IMAGE ANNOTATION USING SEMANTIC SPARSE RECODING OF VISUAL CONTENT

Surender Reddy S [1], Dr.Neeraj Sharma[2], Dr.Y. Mohana Roopa[3]

[1]Research Scholar, Dept. of Computer Science and Engineering

Sri Satya Sai University of Technology and Medical Sciences,

Sehore Bhopal-Indore Road, Madhya Pradesh, India.

[2]Research Guide, Dept. of Computer Science and Engineering

Sri Satya Sai University of Technology and Medical Sciences,

Sehore Bhopal-Indore Road, Madhya Pradesh, India.

[3]Research Co-Guide, Professor. Dept. of Computer Science and Engineering

Institute of Aeronautical Engineering, Dundigal, Hyderabad

ABSTRACT

For semantic segmentation, image datasets with high-quality pixel-level annotations are advantageous: labelling every pixel in an image ensures that uncommon classes and small objects are tagged. Systems for Content-Based Image Retrieval extract and retrieve images based on their low-level characteristics such as colour, texture, and shape. However, these visual elements do not enable users to search for photographs based on their semantic information. Tags are added to an image collection for both training and testing purposes. Our training strategy entails developing classifiers and kernels that account for the similarity of visual features and tags. We implement an automatic semantic segmentation and object recognition methodology that bridges the semantic divide between low-level image content attributes and high-level conceptual meaning. Semantically comprehending an image is critical for modelling autonomous robotics, marketing, and reverse engineering building information modelling in the construction business.

Keywords : Image annotation, Content-based image retrieval, Semantic segmentation.

INTRODUCTION

Although the content-based image retrieval paradigm has been studied for over a decade, it has not been universally recognised as successful. Two factors, we believe, have hampered

user adoption of content-based image retrieval. To begin, the technologies involved, such as feature extraction, indexing, and query processing, all require significant breakthroughs. Second, consumers like to specify queries using keywords. Effective image annotation is crucial to enabling keyword search of images. We address the issue of image annotation in this work. Traditionally, image annotation has been viewed as a machine learning task composed of two key components: feature extraction and feature mapping to semantic labelling. Feature extraction is a technique for extracting meaningful signals from photographs. Following that, signals are mapped to keywords using a machine learning system. We begin this article by describing our overall annotation framework. The issue then shifts to our educational infrastructure: Dirichlet Allocation using Latent Dirichlet. Our work is distinct from conventional methods in two ways. To begin, our system takes a synergistic approach to perceptual features and user logs. Second, our technology operates on parallel processors and is capable of processing data in the millions. Our preliminary analysis in this publication demonstrates that our paradigm is promising. To achieve high retrieval speeds and to truly scale the retrieval system to enormous image collections, an effective multidimensional indexing module is a critical component of the overall system. This module will be advanced by the communities of computational geometry, database management, and pattern recognition. The retrieval engine module is composed of two submodules: a query interface and query processing. The query interface is graphic-based in order to connect with the user in a pleasing manner. The interface gathers the necessary information from the users and presents the retrieval results to them in a relevant manner. The advancement of user psychology and user interface research contributes to the improvement of interface design. Additionally, the same user inquiry can be processed in a variety of ways. The query processing submodule parses the user query and routes it to the most appropriate processing methods.

The database management community has advanced these strategies. Content-based image retrieval is founded on three key principles: visual feature extraction, multidimensional indexing, and retrieval system architecture. The remainder of this article will discuss numerous visual aspects and their associated representation and matching strategies. It evaluates several dimension reduction and multidimensional indexing approaches. Commercial and research systems that are state-of-the-art, as well as their distinctive properties. To address these issues, it was proposed to use content-based image retrieval. That is, rather than being manually labelled with text-based key phrases, photos would be

indexed according to their inherent visual information, such as colour and texture. Since then, numerous methodologies in this research area have been created, as well as numerous image retrieval systems, both scientific and commercial.

