Image Annotation Framework

Avinash R,Apoorva G,GuruShankar SG

***Abstract*—In the recent years Image Annotation has picked up significance in the process of searching, retrieving, computer visions and applying labels or tags for image. For accomplishing the most advantage, the humans observation and description can rewarded from various frameworks. The idea generated is to prearrange the raw image as input applied to a proposed model that can help in recognition of each object in the image that can be further used for placing of appropriate tags using machine learning technique known as Convolution neural network. In this context, a suitable framework will be designed and developed that learns vigorous scene models using noisy tags and pictures from several partially annotated datasets. With reference to the above mentioned details, the model is proposed that will clearly demonstrate the presence of rightness within the structure via annotation using labels (tags) and portioning (segment) pictures from different classes portraying to the sport scenes. The proposed work accomplishes two major tasks: Initially, to feature extract from the image to identify the objects and later to explain with the labels with the comparable area,such that the expectation of the overall model will altogether exceeds in cutting edge calculations.**

.

Keywords*—*Image Annotation,Image Classification, Object Detection,

I. Introduction

When an image/photograph is shown to human being, with his understanding and with the prior knowledge one can identify the objects present in the image just by looking, but when it comes to computer it's difficult to identify the objects present in an image. So our goal is to make computer vision to identify the objects present in the scenario and classify them into appropriate classes. To classify objects in an image into appropriate class we are using Deep Convolutional Networks to develop a framework that can identify an object's and annotate them with appropriate tags/class label. The developed framework will be able to detect an object in an image either manually or automatically.

A. Motivation

With the development of the Internet and digital imaging devices many large image collections are being created. Popular online photo-sharing sites like Flickr contain hundreds of millions of diverse pictures. Many organizations, e.g. libraries, hospitals, governments and commerce have also been creating their large image databases by scanning paintings, manuscripts, prints and drawings. Searching and finding large numbers of images from a database is a challenging problem. Search engine they do not really capture the semantics or meaning of images well. For image retrieval systems based on text queries, the key problem is how to get the metadata such as captions, titles or transcriptions. Manual annotation is not practical for large volumes of image sets. Commercial image search engines for the Internet, e.g. Google image and Yahoo image, use the text surrounding each image as its description. However, these search engines entirely ignore the visual content of the images and the surrounding text doesnt always relate to the visual content of an image. The consequence is that the returned images may be entirely unrelated to users needs.
Vision is the richest sense that a human being has which computer does not have and will consist of a tedious process to achieve the same for a computer. Object recognition and classification play a major role in this field.

B. Scope

Image annotation is a complex job of detecting objects and classifying each objects in a given image. Even though the process is extremely useful in some cases, the complexity of the process limits many novice developers from using image annotation and object detection in their projects. So we are developing a user friendly framework that will do simplest of the image annotation tasks and help novice developers in their projects that might need object detection and classification.

C. Challenges in Image Annotation

Image Annotation is extremely useful in many areas of computer vision such as MultiMedia understanding, machine learning, image processing and analysis, and also in the field of querying and retrieving the proper information. When we try to analyze an image and annotate them with the caption, it's usually done with feature vector calculation and will be required to train a lot of training model to predict words/tag/words using machine learning technique to predict. Annotation of new image will be possible only after training and learning of the model. The task of scene understanding and object recognition for semantic prediction is challenging work. In Image Annotation Process we deal with recognising multiple object.

D. Applications

Following are the application of Object Recognition and Annotation for an image:

1) Android Eyes Object Recognition: Its an advanced object recognition application. If we take picture of an object then machine eye can tell what it is.
2) Automated vehicle parking systems: it is designed to reduce volume or area required for parking cars.
3) Optical Character identification: it is the conversion of image typed,handwritten in to encoded text.

4) Content-Based Image Indexing: Is searching for digital images in huge database. Search analyses the contents of the images.
5) Image watermarking: It is the method of smacking the digital information. Its used to find ownership.
6) Global robot localization: It is constructing and updating a map of environment which is unknown also simultaneously keeping track of the location of the agent.
7) Face detection: Used to recognize human faces. Using this process person locate and attend to faces in a scene.
8) Video Stabilization: helps in cancelling our moves while we are recording a movie.
9) Manufacturing Quality Control: It ensures that our products are safe, pure and effective. And products are released only after though analysis as per specification.

