Image Object Detection & Annotation Techniques

Shivam Agrawal, Shubham Jha, Satyajeet Salunkhe, Sanjeet Jha
Final Year, B. Tech CSE, Dr. Vishwanath Karad MIT World Peace University, Pune, India
Professor - Prakash Waghamode ,
Dr. Vishwanath Karad MIT World Peace University, Pune, India

----------***----------

Abstract - In this world, and the images, are an important piece of information. An image can tell you much more about a subject matter, or thing, than textual information.Recently, research has been conducted on the use of an image on the basis of the information. With the help of the image search query in the text, it is an old method of. Image search is a newer method that does not have a full list of all of the text. This method is useful for finding an exact copy of the original image.Object detection, there are a lot of factors, such as image recognition, and image creation. Object detection is related to computer vision and it is widely used fordetet objects in certain classes. We are able to detect an object in the image of the object on the surface of some of the rectangular frames, and the identification of each and every object and give it its own tag. This also applies to the correctness of each of the object's authentication methods. The world's leading manufacturer in this field is Google. Most recently, Google has changed its photo-storage service, and the creation of a new service called Google Photos, the operating system is a sophisticated automatic semantic analysis of the techniques for photos. This will result in the automatic classification / classification / annotation / labelling of the photos based on their content. It turns out that the categories that have been created with a high degree of accuracy, in line with expectations, but the specificity is low, as expected. Searching for images using the images, it is also very popular.

The pictures on the Internet are exponentially growing fast, and it can act as a very important source of information. But information discovery is still very largely based on textual search. People search for a label associated with images to discover it. In recent years information discovery for images has seen significant gains. Searching for images using images has also become considerably popular. Reverse image search helps us in finding similar images which are present on the internet.

Keywords:***Google Photos,computer vision,Reverse image***

1. INTRODUCTION

1.1 Increasing number of images on internet

- Usage of the internet is increasing day by day and people uploading images on the internet is growing heavily. There are a number of factors, which have been added to the exponential growth of those pictures, all of which are primarily related to the new, easy-to-use technology.
- The quality of the photos from the mobile cameras and professional digital DSLR cameras has been increased by more than 100 pixels in each frame and shows off more of the features.
- Speed of the internet is the most important factor because people upload a large amount of image data on the internet in very less time.In today's anyone can access fast internet.
- With the advent of smartphones and their price-to-quality cameras, the availability and accessibility of digital images have increased dramatically. And thank you as well for the spare wheel in the cloud, you can use this image online, and without having to worry about the size of the image.

Users don't have to worry about storing pictures, as they are accessible to all, with services such as Google Photos, and Google's Cloud Drive, Mega etc. It also allows you to edit images and share them easily online in less time.

1.2 *Organization and annotation of images*

Automatic annotation-is an important area of computer vision is playing an important part in the pictures, the description of the picture.Image annotation, it is a process by which a computer program to automatically generate the metadata is in the form of captions or keywords to a digital image.Automatic image annotation is a problem related to multi -label classification which aims on at associating a set of textual with an image that describe its semantics.

As a result, we can see that the pictures on the Internet, and are well-organized, and be saved from the user's point of view, because this is the Internet, and the possibilities are at a relatively high level, and high level

technology, software, and networks to ensure reliable storage and fast transmission of data, information, or graphics of the client-to-server, on a worldwide scale, for a variety of users, with the exception, of course, users will still not be able to afford constant and convenient access to all of these technologies.

1.3 Searching for images with keywords.

Most users use search engines by writing text queries and consuming text results.While in general each image has specific keywords which are used to find that image and images related to it.Image classes deal with the task of associating multiple labels with single data. It is a difficult problem becauses one needs to consider the intricate correlations between different class labels. Machine learning and neural network techniques help to search images with keywords.

1.4 Object Detection in Images

Object detection means an application of Computer Vision to identify the classes of objects in a digital image. Various algorithms such as Feature matching , template matching , Different Classifiers such as Haar cascades are created and used for that purpose.

First a dataset is prepared with bounding boxes and labels assigned to the observed object. Then the model is trained on that dataset and the to be detected features are extracted from the image. After training the model is tested against the test images and accuracy of the detection is calculated.

Fig 1. Diagram of object detection workflow

In deep learning based object classification such as RCNN we take an input image and divide 2000 different regions in an image. After that convolutional neural network extracts useful features from those 2000 regions. In the next step a linear classifier such as SVM classifies each of the regions into classes of image.

In another approach known as Faster RCNN we give the model an image and several Regions of Interest as input. These go into a convolutional neural network.

Then the pooling layers use max-pooling to convert the features inside regions of interest into small feature maps of a particular height and width. These are then mapped to a feature vector by fully connected (FC) layers.

Finally, the network outputs two vectors per region. One is the softmax probabilities from which we can obtain the image classes. The other one is per-class bounding box regression offsets. This second method is found to be more accurate in terms of detection on a random image.

