OBJECT DETECTION

A Report for the Evaluation 3 of project 2

Submitted by

Anurag Singh

(1613101167/16SCSE101540)

In partial fulfillment for the award of the degree

of

Bachelor of Technology

IN

Computer Science and Technology

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

Under the Supervision of

Dr.D. Ganesh Gopal(CSE

Professor

Galgotias University

April-may 2020

TABLE OF CONTENTS

1. Abstract

Efficient and accurate object detection has been an important topic in the advancement of computer vision systems. The project aims to incorporate state-of-the-art technique for object detection with the goal of achieving accuracy with a real-time performance. A major challenge in many of the object detection systems is the dependency on other computer vision techniques for helping the deep learning based approach, which leads to slow and non-optimal performance. Computer vision is consists of various aspects like image recognition, image generation, object detection. In this project, we are using highly accurate object detection-algorithms and methods such as R-CNNN, Fast-RCNN, Faster-RCNN, RetinaNet and fast yet highly accurate ones like SSD and YOLO. Using these methods and algorithms, based on deep learning which is also based on machine learning require lots of mathematical and deep learning frameworks understanding by using dependencies such as TensorFlow, OpenCV, imageai etc, we can detect each and every object in image by the area object in an highlighted rectangular boxes and identify each and every object and assign its tag to the object. This also includes the accuracy of each method for identifying objects. In this project, we use a completely deep learning based approach to solve the problem of object detection in an end-to-end fashion.

1.Introduction

A few years ago, the creation of the software and hardware image processing systems was mainly limited to the development of the user interface, which most of the programmers of each firm were engaged in. The situation has been significantly changed with the advent of the Windows operating system when the majority of the developers switched to solving the problems of image processing itself. However, this has not yet led to the cardinal progress in solving typical tasks of recognizing faces, car numbers, road signs, analyzing remote and medical images, etc. Each of these "eternal" problems is solved by trial and error by the efforts of numerous groups of the engineers and scientists. As modern technical solutions are turn out to be excessively expensive, the task of automating the creation of the software tools for solving intellectual problems is formulated and intensively solved abroad. In this field of image processing, the required tool kit should be supporting the analysis and recognition of images of previously unknown content and ensure the effective development of applications by ordinary programmers. Just as the Windows toolkit supports the creation of interfaces for solving various applied problems.

Object recognition is to describe a collection of related computer vision tasks that involve activities like identifying objects in digital photographs. Image classification involve activities such as predicting the class of one object in an image. Object localization is refers to identifying the location of one or more objects in an image and drawing an abounding box around their extent. Object detection does the work of combines these two tasks and localizes and classifies one or more objects in an image. When a user or practitioner refers to the term "object recognition", they often mean "object detection".It may be challenging for beginners to distinguish between different related computer vision tasks. So, we can distinguish between these three computer vision tasks with this example:

Image Classification: This is done by Predict the type or class of an object in an image.

Object Localization: This is done through, Locate the presence of objects in an image and indicate their location with a bounding box.

Object Detection: This is done through, Locate the presence of objects with a bounding box and types or classes of the located objects in an image.

One of the further extension to this breakdown of computer vision tasks is object segmentation, also called “object instance segmentation” or “semantic segmentation,” where instances of recognized objects are indicated by highlighting the specific pixels of the object instead of a coarse bounding box. From this breakdown, we can understand that object recognition refers to a suite of challenging computer vision tasks. For example, image classification is simply straight forward, but the differences between object localization and object detection can be confusing, especially when all three tasks may be just as equally referred to as object recognition.

Humans can detect and identify objects present in an image. The human visual system is fast and accurate and can also perform complex tasks like identifying multiple objects and detect obstacles with little conscious thought. The availability of large sets of data, faster GPUs, and better algorithms, we can now easily train computers to detect and classify multiple objects within an image with high accuracy. We need to understand terms such as object detection, object localization, loss function for object detection and localization, and finally explore an object detection algorithm known as “You only look once” (YOLO).

Image classification also involves assigning a class label to an image, whereas object localization involves drawing a bounding box around one or more objects in an image. Object detection is always more challenging and combines these two tasks and draws a bounding box around each object of interest in the image and assigns them a class label. Together, all these problems are referred to as object recognition.

Object recognition refers to a collection of related tasks for identifying objects in digital photographs. Region-based Convolutional Neural Networks, or R-CNNs, is a family of techniques for addressing object localization and recognition tasks, designed for model performance. You Only Look Once, or YOLO is known as the second family of techniques for object recognition designed for speed and real-time use.