II. Literature Review

To solve the problem of captioning the image automatically. In this, a training model is developed with captioned images, and try to predict and caption the image correlation with objects and caption the image with appropriate tags. Here researchers try to find the relationship with the image feature extracted and the keywords, in order to find the appropriate keyword for the new image. The authors done the performance measure with the alternate design on large set of data and the proposed model achieves up to a 45percent relative captioning accuracy over the other design model [1].

Tradition of of manual image annotation for retrieving images and indexing the images for collection is very expensive in terms of labor and due to the millions of images over the internet it's has become a tedious job to annotate every single image on the internet. Hence, author proposes automatic way to annotate photograph for retrieval of images based on a set of pre trained image data set. Author assumes that regions in a multimedia can be segmented into frames and detect the blobs of vocabulary. Blobs are image features, generated from process called clustering. A model will be trained with a set of images with captions, the resulting model will be given a new image to process the model will show the probabilistic set of words for a given set of blob. The resulting words can be used for automatically caption a images with its object and helps retrieve the more accurate image when for a query.This model concludes that it's six times good as any other model based on the word blob co occurrence model and twice good as the machine translation model[2].

A. Object Detection and Recognition

Object detection and recognition are the first step in captioning the image, basically we find to find what all objects are present in the image and then we proceed with recognising the objects that are detected in the image. Here i list some basic MatLab functionality that will be used in the process of image annotation.

1) Object Detection in a Cluttered Scene Using Point Feature Matching: This algorithm works on detecting image based on finding point correspondences between two objects that are reference image and target image. It detect objects image even though the plane change or scale change. It Robust to little amount of out of plane rotation and conclusion. This method of object detection works better on the images which exhibits non-repeating texture patterns,which give rise to unique feature matches.

2) Object Detection Using Deep Learning: This algorithm tells about how train object detector for detecting objects using R-CNN. R-CNN, basically it's a object detection tool box which uses technique called convolutional neural network (CNN) to classify the image regions within an image. This algorithm instead classifying every single pixel using a sliding window, the R-CNN method concentrates on those regions which has more percentage to contain an object. This will help in reducing the computation work and also time and space complexity.

B. Convolutional neural network (CNN)

.

Its Basically, a type of feed forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. Individual cortical neurons respond to stimuli in a restricted region of space known as the receptive field. The receptive fields of different neurons partially overlap such that they tile the visual field. The response of an individual neuron to stimuli within its receptive field can be approximated mathematically by a convolution operation. Convolutional networks were inspired by biological processes and are variations of multilayer perceptrons designed to use minimal amounts of preprocessing. Convolutional neural network applications in image and video recognition, natural language processing and also in recommender system(shopping cart).

1) Image Retrieval Based on Convolutional Neural Network: Retrieving similar image using image retrieval, and the effect of retrieving the image is depended on the image selection and feature extraction to some extent. But the author says based on the deep learning algorithm which has self learning ability of convolutional neural network to extract more helpful to next level semantic feature retrieval using CNN, and also to use as a distance metric function similar image. By surveying this paper we realize that convolution neural network model to extract the high level semantics of the image, and also image retrieval by analyzing the structure of the network Deep convolutional neural network firstly, the image is gradually learning and abstract, each layer can be generated to describe the image content of the underlying feature of the image[19]. Convolution neural network also advances in 3 Dimension Sensing technology making it easily to record the color and depth of the image which indeed helps in the object recognition in 3D modality.A model based on a combination of convolutional and recursive neural networks for learning features and classifying RGB-D images is developed. The CNN layer learns low-level translationally invariant features which are then given as inputs to multiple, fixed-tree RNNs in order to compose higher order features[15].

eep hierarchical architectures accomplish the best published results on benchmarks for object recognition and with error rates of 2.53%, 19.51%, 0.35%, respectively. Deep nets trained by simple back propagation perform much better than more shallow ones. Learning is surprisingly rapid[17].