On the basis of the current state of the art and the requirements of real-world applications, promising future research directions and suggested methodologies are described, along with some concluding observations. The computer vision community has made significant contributions to this research direction. Numerous special issues of prestigious publications have been devoted to this subject. This approach has provided a new framework for image retrieval. However, numerous unresolved research concerns must be resolved before such retrieval systems can be implemented. Concerning content-based image retrieval, it is necessary to assess what has been accomplished in recent years and to identify new research directions that could result in compelling applications. Because comprehensive surveys of text-based image retrieval paradigms already exist, we shall focus exclusively on the content-based image retrieval paradigm in this study. Reranking images is a powerful strategy for optimising the results of web-based image searches. Existing industrial search engines such as Google, Bing, and others have implemented it. The search engine re-ranks the collection of photos based on the query keyword. The user then selects single image from the set, and the other images are re-ranked in relation to the user-selected image. The exponential growth of digital photos on the web necessitated the development of the best image retrieval technique capable of increasing the retrieval accuracy of the images. As a result, the research focus has turned away from the development of complex algorithms capable of closing the semantic gap between visual characteristics and the richness of human semantics. As a result, numerous image reranking algorithms have been developed to improve text-based image results by utilising visual information present in the images.

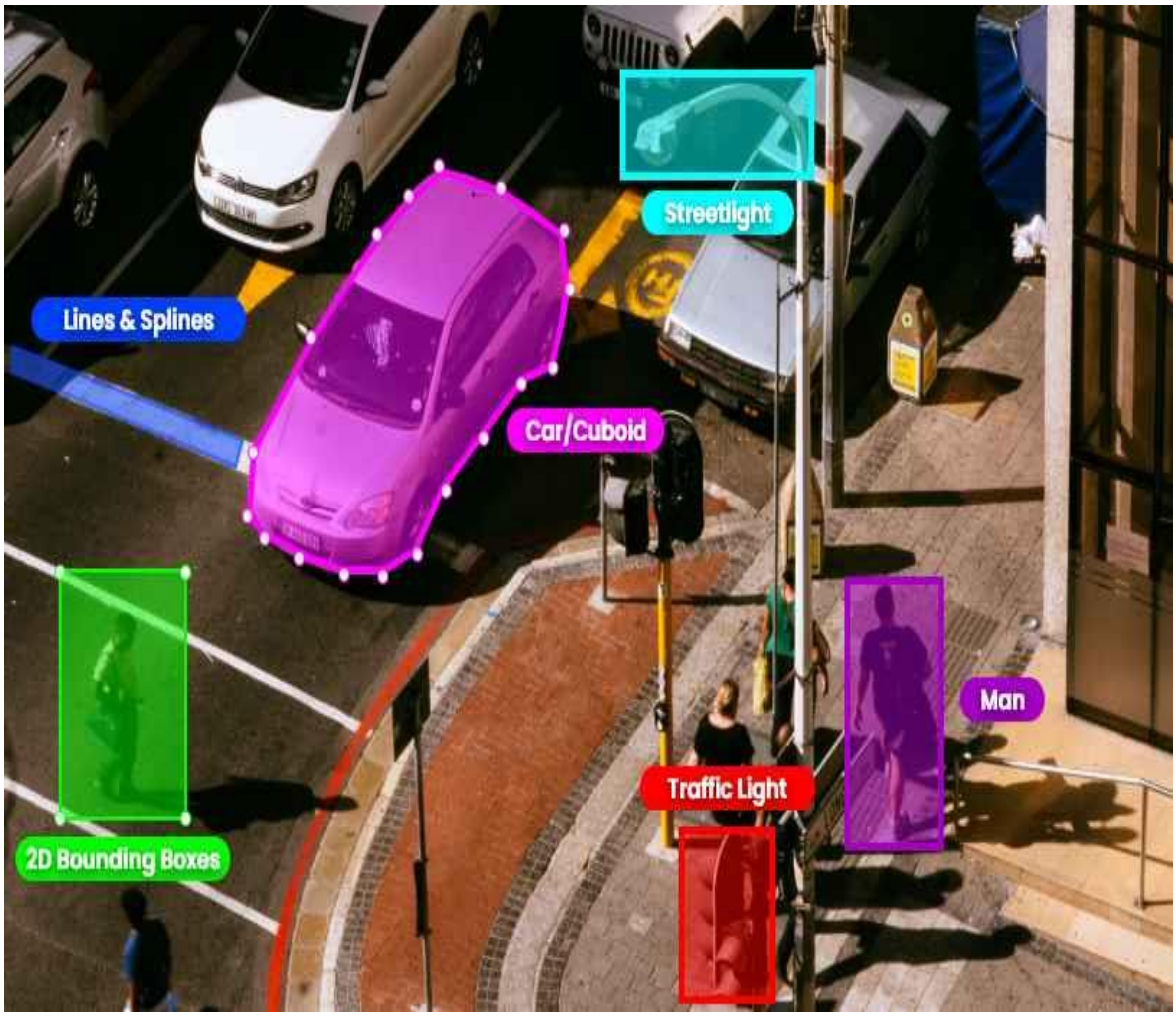

Figure 1 Image Annotation

Image search is a rapidly developing function of popular search engines such as 'Google,' 'Yahoo,' and 'Bing,' among others. For each text query, the search engine must search millions of images in order to return the most relevant images as quickly as possible. In general, search engines rely on text meta-data, such as keywords, tags, or text descriptions adjacent to images. Web image reranking is a procedure in which photos are retrieved and sorted according to their characteristics and user requirements. To make it easier for users to access the continuously growing collection of photos on the Web and to maximise their utility, image search has grown in importance as a study issue. Web-scale image search engines mostly use keywords as queries and index photos based on their surrounding text. In general, an image search engine runs in two stages: offline index generation and online index serving. Effective online search requires the retrieval of meaningful and significant images. In image retrieval, uncertainty occurs when established techniques are used. Oftentimes, users have difficulty adequately defining the visual content of target images

using only keywords. Due to the absence of query and visual features, the photos obtained are less relevant to the user's requirements. Then, by re-ranking the initial batch of returned photos, the precision and accuracy can be increased. In recent years, consumers' efforts have been minimised by online image re-ranking methods that require only one-click feedback to improve search results.

LITERATURE REVIEW