Object detection is an important task, yet challenging vision task. It is a critical part of many applications such as image search, image auto annotation and scene understanding, object tracking. Moving object tracking of video image sequences was one of the most important subjects in computer vision. It had already been applied in many computer vision fields, such as smart video surveillance (Arun Hampapur 2005), artificial intelligence, military guidance, safety detection and robot navigation, medical and biological application.

2.Proposed System

SqueezeNet:

SqueezeNet is name of a DNN for computer vision. This original version of SqueezeNet was implemented on top of the Caffe deep learning software framework SqueezeNet ships as part of the source code of a number of deep learning frameworks such as PyTorch, Apache MXNet, and Apple CoreML.

InceptionV3:

Inception v3 is widely used as image recognition model that has showed to obtain accuracy of greater than 78.1% on the ImageNet dataset.

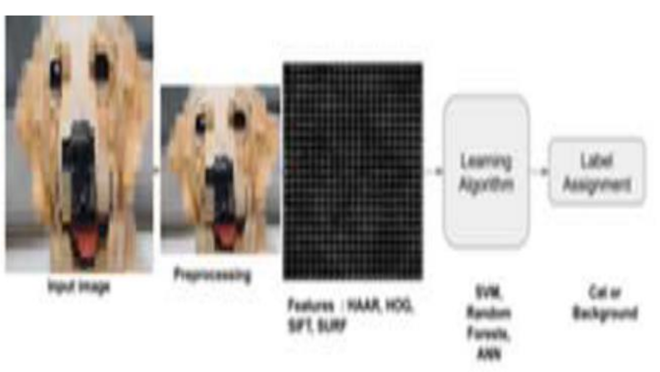
DenseNet:

DenseNet stands for Densely Connected Convolutional Networks it is one of the latest neural networks for visual object recognition. It is similar to ResNet but has some fundamental differences.

3.Implementation

PREPROCESSING

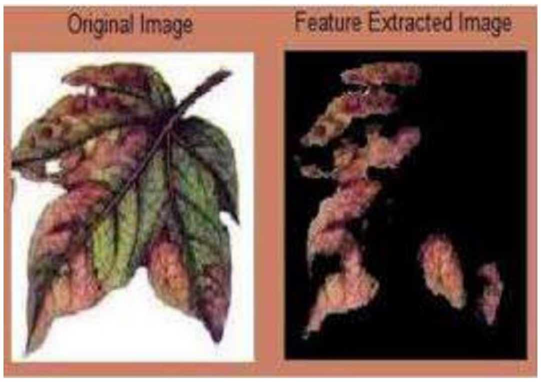
It is the lowest level of abstraction. The process preprocessing improves the image intensity by suppressing the unwanted features or enhancing them for further processing.[2].It resizes the image size to 448*448 and also normalizes the contrast and brightness effects. The image is also cropped and resized so that feature extraction can be performed easily. The input images are pre-processed and very easily normalize the contrasts and brightness. The Preprocessing step can be done by subtracting the mean of image intensities and divide by the standard deviation. New brightness value can be found by using the neighborhood of a pixel in the input image. The fig 2 below shows the preprocessing of the image.



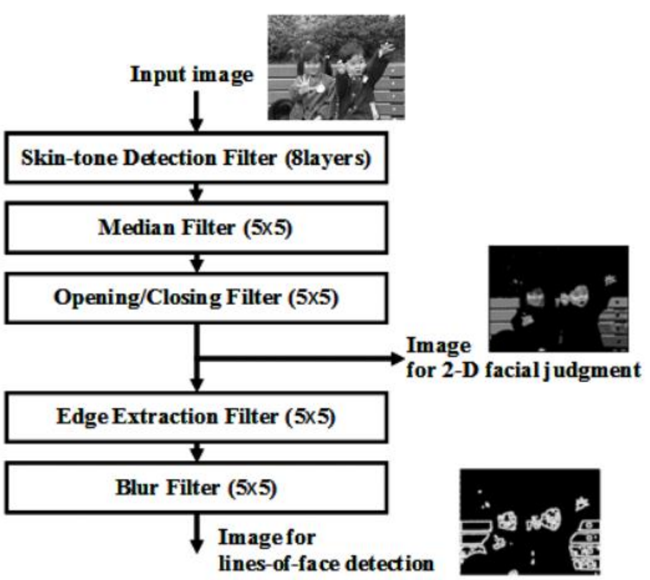
Preprocessing

FEATURE EXTRACTION

Its main motive is to simplify the image by considering only the important information and leaving out the extra information which is not necessary for recognition. It uses the method of edge detection which can only retain the essential information .it represents the reduced part of an image as a feature vector. This approach is used when the size of the image is very large. Hence, by this process, image recognition becomes easier. It starts from the already measured data and features which provides some kind of information facilitating the further steps.



