEEE Transactions on Multimedia, vol. 10, no. 2, pp. 167–187
14. Galar, M, Fernandez, A, Barrenechea, E, Bustince, H & Herrera, F 2011, ‘An overview of ensemble methods for binary classifiers in multi-class problems: an experimental study on one-vs-one and one-vs-all schemes’, Pattern Recognition, Elsevier, vol. 44, no. 8, pp. 1761-1776
15. Hanjiang, L, Yan, P, Ye, L & Shuicheng, Y 2015, 'Simultaneous Feature Learning and Hash Coding with Deep Neural Networks', Proceedings of the CVPR2015, pp. 3270 -3278
16. Hao, M, Jianke, Z, Michael, R & Irwin, K 2010, ‘Bridging the semantic gap between image contents and tags’, IEEE Transaction on Multimedia, vol. 12, no. 5, pp. 462 - 473.

DOI: https://doi.org/10.17762/turcomat.v9i3.13917 *Vol.09 No.03 (2018), 1245-1252*

17. Hsu, CW & Lin, CJ 2002, 'A comparison of methods for multiclass support vector machines', IEEE Transactions on Neural Networks, vol. 13, no. 2, pp. 415–425
18. vasic-Kos, M, Pobar, M & Ribaric, S 2016, 'Two-tier image annotation model based on a multi-label classifier and fuzzyknowledge representation scheme', Elsevier-Pattern Recognition, vol. 52, pp.287–305
19. Jason, KG & Adriaan, VN 2017, 'Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning', Computers and Electronics in Agriculture, vol. 142, pp. 50-58
20. Jing, XY, Wu, F, Li, Z, Hu, R & Zhang, D 2016, 'Multi-label dictionary learning for image annotation', IEEE Transactions on Image Processing, vol. 25, no. 6, pp. 2712-2725.