Suvarn et al., (2020) Numerous search engines, such as Bing, Google, Cydral, Yahoo, AltaVista, and Ask, are available today that cater to image search requests. Users express their requirements through search engines, and the resulting photos are shown. Two strategies are utilised to improve search results: image annotation and web image search re-ranking. With the ever-increasing volume of digital photographs available on the Internet, finding suitable images from a diverse variety of databases has developed into a significant research project. Numerous image retrieval systems have been created during the last few years, including text-based image retrieval, content-based image retrieval, and a hybrid approach. Due to the fact that meta-data are not always associated with the visual term of the photos, retrieval of images is frequently combined with irrelevant images. Nonetheless, it has been determined that the retrieved images contain sufficient relevant images and are structured for users who are more concerned with precision than recall.

Seema et al., (2018) In document and image search, text-based search is more successful and efficient. The user enters a text query. These text-based queries can be expressed in free text and compared to text descriptors such as description, subjects, title, or the text surrounding an embedded image utilising text retrieval algorithms. Text data contained within multimedia files contain valuable information that can be used for automatic annotation and indexing. The primary issue and challenge with web image search is the disparity between image content and web page language. The TBIR has been extensively utilised in famous image search engines including as Google, Bing, and Yahoo! When a user submits a textual query to the retrieval system, the system returns ranked images whose adjacent texts contain the given query keyword, with the ranking score determined by the degree of similarity and correspondence between the user's query keyword and the textual features of relevant images. While text-based search strategies have been demonstrated to perform effectively on textual materials, they frequently produce inconsistent results when applied to query image searches due to the metadata's inability to

reflect the semantic content of photos. However, focusing exclusively on global attributes obscures the spatial placement and orientation of objects in photographs.

Citation: Yang et al (2015) Although the majority of search engines use text-based approaches, there are alternatives. Content-based image retrieval (CBIR) extracts visual aspects such as colour, texture, and shape from images automatically and compares them using feature distances. Here, implementation is straightforward and retrieval is lightning quick. Appropriate feature representation and a similarity metric are required for ranking photos in response to a query. As a preprocessing phase, the majority of content-based image retrieval algorithms extract features from images. It demonstrates the CBIR technique. Textual features such as key words, annotations, and tags may be combined with visual features such as colour, texture, shape, and faces. Image retrieval based on content, which needs the user to submit a query image and returns photos that are comparable and related in content. Google is one of the search engines involved in this process of image reranking. This method extracts natural and objective visual information, but completely disregards the function of human comprehension or knowledge in the interpretation process. The selection of features is a significant impediment to content-based image analysis. Despite the fact that numerous features have been presented over the years, none have come close to bridging the semantic divide between an image's low-level visual content and what people perceive as the image's high-level semantic.

