ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

A Survey on Object Detection Methodologies

Meenakshi Chandak[1], Dr. A. S. Ghotkar[2]

Department of Computer Engineering, PICT, Pune[1,2]

Abstract: Visual content is becoming a major medium for social interaction on the internet, including various popular platforms like Flickr, Instagram and Facebook. Increased attention has been obtained by the visual sentiment analysis due to the rapidly grown number of images in online interactions. Several applications such as advertisement, education and entertainment emerged because of the development in this field. Most work reported in the literature focuses on competent techniques for object recognition and its applications. This paper includes various approaches that have been used by different researchers for object detection.

Keywords: Visual Content, Computer Vision, Social Multimedia, Object Detection.

I.INTRODUCTION

The modern world is enclosed with gigantic masses of digital visual information. To analyze and organize this devastating ocean of visual information image analysis techniques are major requisite. In particular it would be useful, methods that could automatically analyze the semantic contents of images or videos. The content of the image determines the significance in most of the potential uses. One important aspect of image content is the objects in the image. So there is a need for object recognition techniques. Object detection describes the task of processing an image in a certain way to localize and to classify objects. There exists a huge variety of object recognition approaches, but the general concept remains the same. An object recognition system uses training datasets containing images with known and labelled objects and it extracts different types of information based on the chosen algorithm. This can be information about colours, edges, geometric forms and so on. Generally, for any new image the same information is gathered and compared to the training dataset to find the most suitable classification. Many applications using object recognition can be found in everyday life. Starting with robots in industrial environments, face or handwriting recognition and up to autonomous systems such as modern cars which use object recognition for pedestrian detection, emergency brake assistant and so on.

Problems and Challenges with object Detection:
Variable number of objects: When training machine learning models, usually representation of data need to be into fixed-sized vectors. Since the number of objects in the image is not known beforehand, correct number of outputs is unknown. Because of this, some post-processing is required, which adds complexity to the model.

Sizing: Another big challenge is the different conceivable sizes of objects. When doing simple classification, expectation is to classify objects that cover most of the image. On the other hand, some of the objects may want to find could be a small as a dozen pixels (or a small percentage of the original image).

Modeling:A third challenge is solving two problems at the same time. How do we combine the two different types of requirements: location and classification into, ideally, a single model.

II.APPROACHES FOR OBJECT DETECTION

A. Classical Approaches

Although there have been many different types of methods throughout the years, lets focus on the two most popular ones (which are still widely used). The first one is the Viola-Jones framework proposed in 2001 by Paul Viola and Michael Jones in the paper Robust Real-time Object Detection. The approach is fast and relatively simple, so much that it's the algorithm implemented in point-and-shoot cameras which allows real-time face detection with little processing power.

We won't go into details on how it works and how to train it, but at the high level, it works by generating different (possibly thousands) simple binary classifiers using Haar features. These classifiers are assessed with a multi-scale sliding window in cascade and dropped early in case of a negative classification. Another traditional and similar method is using Histogram of Oriented Gradients (HOG) features and Support Vector Machine (SVM) for classification. It still requires a multi-scale sliding window, and even though it's superior to Viola-Jones, it's much slower.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

B. Deep Learning Approaches

It's not news that deep learning has been a real game changer in machine learning, especially in computer vision. In a similar way that deep learning models have crushed other classical models on the task of image classification, deep learning models are now state of the art in object detection as well. Lets do an overview on how the deep learning approach has evolved in the last couple of years.

OverFeat: One of the first advances in using deep learning for object detection was OverFeat from NYU published in 2013. They proposed a multi-scale sliding window algorithm using Convolutional Neural Networks (CNNs). [5]

R-CNN: Quickly after OverFeat, Regions with CNN features or R-CNN from Ross Girshick, et al. at the UC Berkeley was published which boasted an almost 50% improvement on the object detection challenge. What they proposed was a three stage approach:

- Extract possible objects using a region proposal method (the most popular one being Selective Search).
- Extract features from each region using a CNN.
- Classify each region with SVMs.

