Image Segmentation and Text Recognizer Using Deep Learning:

A PROJECT SYNOPSIS

SUBMITTED BY

Ishan Modi (N18018)

Ishan Srivastava (N18068)

Janhavi Panambor (N18057)

UNDER THE GUIDANCE OF

DR. VARSHA DEGAONKAR

BE (ELECTRONICS AND TELECOMMUNICATION)

DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATION

HOPE FOUNDATION's

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY,

HINJAWADI, PUNE(MH)-411057

SAVITRIBAI PHULE PUNE UNIVERSITY

A.Y. 2021-22

Project Details

Title

Image Segmentation and Text Recognizer Using Deep Learning.

Introduction

Image segmentation is an essential component in many visual understanding systems. It involves partitioning images (or video frames) into multiple segments or objects [2]. Segmentation plays a central role in a broad range of applications [3], including medical image analysis (e.g., tumor boundary extraction and measurement of tissue volumes), autonomous vehicles (e.g., navigable surface and pedestrian detection), video surveillance, and augmented reality to count a few. Numerous image segmentation algorithms have been developed in the literature, from the earliest methods, such as thresholding [4], histogram-based bundling, regiongrowing [5], k-means clustering [6], watersheds [7], to more advanced algorithms such as active contours [8], graph cuts [9], conditional and Markov random fields [10], and sparsitybased [11] - [12] methods. Over the past few years, however, deep learning (DL) networks have yielded a new generation of image segmentation models with remarkable performance improvements often achieving the highest accuracy rates on popular benchmarks resulting in what many regard as a paradigm shift in the field. For example, Figure 1 presents sample image segmentation outputs of a prominent deep learning model, DeepLabv3 [13].

Figure 1: Segmentation results of DeepLabV3 [13] on sample images

Image segmentation can be formulated as a classification problem of pixels with semantic labels (semantic segmentation) or partitioning of individual objects (instance segmentation). Semantic segmentation performs pixel-level labeling with a set of object categories (e.g., human, car, tree, sky) for all image pixels, thus it is generally a harder undertaking than image classification, which predicts a single label for the entire image. Instance segmentation extends semantic segmentation scope further by detecting and delineating each object of interest in the image (e.g., partitioning of individual persons).

In the digital era of technology, everyday large volumes of data is generated in the form of text, images, and videos. which are raw data. Applying artificial intelligence to them we can classify their domain applications and working fields. We will be analyzing the images and video files encountered by the camera which can be scaled upto humans and scaled down to text's written on paper, the model will be able to comprehensively review and then process the image into segments with classification into different objects. Through this , we will be able to get number of insights for images and videos around us. Thus, this can be implemented at places around us for convinient usage and needs.

Objective and Scope

- To, simplify and enhance the representation of an image in more meaningful way to analyze.
- To learn more about Data form analysis and dive into Aritifical intelligence.

Abstract

Deep learning has been consistently developed day by day for its various application into the area of healthcare, security, education etc. Data has been the key of today's modern word with digital era. Image, Video, Audio and Text are the major forms of data which are generated in large volumes on a daily basis, using all these raw data we can apply Deep learning algorithms to it and get valuable insights. Here we have performed the comprehensive review on the same, model is been trained with

different datasets of objects which inlcudes humans on the bigger scale and texts on the smaller scale. This model is trained extensively using TensorFlow using Python and then later implemented through a proper GUI at Flask Web Application. Web application consists a basic image input through the user and then image is segmented using the model training. This is entirely based on Python and thus finds many robust applications in different domains. [1]

Project Category

Digital Image Processing, Deep Learning.

Process Description

The system will work on pre trained models that are able to analyze the input images. Once the system has input images, it performs pre-processing operations on it. Pre-processing is the name for operations on images at the lowest level of abstraction whose aim is an improvement of the image data that suppress undesired distortions or enhances some image features important for further processing. Once the Pre-processing is done the system segments the objects and labels them accordingly. After image segmentation the system then does post processing on it and displays the final output image or output data.

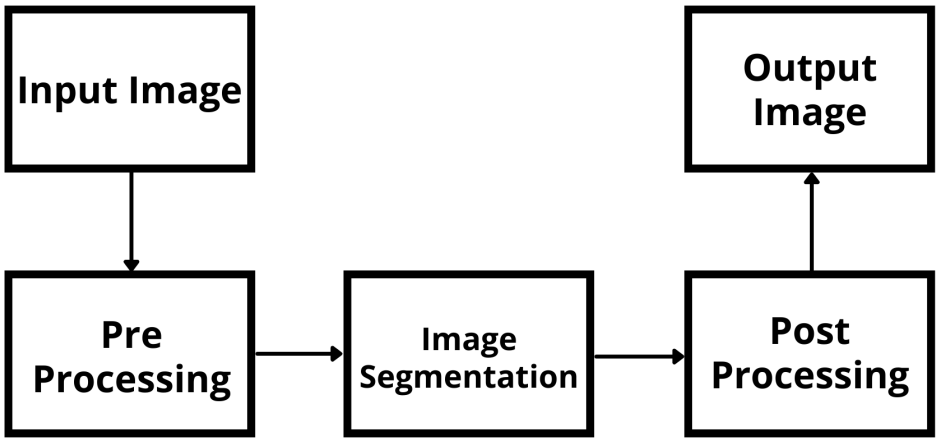

Figure 2: Block Diagram

Requirement of Resources and Limitations

Requirement: Python IDE, Datasets, Device with camera functionality, Software to develop GUI for user end.

