Performance Analysis of Prominent Object Detection Algorithms

Prepared by

Group – 20

Mst. Sumiya Siddika	----	1975
Md. Masud Rana	----	2015
Shariful Islam	----	2020

Thesis Supervisor

Dr. Fahima Tabassum

Professor

Institute of Information Technology

Contents

- Introduction
- Deep Learning
- Methodology
- Algorithms
- Conclusion and Future Work

Introduction

A computer vision task called object detection includes finding and recognizing things within a picture. Object detection aims to build bounding boxes around objects in order to offer information about their exact positions in addition to identifying which objects are there in the visual input. Through object detection, machines can identify and locate items in photos. When it comes to automating procedures that call for interaction with the real world, object detection is essential. There are different techniques that are used in object detection like: CNN (Convolutional Neural Network), YOLO (You Only Look Once), D2Det (The deformable two-stage detector), SSD (Single Shot Multibox Detector).

Problem Statement

- The main purpose of our study is to detect objects using four popular algorithms named CNN, YOLO, SSD and D2det.
- We will find the strengths, limitations, and comparative effectiveness of these algorithms in addressing the evolving challenges of accuracy, speed, and versatility.

Objective

The main objectives are to:

- Use four algorithms named CNN, YOLO, SSD and D2Det in object detection.
- Discuss those algorithms in a detailed way.
- Evaluate and compare selected algorithms on the basis of their performance, Speed and Accuracy.

Deep Learning

Deep learning is a subfield of machine learning that focuses on the development and training of artificial neural networks, which are computational models inspired by the structure and function of the human brain. The term "deep" refers to the use of multiple layers in these neural networks. Deep learning has gained significant attention and popularity due to its ability to automatically learn hierarchical representations of data through the training process.

Concepts of Deep Learning include:

- **Neural Networks**: Deep learning relies on neural networks, which are composed of layers of interconnected nodes or artificial neurons. These layers include an input layer, one or more hidden layers, and an output layer.
- **Deep Architectures**: Deep learning models often have many hidden layers, allowing them to automatically learn intricate and hierarchical features from the input data.
- **Training**: Deep learning models are trained on large datasets using optimization algorithms. During training, the model learns to adjust its parameters (weights and biases) to minimize the difference between predicted outputs and actual outputs (labels) in the training data.
- **Activation Functions**: Activation functions introduce non-linearity into the neural network, enabling it to learn complex relationships in the data. Common activation functions include sigmoid, tanh, and rectified linear unit (ReLU).

- **Convolutional Neural Networks (CNNs):** CNNs are a specific type of deep learning architecture designed for processing grid-like data, such as images. They use convolutional layers to automatically learn spatial hierarchies of features.

- **Recurrent Neural Networks (RNNs):** RNNs are designed for processing sequential data. They have connections that form cycles, allowing them to capture dependencies over time. Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) are popular types of RNNs.

- **Transfer Learning:** Deep learning models can benefit from transfer learning, where a pre-trained model on one task is adapted for a different but related task. This is especially useful when working with limited labeled data.

Methodology

- Algorithms
- System Models

Algorithms

We will use four different algorithms for object detection:-

- CNN (Convolutional Neural Network)
- D2Det (The deformable two-stage detector)
- YOLO (You Only Look Once)
- SSD (Single Shot Multibox Detector)

Basic theory of Algorithms

CNN (Convolutional Neural Network): A subclass of deep neural networks called CNNs is specially made to handle organized grid data. They have been widely used in the processing, analysis, and recognition of images and videos. CNNs use a type of perceptron with multiple layers that are intended to require very little preprocessing.

1. Convolutional Layer: The convolutional layer, which produces the majority

of the network's computations, is the fundamental layer used to build convolutional neural networks.

Basic theory of Algorithms

2.Activation Layer: Artificial neural networks can be filled with an Activation Function to aid in the network's ability to recognize complex patterns in data. Rectified Linear Units (ReLU), Randomized LeakyReLUs (RReLU), Exponential Linear Units (ELU), and others are examples of common activation functions.

3.Pooling Layer: Convolutional neural networks frequently use it as one of their constituent parts. In order to reduce overfitting, the amount of data and parameters are compressed by placing the pooling layer between successive convolutional layers. The pooling layer's primary job is to compress images if that's the input type.

Basic theory of Algorithms

Working procedure of CNN:

- Import the necessary libraries.
- Set the parameter.
- Defines the kernel.
- Load the image and plot it.
- Reformat the image.
- Apply convolution layer operation and plot the output image.
- Apply activation layer operation and plot the output image.
- Apply pooling layer operation and plot the output image.