Feature Extraction.



Combined process of Preprocessing and feature extraction.

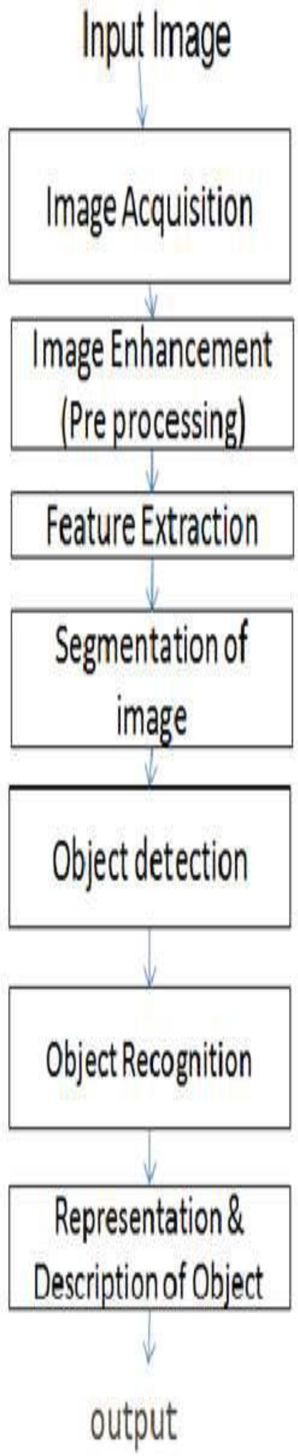
Template Matching

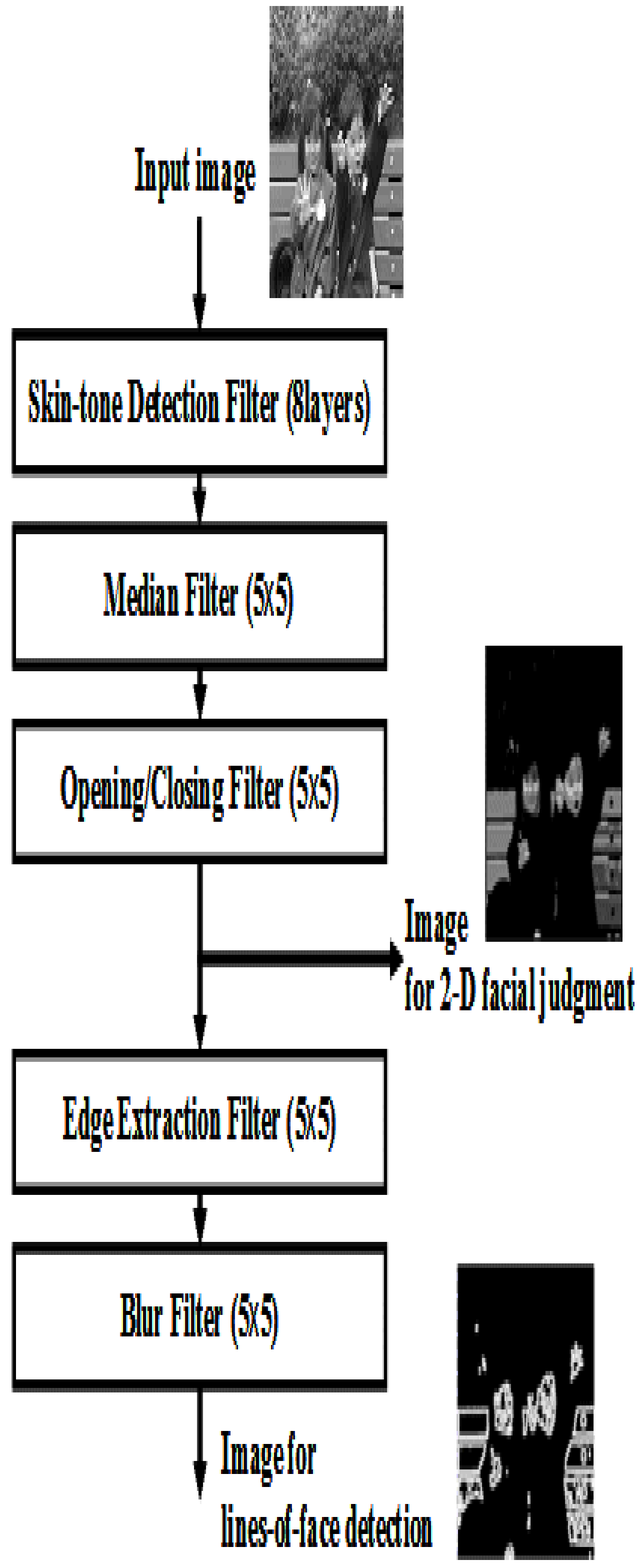
Template matching is a technique for finding small parts of an image which match a template image. It is a straight forward process. Template Matching is the technique of finding small parts of an image which match a template image. It slides the template from the top left to the bottom right of the image and compares for the best match with the template. The template dimension should be equal to the reference image or smaller than the reference image. It recognizes the segment with the highest correlation as the target. Given an image S and an image T, where the dimension of S was both larger than T, output whether S contains a subset image I where I and T are suitably similar in pattern and if such I exists, output the location of I in S as in Hager and Bellhumear (1998). Schweitzer et al (2011), derived an algorithm which used both upper and lowers bound to detect 'k' best matches. Euclidean distance and Walsh transform kernels are used to calculate match measure. The positive things included the usage of priority queue improved quality of decision as to which bound-improved and when good matches exist inherent cost was dominant and it improved performance.