2) ImageNet DataSet Classification with Deep Convolutional Neural Networks: Deep convolutional neural network to classify the 1.2 million high quality images in the ImageNet LSVRC-2010 contest into the 1000 different classes/tags. On the test data, we obtained top-1 and top-5 error rate of 37.4 and 17.2 which is much better than the previous state of the art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently advanced regularization method called dropout that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

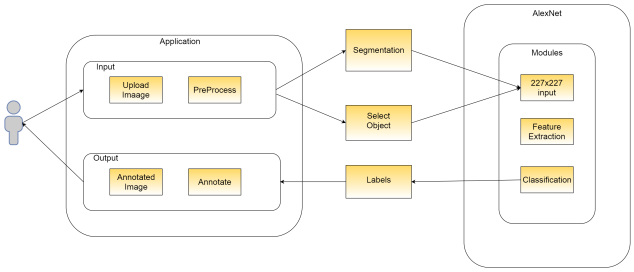
We try investigate the effect of the convolutional network depth on its exactness in the large scale image recognition setting. Thorough evaluation of networks of increasing depth using an architecture with very small (3 3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 1619 weight layers[12].

We evaluate whether features extracted from the activation of a deep convolutional network trained in a fully supervised fashion on a large, fixed set of object recognition tasks can be re- purposed to novel generic tasks. Our generic tasks may differ remarkably from the originally trained tasks and there may be inadequate labeled or unlabeled data to conventionally train or adapt a deep architecture to the new job. We explore and visualize the semantic clustering of deep convolutional features with respect to a variety of such tasks, including scene recognition, domain adaptation, and fine-grained recognition challenges.we analyze the use of deep features applied in a semi-supervised multi-task framework[14].

How Convolutional Neural Networks, trained to know objects primarily in photos, perform when applied to more abstract representations of the same objects. Main goal is to better understand the generalization abilities of these networks and their learned inner representations. So both GoogLeNet and AlexNet networks are largely unable to recognize abstract sketches that are easily recognizable by humans. Here show that the measured efficacy vary considerably across different classes.The work presented here contributes to the understanding of the applicability of CNN in domains that are different but related to that of the training set [18].

III. Implementation

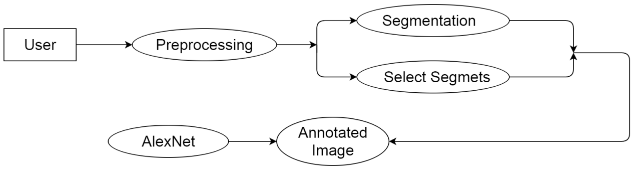
Image annotation framework using neural network uses AlexNet. In this framework user will be uploading the image through MatLab GUI, user will be given two option in the GUI whether user wants to detect the objects either manually dragging a rectangle where the user wants to detect the object or will be having a feature to detect the objects automatically where it takes some time to processes the image in the background and it will automatically detect the objects in the image. And also we can detect the objects in the image using live webcam. While designing the system we wanted to develop a framework which is simple and efficiently fast to detect the objects and annotate them.



A. Data Flow Diagram

The Data Flow Diagram is clear graphical formalism that can be used, to address a system, to the extent the data to the structure; diverse get ready did on this data and the yield data made by the structure. A Dataflow Diagram model uses an incredibly foreordained number of primitive pictures to address the limits performed by a system and the data stream among the limits.

The Framework gives user a opportunity to upload a image of his choice and the framework will starts pre processing the image, pre processing involves image to conversion to black and white and binary image etc and then image will be segmented and applied to the alexnet neural network model for object recognition and categorization and the annotated image will be presented to the user with the tag.



B. Preprocessing

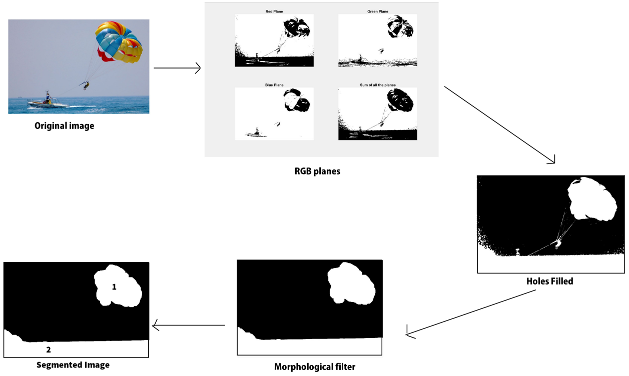
Goals:

- It enhances the image features and appearance
- Improves the image facts and suppress the unwanted features like noise.