RESEARCH METHODOLOGY

The purpose of this project is to establish a computationally efficient framework for object identification in commercial building environments. The purpose of this dissertation is to combine data in order to maximise accuracy and efficiency by utilising both types of approaches. A image is composed of items and structures. Segmentation is a critical step in this methodology, and any error in this step has a detrimental effect on semantic image processing. As a result, we created a method for capturing and labelling perceptually significant sections, which frequently reflect the image's global representation and comprehension.

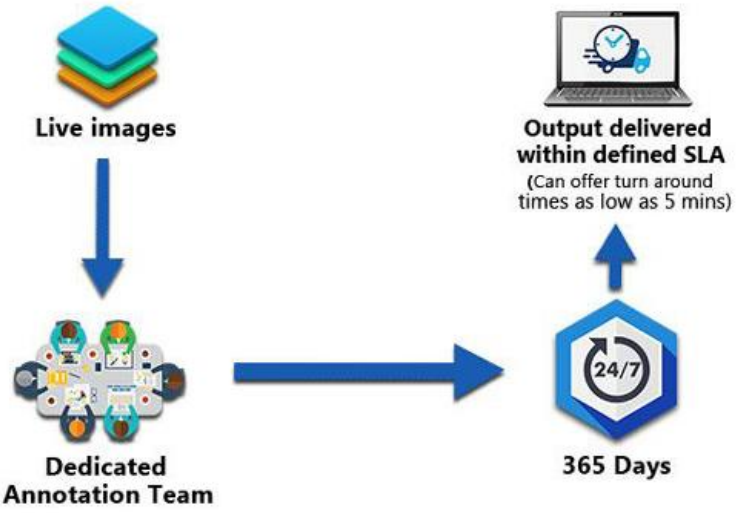

Figure 2 Framework Architecture

The following issues were addressed throughout the development of our model:

- Creation of a novel segmentation method based on the generation of features for each pixel. For each pixel, the feature vector is a mix of graph-based segmentation, the surface index, and the feature vector.
- Object recognition via supervised classification algorithms, with each segment including a unique collection of features.
- Creating a scene detection model based on global features such as colour, texture, and a gradient's local histogram.
- Developing a model for merging evidence from several learning models.

This model incorporates data from the scene, objects, and commonsense knowledge.

On photos from indoor and outdoor building sites, this dissertation employs an image understudying approach. To our knowledge, no corpus has photographs of this type, and hence we developed our own dataset for scene detection. A dataset of 764 photos of classrooms, bathrooms, computer laboratories, corridors, and the outdoors was collected for scene detection. Each of these photographs was annotated and assigned a scene tag. Two annotators completed the annotations manually, and because the scenes were distinguishable by humans, there was 100 percent agreement between the annotators.

Additionally, there is no image dataset with RGBD information of indoor locations for object detection.

To our knowledge, only video files with RGBD data are available, which are unsuitable for this research due to the data being skewed and redundant. As a result, training and testing corpora have to be constructed manually. A collection of 200 single object photos was collected for the object recognition test. The collection contains photos of chairs, tables, couches, garbage cans, computers, sinks, and toilets, as well as automobiles. The photographs were recorded with a Kinect sensor and then processed in Photoshop to remove the backdrop, leaving only one object on a white background. To aid with classification, a "unknown" class was also developed utilising randomly coloured, scaled, oriented, and shaped items.