Background Subtraction

The background and foreground subtraction method used the following approach. A pixel was modelled by a 4-tuple [Ei, si, ai, bi], where Ei- a vector with expected colour value, si - a vector with the standard deviation of colour value, ai the variation of the brightness distortion and bi was the variation of the chromaticity distortion of the ith pixel. In the next step, the difference between the background image and the current image was evaluated. Each pixel was finally classified into four categories: original background, shaded background or shadow, highlighted background and moving foreground object.This algorithm mainly used two steps Firstly, a sub-sampling scheme based on background subtraction techniques was implemented to obtain stationary foreground regions. This detects foreground changes at different time instants in the same pixel locations. This was done by using a Gaussian distribution function. Secondly, som modifications were introduced on this base algorithm such as thresh holding the previously computed subtraction. The main purpose of this algorithm was reducing the amount of stationary foreground detected.

4.Result


Input Image
Image Acquisition
Image Enhancement
(Pre processing)
Feature Extraction
Segmentation of
image
Object detection
Object Recognition
Representation &
Description of Object
output


Input image
Skin-tone Detection Filter (8layers)
Median Filter (5X5)
Opening/Closing Filter (5X5)
Image
for 2-D facial judgment
Edge Extraction Filter (5X5)
Blur Filter (5X5)
Image for
lines-of-face detection

6. Conclusion

The basic purpose behind this topic is that it is something that will overdo all the physical tasks. Object recognition and tracking reduces human efforts and provides efficiency. It is of interest as it may help humans to be aware of minute information about objects and reduce human tasks. Automatic recognition and extraction adds to the smart systems used today.

An accurate and efficient object detection system has been developed which achieves comparable metrics with the existing state-of-the-art system. This project uses recent techniques in the field of computer vision and deep learning. Custom dataset was created using labelImg and the evaluation was consistent. This can be used in real-time applications which require object detection for pre-processing in their pipeline.

7.References

• Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. IEEE Trans. Pattern Anal. Mach. Intell. 26,1475–1490. doi:10.1109/TPAMI.2004.108

• Alexe, B., Deselaers, T., and Ferrari, V. (2010). “What is an object?,” in ComputerVision and Pattern Recognition (CVPR), 2010 IEEE Conference on (San Francisco,CA: IEEE), 73–80. doi:10.1109/CVPR.2010.5540226

• Aloimonos, J., Weiss, I., and Bandyopadhyay, A. (1988). Active vision. Int. J.Comput. Vis. 1, 333–356. doi:10.1007/BF00133571models,” in Computer Vision-ECCV 2012 (Florence: Springer),836–849. 6. Azzopardi, G., and Petkov,