Preprocessing step can increase the Reliability and it has several methods in it. The different sorts of pre-processing steps are:

1) Image resize

2) Image grayscale conversion
3) Binary conversion
4) Edge detection
5) Noise removal operation
6) Filters
7) Pixel brightness correction



C. Feature Extraction

The extraction of image content description is very important. This step as name indicates, it will extract the features which we want from image, which will help in recognize the objects. It will reduce the amount of resources that is needed for describing a large dataset. There are many algorithms for extracting features like SIFT, SURF, HOG, ORB etc.

Morphology is a broad set of image processing operations that process images based on shapes. Morphological operations apply a structuring element to an input image, creating an output image of the same size. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors. By choosing the size and shape of the neighborhood, you can construct a morphological operation that is sensitive to specific shapes in the input image. The most basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries.

RegionProps will find and describe features, which will be stored in form of vectors. RegionProp function is a key point detector and store the result in binary format. So basically, keypoints from the object are extracted first and stored in database or in text file. Regionprop function is used to calculate the centriod.

Regionprops measures a variety of image quantities and features in a black and white image. One of these particular properties is the centroid. This is also the centre of mass. You can think of this as the ”middle” of the object. This would be the (x,y) locations of where the middle of each object is located. As such, the Centroid for regionprops works such that for each object that is seen in your image, this would calculate the centre of mass for the object and the output of regionprops would return a structure where each element of this structure would tell you what the centroid is for each of the objects in your black and white image. Centroid is just one of the properties.

D. Segmentation

In computer vision, image segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze.

It is a process of dividing an image in to multiple blocks; it is basically used for locating an object and boundaries in an image. And it assigns labels to pixels in images. Each pixel in divided blocks will have similar attributes.

Techniques used in Segmentation:

- Segmenting Foreground and background

Application of Segmentation:

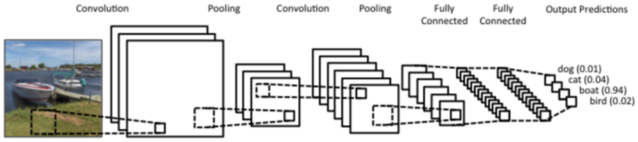
- Object Detection
- Recognition systems
- Video surveillance, etc.

Thresholding is the simplest method of image segmentation. From a grayscale image, thresholding can be used to create binary images. The simplest thresholding methods replace each pixel in an image with a black pixel if the image intensity is less than some fixed constant T or a white pixel if the image intensity is greater than that constant.

The color thresholding technique is being carried out based on the adaptation and slight modification of the grey level thresholding algorithm. Multilevel thresholding has been conducted to the RGB color information of the object extract it from the background and other objects. Different natural images have been used in the study of color information. The results showed that by using the selected threshold values, the image segmentation technique has been able to separate the object from the background.

E. Object Detection

AlexNet is one of the deep ConvNets designed to deal with complex scene classification task on Imagenet data. The task is to classify the given input into one of the 1000 classes. AlexNet has 5 convolutional layers, 3 sub sampling layers, 3 fully connected layers and its total trainable parameters are in crores. The non linearity used in the feature extractor module of the AlexNet is ReLU. AlexNet uses dropout. A fully trained AlexNet on ImageNet data set can not only be used to classify Imagenet data set but it can also be used without the output layer to extract features from samples of any other data set.



F. Annotation

The output of the last layer of the AlexNet will give the mapped label in the data set which will be used by our framework to add tags to the given image and form an annotated image.

IV. Conclusion

Here a novel framework is proposed for annotating the images using ImageNet as Dataset. For every image in the dataset, Annotations are produced and are displayed. When we input an image, it will be segmented to detect objects by Image Retrieval Algorithm Based on Convolutional Neural Network. And the input image is compared to correctness of the objects detected.The proposed method gives us a robust methodology for extracting different object of images at low time complexity.We have tested our result with other methods such "Image Classification with Bag of Visual Words" which uses K means clustering technique to classify the image which takes lot of time and space. Our proposed method takes comparatively less time than the other method to detect an object in the image.This mechanism can be used for various purposes in the field of Object detection.