We propose to conduct multi-label learning in two spaces concurrently in order to improve feature representation in the input feature space and label propagation in the output label space. We present a unique technique to image annotation, namely multi-label dictionary learning with regularisation for label consistency and partial-identical label embedding. We include the dictionary learning technique into multi-label learning in the input feature space. Dictionary learning seeks to construct an exhaustive dictionary from the space of training images in which the supplied input signal is well represented. Additionally, we design the label consistency regularisation term to learn a discriminative lexicon. Thus, in order to generate a robust and discriminative representation of features in the input feature space, we construct the partial-identical label embedding, which is based on the samples' advantageous collaborative representation capabilities while using partial-identical label sets. It enables clustering of photos with same label sets and collaborative representation of images with partial-identical label sets. We develop a joint objective function that enables simultaneous multi-label dictionary learning and partial-identical label embedding.

DISCUSSION & RESULTS

Annotations on blocks serve as both hints and targets. This means that training the block-inpainting model requires no new data (or human annotation labour). We conduct our studies using (synthetically generated) Block50% annotations. Half of the annotated blocks in each image are randomly selected online during training to serve as suggestions. Each annotated block is used as a target. This enables the network to "paste" suggestions into

the final output while using the hints as context to inpaint labels for regions without hints. The block-inpainting model generates labels with a human-agreement level comparable to that of human annotators.