Limitations: Limited detection due to less datasets being used in training.

Impact analysis

1. Impact of project on society:

Positive Impact of project on society:

- The system can be very helpful in the healthcare sector, as when properly trained can detect various types of tumours.
- The system can be implemented into autonomous systems for better decision making.

Negative Impact of project on society:

There is no such negative impact of project on society, however if not properly trained can lead to false results, which in turn can hamper the specific application where it is being implemented.

2. Impact of project on environment

Positive Impact of project on environment:

- Deep Learning-based Image Segmentation can be successfully applied to segment satellite images in the field of remote sensing, including techniques for urban planning or precision agriculture. This in turn helps preserving the environment.
- Images collected by drones (UAVs) can be segmented using Deep Learning based techniques, offering the opportunity to address important environmental problems related to climate change.

Negative Impact of project on environment:

No negative impact of the project on environment has been observed as of now, as the project is more oriented for society, and towards preserving the environment.

Professional ethical practices to be followed

1. Giving credits.

2. Keeping technical guidelines in mind.

3. Following the norms of engineering Practices.

4. Liability for outcome caused by one's actions or decisions.

Project Planning (Timeline analysis and Finance management

Table 1: PROJECT PLANNING: (Timeline Analysis)

Parameters	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12
Formation of group	Y											
Identification of problem statement		Y										
Literature survey		Y	Y									
Objectives and outcomes			Y									
Proposed system (methodology)				Y								
Simulation/ Testing)					Y	Y	Y					
Developing final product								Y	Y			

Table 1: PROJECT PLANNING: (Timeline Analysis)

Parameters	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12
Testing and verification										Y		
Report writing											Y	Y
Final presentation/ demo												Y
(color the cells selected for the task/parameter to be perform												

Future scope and further enhancement

The developed model, can be used in many areas:

1. It can be implemented into security systems for enhanced security surveillance as it will be much more effective when the faces are distinguishable from noisy objects.

2. The system can be developed further to be implemented into satellite systems. This in turn can enhance GPS applications

3. The system when implemented for healthcare applications, can be used for medical imaging and hence be used to detect different types of tumours

4. The system can be used in self-driving cars for giving easy distinctions between various objects. Be it traffic signals, signboards, humans, and cars. It can help the driving instruction algorithm to better assess the surrounding before generating the next instruction.

5. The system can be used to detect defects in circuit board. A company has to bear the responsibility of defected devices. If a camera backed with an Image Segmentation model keeps scanning for defects produced in the final product, a lot of money and time can be saved in fixing a defective device.

Expected Outcome

A web based application which can analyze the images captured by the camera and detect their classification using the machine learning model into obbjects, humans, and various other types.

References

[1] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos, "Image Segmentation Using Deep Learning: A Survey" *IEEE Transactions on Pattern Analysis and Machine Intelligence* , February,2021.

[2] R. Szeliski, "Computer vision: algorithms and applications:" *Springer Science and Business Media*, 2010.

[3] D. Forsyth and J. Ponce, "Computer vision: a modern approach: "*Prentice Hall Professional Technical Reference,* 2002.

[4] N. Otsu, "A threshold selection method from gray-level histograms:" , *IEEE transactions on systems, man, and cybernetics* , vol. 9,no. 1, pp. 62 to 66, 1979.

[5] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Transactions on pattern analysis and machine intelligence* , vol. 26, no. 11, pp. 1452 to 1458, 2004.

[6] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using k-means clustering algorithm and subtractive clustering algorithm,"*Procedia Computer Science,* vol. 54, pp. 764 to 771, 2015.

[7] L. Najman and M. Schmitt, "Watershed of a continuous function," *Signal Processing,* vol. 38, no. 1, pp. 99 to 112, 1994.

[8] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International journal of computer vision,* vol. 1, no. 4, pp. 321 to 331, 1988.

[9] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 23, no. 11, pp. 1222 to 1239, 2001.

[10] N. Plath, M. Toussaint, and S. Nakajima, "Multi-class image segmentation using conditional random fields and global classification," *in Proceedings of the 26th Annual International Conference on Machine Learning.* ACM, 2009, pp. 817 to 824.

[11] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE transactions on image processing,* , vol. 14, no. 10, pp. 1570 to 1582, 2005

[12] S. Minaee and Y. Wang, An admm approach to masked signal decomposition using subspace representation, *IEEE Transactions on Image Processing* , vol. 28, no. 7, pp. 3192 to 3204, 2019.

[13] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:* 1706.05587, 2017