V. Future Work

Future work to improve the accuracy of the system can take many directions. First, the incorporation of 3-D information in the learning process may improve the models, perhaps through learning via stereo images or 3-D images. Additionally, shape information can be utilized to improve the modeling process. Second, better and larger amounts of training images per semantic concept may produce more robust models. Contextual information may also help in the modeling and annotation process. Third, this method holds promise for various application domains, including bio-medicine. Finally, the system can be integrated with other retrieval methods to improve usability in the field of computer vision.

Acknowledgment

This project would not have been possible without the support of many people, each of whom, in their own capacities has helped us in carrying out our work. We would like to take this opportunity to thank them all. First and foremost we extend our hand to thank our
HOD, Dr. Vijay Kumar B.P, Professor and Head, Department of Information Science Engineering, for permitting us to take up this project. We also wish to express our gratitude to our project guide, Mr. Prashanth kambli who was abundantly helpful and offered invaluable assistance, exemplary guidance, monitoring and constant encouragement.

We would also like to thank all teaching and non-teaching staff of our department for their kind cooperation. Although there are many who remain unacknowledged in this humble note of gratitude, there are none who remain unappreciated.

References

1) Jia-Yu Pan, Hyung-Jeong Yang, Pinar Duygulu and Christos Faloutsos,Automatic Image Captioning,2104.
2) Jeon J, Lavrenko V, Manmatha R, Automatic image annotation and retrieval using cross-media relevance models, Jul.2003.
3) Ilaria Bartolini and Paolo Ciaccia,Imagination: Exploiting Link Analysis for Accurate Image Annotation.2012
4) Chong Wang, David Blei, Li Fei-Fei,Simultaneous Image Classification and Annotation,2012.
5) Yashaswi Verma and C. V. Jawahar,Image Annotation Using Metric Learning in Semantic Neighbourhoods,2012
6) Yasuhide MORI, Hironobu and Ryuichi OKA, Image-to-word transformation based on dividing and vector quantizing images with words,1999.
7) Duygulu, P., Barnard, K., Freitas, J., Forsyth, D.A., Object recognition as machine translation: learning a lexicon for a fixed image vocabulary".
8) Guillaumin, M,Ferrari, V. Large scale knowledge transfer for object localization in ImageNet,2012
9) Jeon J, Lavrenko V, Manmatha R, Automatic image annotation and retrieval using cross-media relevance models, Jul.2003.
10) Shaoting Zhang, Junzhou Huang, Hongsheng Li, and Dimitris N. Metaxas , Automatic Image Annotation and Retrieval Using Group Sparsity, IEEE June 2012 Sparsity, IEEE June 2012
11) Rong Jin, Joyce Y. Chai, Luo Si, Effective Automatic Image Annotation Via A Coherent Language Model and Active Learning, ACM, 2004.
12) Karen Simonyan, Andrew Zisserman,Very Deep Convolutional Networks for Large-Scale Image Recognition,2014.
13) Krizhevsky, A., Sutskever, I. and Hinton, G. E,ImageNet Classification with Deep Convolutional Neural Networks.NIPS 2012.
14) Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, Trevor Darrell,DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition,2013.
15) Richard Socher, Brody Huval, Bharath Bhat, Christopher D. Manning, Andrew Y. Ng, Convolutional-Recursive Deep Learning for 3D Object Classification.April 2014.
16) Ronan Collobert, Ronan Collobert , A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning.
17) Collobert, Ronan; Weston, Jason, A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning.
18) Pedro Ballester and Ricardo Matsumura Araujo, On the Performance of GoogLeNet and AlexNet Applied to Sketches.
19) Hailong Liu, Baoan Li, Xue Qiang Lv and Yue Huang,Image Retrieval Algorithm Based on Convolutional Neural Network, august 2016.
20) Ando, R. and Zhang, T, A framework for learning predictive structures from multiple tasks and unlabeled data. 2005.