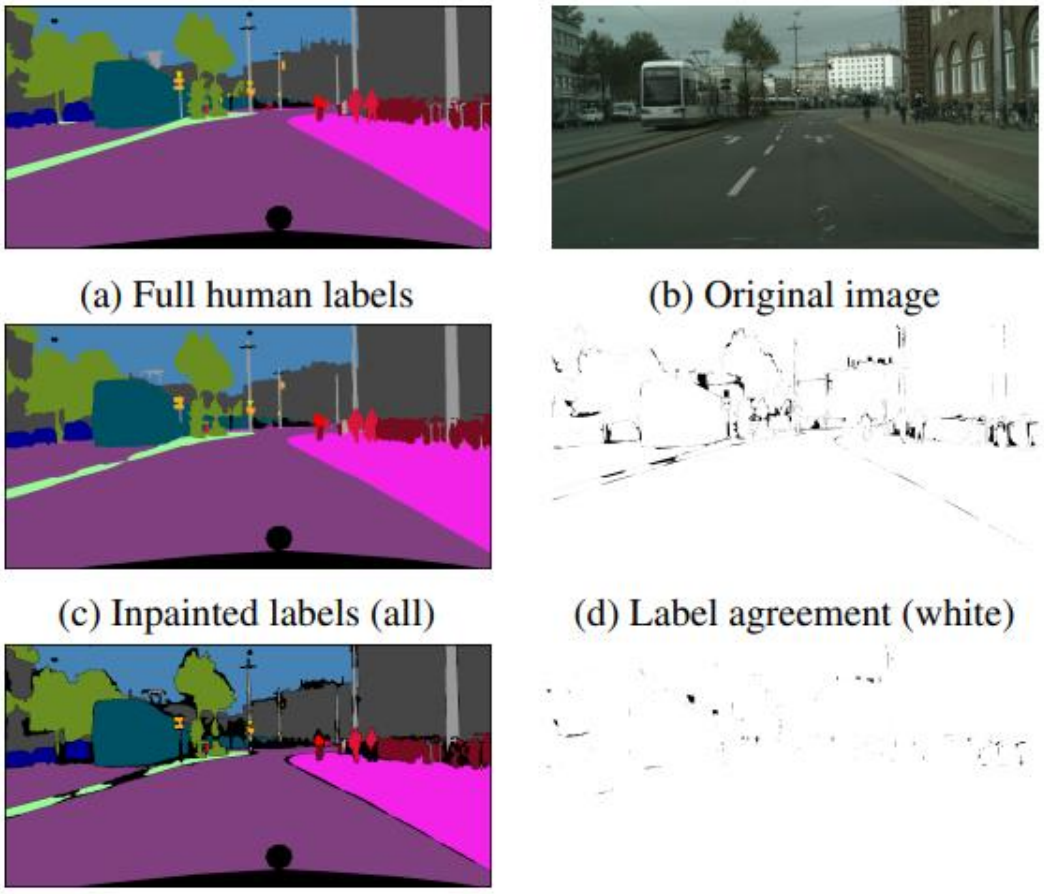

Figure 3 Results of Image

In this experiment, we inpaint Block-50 percent annotations. On Cityscapes (ADE20K), almost 94 percent of the pixels are labelled at a relative uncertainty level of 0.2 (0.4). The mean agreement between pixels is 99.8% (98.7%), while the class-balanced error rate is 3.1 percent (28 percent). Previous research indicates that human label agreement between annotators is between 66.8 and 73.6 percent, while annotator self agreement ranges between 82.4 and 97.0 percent. The failure of human annotators to agree in a non-trivial manner demonstrates that annotator self-agreement fails in three ways: variances in complex boundaries (32%), improper naming of ambiguous classes (34%), and failure to segment small objects (34 percent). Figure 7 illustrates the labels generated by the block-inpainting approach. With a higher uncertainty threshold, the frequency of pixel conflicts reduces.

CONCLUSION

A novel approach to content-based image retrieval based on database categorization and the Density-Scale Invariant Feature Feature selection for transform and shape, texture and colour string coding. In summary, our contribution to the ever-growing field of image databases is an effective method for retrieving photos from massive databases. Two primary considerations were made. The retrieval procedure must be novel and participatory. The method's resilience is critical when there is some degree of noise present, as there may be in the case of basic photographs. Without change, the drawn image cannot be compared to a colour image or its edge representation. Alternatively, a step for distance transformation was added. The effectiveness of the Text Based Image Retrieval System and the implementation of the dynamically parameterized Content Based Image Retrieval System were compared during the tests. It was evaluated using additional databases. In our experience, content-based image retrieval performed significantly better than text-based retrieval in the majority of circumstances. By utilising database classification, we can significantly increase the performance of content-based image retrieval when compared to non-classified content-based image retrieval. Finally, this database classification and feature selection for shape, texture, and colour string coding will produce the best results.

REFERENCES

1) Suvarn V. Jadhav A. M. Bagade, Department of Information Technology, Pune Institute of Computer Technology, Pune,(India), "Comparative Analysis Of Reranking Techniques For Web Image Search", International Conference on Recent Innovations in Engineering and Management, Dhananjay Mahadik Group of Institutions (BIMAT) Kolhapur Maharashtra, 23rd March 2020.
2) Seema Ranuji Ghule, “A Survey on Multimodal Visual Search Methods”. International Journal of Science and Research (IJSR), Pune Institute of Computer Technology, Savitribai Phule Pune University, India. 24(6): vol. 30, no. 11, pp.1877–1890, Nov. 2018.

3) X.Yang, X. Qian, Y. Xue "Scalable Mobile Image Retrieval by Exploring Contextual Saliency". IEEE Transactions on Image Processing 24(6): 1709-1721 (2015).

4) Y. Wang, X. Lin, L. Wu, and W. Zhang, "Effective Multi-Query Expansions: Collaborative Deep Networks for Robust Landmark Retrieval," IEEE Trans. Image Process., vol. 26, no. 3, pp. 1393–1404, Mar. 2017.

5) L. Ballan, M. Bertini, G. Serra, and A. Del Bimbo, "A data-driven approach for tag refinement and localization in web videos," Comput. Vis. Image Underst., vol. 140, pp. 58–67, Nov. 2015.

6) T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "NUS-WIDE: a real-world web image database from National University of Singapore," in Proceeding of the ACM International Conference on Image and Video Retrieval - CIVR '09, 2009.

7) F. Tian, X. Shen, and X. Liu, "Multimedia automatic annotation by mining label set correlation," Multimed. Tools Appl., vol. 77, no. 3, pp. 3473–3491, Feb. 2018.

8) Maihami V, Yaghmaee F. Automatic image annotation using community detection in neighbor images. Physica A: Statistical Mechanics and its Applications. 1;507:123-32, 2018.

9) Tian and Z. Shi, "Automatic image annotation based on Gaussian mixture model considering cross-modal correlations," J. Vis. Commun. Image Represent., vol. 44, pp. 50–60, Apr. 2017.

10) K. Akhilesh and R. R. Sedamkar, "Automatic image annotation using an ant colony optimization algorithm (ACO)," in 2016 IEEE 7th Power India International Conference (PIICON), 2016, pp. 1–4.