While it achieved great results, the training had lots of problems. To train it you first had to generate proposals for the training dataset, apply the CNN feature extraction to every single one (which usually takes over 200GB for the Pascal 2012train dataset) and then finally train the SVM classifiers. [6]

Fast R-CNN: This approach quickly evolved into a purer deep learning one, when a year later Ross Girshick (now at Microsoft Research) published Fast R-CNN. Similar to R-CNN, it used Selective Search to generate object proposals, but instead of extracting all of them independently and using SVM classifiers, it applied the CNN on the complete image and then used both Region of Interest (RoI) Pooling on the feature map with a final feed forward network for classification and regression. Not only was this approach faster but having the RoI Pooling layer and the fully connected layers allowed the model to be end-to-end differentiable and easier to train. The biggest downside was that the model still relied on Selective Search (or any other region proposal algorithm), which became the bottleneck when using it for inference. [7]

YOLO: Shortly after that, You Only Look Once: Unified, Real-Time Object Detection(YOLO) paper published by Joseph Redmon (with Girshick appearing as one of the co-authors). YOLO proposed a simple convolutional neural network approach which has both great results and high speed, allowing for the first time real time object detection.[8]

Faster R-CNN: Subsequently, Faster R-CNN authored by Shaoqing Ren (also co-authored by Girshick, now at Facebook Research), the third iteration of the R-CNN series. Faster R-CNN added what they called a Region Proposal Network (RPN), in an attempt to get rid of the Selective Search algorithm and make the model completely trainable end-to-end. We won't go into details on what the RPNs does, but in abstract it has the task to output objects based on an "objectness" score. These objects are used by the RoI Pooling and fully connected layers for classification.[9]

SSD and R-FCN: Finally, there are two notable papers, Single Shot Detector (SSD) which takes on YOLO by using multiple sized convolutional feature maps achieving better results and speed[10], and Region-based Fully Convolutional Networks (R-FCN) which takes the architecture of Faster R-CNN but with only convolutional networks.[11]

III. SURVEY OF PREVIOUS PAPERS

Hsu, Liao et al presented a general framework for predicting likely affective responses of the viewers in the social media environment after an image is posted online. The approach emphasizes a mid-level concept representation, in which intended effects of the image publisher is characterized by a large pool of visual concepts termed as Publisher affect concept detected from image content directly instead of textual metadata, evoked viewer affects are represented by concepts mined from online comments termed as Viewer affective concept, and probabilistic methods are used to model the co-relations among these two types of concepts.[1]

Chen, Chang et al proposed object-based visual concepts such as happy dog and yummy food with a goal to extract emotion related information from social multimedia content. Detection of adjective noun pairs is the main focus because of their strong co-occurrence relation with image tags about emotions. Highly subjective nature of the adjectives like happy and yummy and the ambiguity with the annotations makes the problem very challenging. Associated adjectives with physical nouns have made the combined visual concepts more tractable and detectable. A

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

hierarchical system to handle the concept classification in an object specific manner and decomposing the hard problem into sentiment related concept model and object localization is proposed in the paper.[2]

Taikun Liu, Hong-Yuan Mark Liao et al presented a general framework and working system for predicting likely affective responses of the viewers in the social media environment after an image is posted online. Their approach emphasizes a mid-level concept representation, in which intended affects of the image publisher is characterized by a large pool of visual concepts (termed PACs) detected fromimage content directly instead of textual metadata, evoked viewer affects are represented by concepts (termed VACs) mined from online comments, and statistical methods are used to model the correlations among these two types of concepts. Demonstration of the utilities of such approaches are shown by developing an end-to-end Assistive Comment Robot application, which further includes components for multi-sentence comment generation, interactive interfaces, and relevance feedback functions.[3]

Machajdik and Hanbury addressed the challenge of sentiment analysis from visual content. In contrast to the methods that have been existed which infer sentiment or emotion directly from low level features, they have proposed a novel approach based on the semantics of images. The key contribution is two-fold, they present a psychology theory based method to automatically construct a large scale Visual Sentiment Ontology (VSO) consisting of more than 3000 Adjective Noun Pairs (ANP). Second, they propose SentiBank, a novel mid level representation framework built upon the VSO encoding the concept presence of 1200 ANPs from visual content. The Visual Sentiment Ontology and SentiBank are distinct from existing works and can be used in various high level applications.[4]