2. METHODOLOGY

2.1 Uploading images of different classes to google photos

For the classification of images we needed a dataset consisting of multiple classes. We created our own dataset by compiling images from multiple random sources like kaggle, google images, Reddit etc.

We uploaded images of each class in bulk.

2.2 Searching of classes with their synonyms

After uploading, Google photos automatically classifies images into different classes and stores as metadata.

We find the synonyms of each classes which we could use to identify the uploaded class, for eg. “selfie” can be synonymous with face, human, people etc.

Then we type in the search bar the keywords we want to search by. Each keyword corresponds to a specific class and is a synonym or grammatical number of a particular class.

Fig 2. Screenshot of querying using the keyword selfie.

2.3 Evaluation of classes

We evaluated the quality of automatic classification. For each image uploaded to Google Photos, each image that was classified in the class was checked whether it was correct or not, and also the image which was not detected and we made a list of all the images that were detected and the percentage of images google photos was not able to detect properly.

The classification takes quite a while, as we noticed that it is not an instantaneous process. It takes a few minutes

if we do a bulk upload with upto 100 images. It takes roughly 5-7 minutes if we do a bulk upload of 500 images for all the images to be classified into their respective classes.

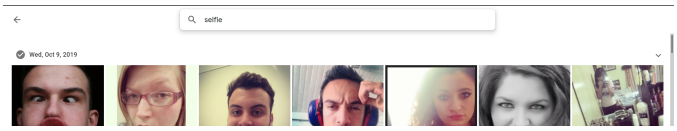

Fig 3. Search results using a keyword.

3. RESULTS

Uploading 570 images of "selfie"

Keywords used in search	Precision (in numbers)	Precision (in %)
Selfie	533/570	93.5
Face	498/570	87.36
Human	121/570	21.22
People	2/570	0.3

Uploading 111 images of "buildings"

Keywords used in search	Precision (in numbers)	Precision (in %)
Building	99/111	89.18
Building	99/111	89.18

Uploading 266 images of "Animal"

Keywords used in search	Precision (in numbers)	Precision (in %)
Animal	224/266	84.21
Animals	255/266	95.86

Uploading 359 images of "Bird"

Keywords used in search	Precision (in numbers)	Precision (in %)
Bird	352/359	98.05
Birds	351/359	97.77

From the above result we can see in Google photos singular and plural of the same word have different precision percentages. Also google photos differentiate in classes like selfie and humans which should be synonymous to each other for object detection.

5. CONCLUSION

Based on experimental results we are able to detect objects more precisely. Save it, and upload it is much easier, due to the uploading of the images in the library. Users will be able to organize and categorize your photos by albums. The main conclusion is that the accuracy of object detection is very high and it is growing very rapidly, it has the potential to automatically organize photos into albums in the digital collection, as well as the quality of the information of the user, detecting, and this may provide an additional order, in addition to the manual in order to be performed.

6. REFERENCES

1. Bernardi, Raffaella, Ruket Cakici, Desmond Elliott, Aykut Erdem, Erkut Erdem, Nazli Ikizler-Cinbis, Frank Keller, Adrian Muscat, Barbara Plank. "Automatic Description Generation from Images: A Survey of Models, Datasets, and Evaluation Measures." Journal of Artificial Intelligence Research 55, 409-442, 2016. See also arXiv:1601.03896v2 [cs.CL] and https://www.jair.org/media/4900/live-4900-9139-jair.pdf
2. Davi Oliveira Serrano de Andrade, Luis Fernando Maia, Hugo Feitosa de Figueirêdo, Windson Viana, Fernando Trinta, Cláudio de Souza Baptista, "Photo annotation: a survey." Multimedia Tools Appl., 2016 Springer, DOI 10.1007/s11042-016-4281-6.
3. Marques, Oge, "Visual Information Retrieval: The State of the Art." IT Professional, Volume: 18, Issue: 4, DOI: 10.1109/MITP.2016.70 2016 (accessed in 2016).
4. Nieuwenhuysen, P., "WWW image searching delivers high precision and no misinformation: reality or ideal?", Issues in Informing Science and Information Technology, Vol. 7, pp. 109-131. ISSN: On-line 1547-5867; CD 1547-5859; Print 1547-5840, peer reviewed journal, available online

free of charge from http://iisit.org/Vol7/IISITv7p109-131Nieuwenhuysen770.pdf 2010. and from http: //www.vub.ac.be /BIBLIO /nieuwenhuysen/presentations/

5. Nieuwenhuysen, P. “Search by Image through the WWW: an Additional Tool for Information Retrieval.” Full text published in proceedings of the international conference on Asia-Pacific Library and Information Education and Practices = A-LIEP 2013 “Issues and challenges of the information professions in the digital age” held at Pullman Khon Kaen Raja Orchid Hotel, in Khon Kaen City, Isan, Thailand, 10-12 July 2013 [online] http://aliep2013.com/index.php/table-of-contents, http://aliep2013.com/images/download/pdfs1/PaperNo41.pdf, and also available free of charge from http://www.vub.ac.be /BIBLIO /nieuwenhuysen /presentations/