Project Proposal: Pattern Recognition (EEL 6825) Object Detection Within Digital Images

Vikas Chaubey
UFID: 3511 5826
Email: vikas.chaubey@ufl.edu
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, Fl 32608

Abstract—The project aims at implementing a convolution neural network which can perform task of object detection within a given digital image. Object detection techniques are used to identify objects in a given image, classify them (identify them) and locate them with a bounding box within image. The project focuses on implementing a neural network using weakly supervised object detection techniques.

Keywords-component; pattern recognition, project proposal, object detection, weakly supervised learning

I. INTRODUCTION

Object detection is a task in computer vision that involves identifying the presence, location, and type of one or more objects in a given image.It is widely used in computer vision tasks such as image annotation, activity recognition, face detection, face recognition, video object co-segmentation etc. When humans look at images or video, we can recognize and locate objects of interest within a matter of moments. The goal of object detection is to replicate this intelligence using a computer program.

An Object detection computer program has to perform following tasks: object recognition (recognizing the presence of objects in a given image/video), Object classification (classifying the category of different objects) and Localization (determining the location of the objects in the image by putting a bounding box around them)

Methods for object detection generally fall into either machine learning-based approaches or deep learning-based approaches. For example, some popular Machine Learning approaches are Viola Jones object detection framework based on Haar features, Scale-invariant feature transform (SIFT), Histogram of oriented gradients (HOG) features. Similarly, some Deep Learning approaches are Region Proposals (R-CNN, Fast R-CNN, Faster R-CNN), Single Shot Multi Box Detector (SSD), You Only Look Once (YOLO), Retina-Net, Deformable convolutional networks. In present for object detection in images convolution neural networks are most popular because they do not require feature engineering. In most of the traditional Machine Learning algorithms the feature selection is done first then these features are fed to a selected classifier which does the classification. In algorithm based image classification approaches, we need to select the features (local, global) and classifiers. On the other hand, deep learning based convolutional neural networks (CNN) are able to do end-to-end object detection without specifically defining features, when we compare handcrafted features with CNN, CNN performs better, and it gives better accuracy. It also covers local and global features. It can also learn different features from images.

Even though Convolutional Neural Networks based approach for object detection in images are comparatively faster and accurate, but the main challenge lies in the data preparation. In order to train these neural networks, the training image data sets needs special preparation which includes annotating objects using bounding boxes and labeling them in all the images. In case of large data set this preparation is very time consuming and expensive. To overcome these challenges concept of weakly supervised learning is gaining traction in present time.

What is weakly supervised learning? Most of the deep learning approaches for object detection require that the objects should be annotated within bounding boxes in the training data sets in order to train model. This is an effective approach to train models with high accuracy. But in case of very large data sets cost of labelling bounding boxes annotations around objects is very expensive, also annotations produced by humans are prone to errors. Hence weakly supervised object detection techniques could be used to train models which do not require training data to be bound boxed. In these techniques labelled data without bounding box annotation could be used for training of models which can reduce data preparation effort and reduce development costs. I plan to implement object detection using a neural network which will be using weakly supervised learning techniques. Such neural networks could be implemented using methods like Gradient based class activation mapping for classification and localization of objects within a given image.

II. Project Plan

A. Project Subject Search

After reading many blogs on web and posts on websites like Quora, looking at the projects on websites such as Kaggle, and listening to famous machine learning enthusiasts on YouTube I decided to do my project in the field of Object detection. The main reason and motivation behind choosing object detection for my project is that computer vision as a field has always fascinated me. The wide range of use cases in different fields makes this domain really interesting for example object detection is currently being used to build self-driving cars (built by Tesla), Delivery drones (being developed by Amazon), face recognition and Pedestrian tracking etc. I think this project will give me good exposure to learn and understand the advancements in the field of object detection. As per project plan I have kept the deadline for topic research task as 10th March 2020.

B. Initial Literature Research

With this project I plan to explore the field of object detection within images in general that includes exploring all popular advancements in this field from traditional object detection approaches to new machine learning based methods and finally deep learning based effective object detection solutions. I plan to do this research for topics like Viola Jones object detection framework based on Haar features, Scale-invariant feature transform (SIFT), Histogram of oriented gradients (HOG) features and Deep Learning approaches like Region Proposals (R-CNN, Fast R-CNN, Faster R-CNN), Single Shot Multi Box Detector (SSD), You Only Look Once (YOLO), Retina-Net, Deformable convolutional networks. In this project I plan to Implement the convolutional neural network using weakly supervised learning techniques which could be achieved using Gradient based class activation mapping, hence a deep understanding of GRAD-CAM and GRAD-CAM ++ algorithms is required , I have planned to read multiple research papers as well as web based articles to gain insight in those methodologies. As per project plan I have kept the deadline for literature research topic exploration task as 20th March 2020.

C. Data Set Preparation for Neural Network Training

The CNN model will be trained and tested using CIFAR-10 data set. It consists of 60,000 images of 10 classes (each class is represented as a row in the above image). In total, there are 50,000 training images and 10,000 test images. The dataset is divided into 6 parts – 5 training batches and 1 test batch. Each batch has 10,000 images. Project plan includes doing a sanity check on the downloaded data set in order to remove irrelevant or very bad quality images using a python script from the data set. This task is planned to be finished by 22nd March 2020.

D. Coding, Implementation and Testing

The project will be developed using Python 3.7. Libraries like TensorFlow, Keras and NumPy will be used to implement a convolutional neural network with GRAD-CAM. after development, model will be trained using CIFAR-10 Training data set. After completion of training, model accuracy will be tested on 10000 test images which data set provides. This task is planned to be finished by 5th April 2020.

E. Project Report and Presentation

After the completion of the project implementation, I will be working on the project report, a power point presentation explaining the project idea, its development process and results obtained, I will also be preparing a video which will showcase the functionality of the neural network. This task is planned to be finished by 15th April 2020.

III. References

[1] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning Deep Features for Discriminative Localization," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2921-2929.

[2] A. Bergamo, L. Bazzani, D. Anguelov, and L. Torresani. Self-taught object localization with deep networks. arXiv preprint arXiv:1409.3964, 2014. 1, 2

[3] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229, 2013.

[4] A. Chattopadhay, A. Sarkar, P. Howlader and V. N. Balasubramanian,"Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, 2018, pp. 839-847.

[5] Francois Chollet. 2017. Deep Learning with Python (1st. ed.). Manning Publications Co., USA.

[6] R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra and J. M. Z. Maningo, "Object Detection Using Convolutional Neural Networks," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 2023-2027.

[7] J. Huang, A. Fathi, V. Rathod, I. Fischer, C. Sun, Z. Wojna, M. Zhu, Y. Song, A. Korattikara, S. Guadarrama, K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors", pp. 1-21, April 2017