AfsaneRajaei and Hamidrezashayegh explained in their paper that the shape of an object can be well represented by a distribution of local intensity gradients or edge directions. This is done by dividing the picture into small spatial parts and finding the edge histograms orientations on all pixels of the cell. The combined histogram arrives form the featurerepresentation after local contrast normalization in overlapping descriptor blocks. For classification, a set of data is created from human and non-human examples, and a linear one classifies is made using SVM on the features of the gradient histogram from both classes. This classifier can then switch to a new one multi-scale input image for recognizing people.[14]

IV. DATASETS

Name	Brief Description	Images	Classes	Created (Updated)
ImageNet	Labelled object image database, used in the ImageNet Large Scale Visual Recognition Challenge	450k	200	2009 (2014)
Common Objects in Context (COCO)	Complex everyday scenes of common objects in their natural context.	120k	80	2014 (2015)
Pascal VOC	Large number of images for classification tasks.	12k	20	2010 (2012)
KITTI Vision	Autonomous vehicles driving through a mid-size city captured images of various areas using cameras and laser scanners.	7k	3	2012 (2014)

IV. CONCLUSION

In conclusion, there are many opportunities regarding object detection, both in unseen applications and in new methods for pushing state of the art results. This paper presents an extensive survey of object detection approaches and also gives a brief review of each approach. There are two approaches such as Classical approaches and Deep learning approaches. Deep leaning model is very efficient for object detection where we discussed on some methods like R-CNN, YOLO, faster R-CNN, SSD and R-FCN of Convolutional Neural Network.This survey on approaches of object detection can give valuable insight into this important research topic and encourage new research.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

REFERENCES

[1] Yan-Ying Chen,Tao et al, "Predicting Viewer Affective Comments Based on Image Content in Social Media", ACM 2014.
[2] Tao Chen, Felix X. Yuet al, "Object-Based Visual Sentiment Concept Analysis and Application", ACM 2014.
[3] Yan-Ying Chen, Tao Chen et al, "Assistive Image Comment Robot a novel mid-level concept based Representation", IEEE 2015.
[4] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory", ACM Multimedia, 2010.
[5] Pierre Sermanet, David Eigenet al, "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks", ILSVRC 2013.
[6] Girshick, Ross, et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", ACM 2014.
[7] Ross and Girshick,"Fast R-CNN", IEEE 2016.
[8] Redmon, Joseph, et al, "You only look once: Unified, real-time object detection", IEEE 2016.
[9] Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks", 2015.
[10] Wei Liu, Dragomir Anguelov, et al," SSD: Single Shot MultiBox Detector", 2015
[11] Jifeng Dai, Yi Li, et al, " R-FCN: Object Detection via Region-based Fully Convolutional Networks", 2016.
[12]https://tryolabs.com/blog/2017/08/30/object-detection-an-overview-in-the-age-of-deep-learning/
[13] D. Borth, R. Ji, T. Chen, T. Breuel, and S.F. Chang,"Large-scale visual sentiment ontology and detectors using adjective noun pairs", ACM Multimedia, 2013.
[14] AfsaneRajaei and Hamidrezashayegh, "Human Detection in Semidense Scenes Using HOG descriptor and Mixture of SVMs", in 3rd International Conference on Computer and Knowledge Engineering (ICCKE 2013)
[15] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining",LREC, 2006. [16] Bird, Steven, E. Loper, and E. Klein," Natural language processing with python", 2009.
[17] Hillip Isola, Jian xiong Xiao, Antonio Torralba, Aude Oliva,"What makes an image memorable?", CVPR,2011.
[18] S. Siersdorfer, S. Chelaru, W. Nejdl, and J. San Pedro, "How useful are your comments? Analyzing and predicting youtube comments and comment ratings",2010.
[19] R. Datta, D. Joshi, J. Li, and J. Z. Wang,"Studying aesthetics in photographic images using a computational approach", ECCV 2006.
[20] P. Lang, M. Bradley, and B. Cuthbert, "Internationalective picture system (IAPS): Affective ratings of pictures and instruction manual", ACM 2008.
[21] Target Advertising Wikipedia Free Encyclopedia.