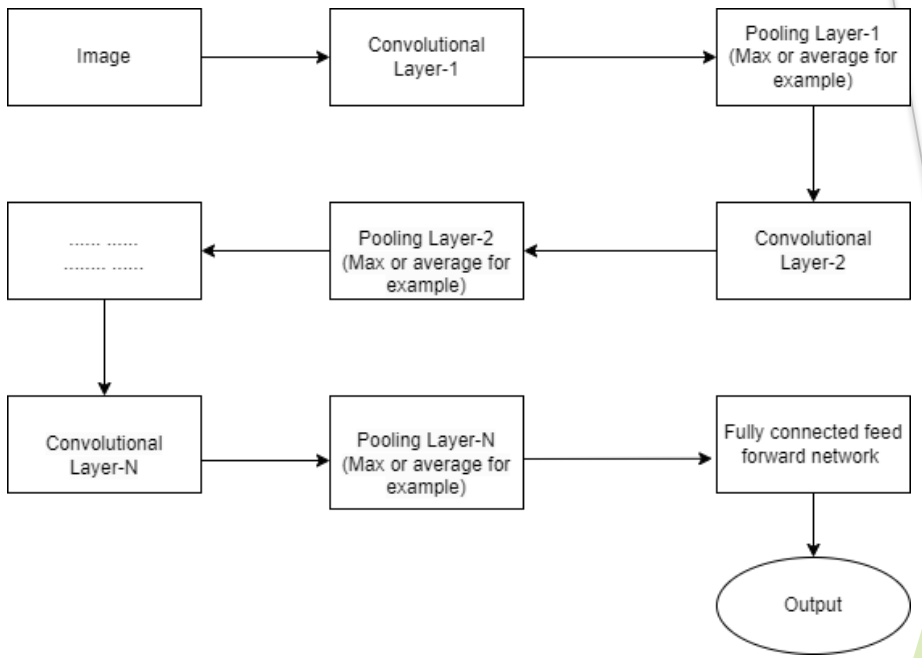

Figure : Flow chart of CNN network

Basic theory of Algorithms

Applications of CNN:

Image Classification: CNNs are frequently used in image classification applications, where the network is trained to classify images into various groups.

Object Detection: CNNs are useful for locating and identifying objects in pictures. CNNs are used in popular object detection frameworks such as Faster R-CNN and YOLO (You Only Look Once).

Basic theory of Algorithms

D2Det:

A deformable two-stage detector is a type of object detection model used in deep learning for computer vision tasks. It combines two key concepts:

1.Two-stage architecture: Object detection models typically follow a two-stage approach. In the first stage, they generate region proposals, which are potential bounding boxes that may contain objects. In the second stage, these proposed regions are classified and refined to produce the final object detections.

2.Deformable Convolutional Networks (DCN): Deformable Convolutional Networks are a type of convolutional layer that can adaptively adjust their receptive fields based on the features within an image. This allows them to capture more accurate and flexible information about object shapes,

Basic theory of Algorithms

Working procedure of D2Det algorithm:

- Input Image
- Feature Extraction
- Region Proposal Network (RPN)
- Anchor Scoring
- Anchor Refinement Predict
- Anchor Selection Filter
- Region-of-Interest (RoI)
- Deformable Convolutional Networks (DCN)
- Object Classification
- Bounding Box Regression
- Post-processing
- Output

Basic theory of Algorithms

YOLO:

YOLO is an algorithm that provides real-time object detection using neural networks. The accuracy and speed of this algorithm make it popular. The abbreviation YOLO stands for "You Only Look Once." This algorithm finds and recognizes different objects in an image. YOLO employs object detection as a regression problem, resulting in the class probabilities of the identified images.

Working procedure of YOLO algorithm:

YOLO algorithm works using the following three techniques:

- Residual blocks
- Bounding box regression
- Intersection Over Union (IOU)

Basic theory of Algorithms

Residual blocks

First, the image is divided into various grids. Each grid has a dimension of S x S.

The following image shows how an input image is divided into grids. Bounding box regression A bounding box is an outline that highlights an object in an image. Every bounding box in the image consists of the following attributes:

- Width (bw)
- Height (bh)
- Class (for example, person, car, traffic light, etc.)- This is represented by the letter c.
- Bounding box center (bx,by)

Basic theory of Algorithms

Intersection over union (IOU)

In object detection, the phenomenon known as intersection over union (IOU) characterize show boxes overlap. YOLO creates an output box that exactly covers the objects by using IOU. The task of predicting the bounding boxes and their confidence scores falls on each grid cell. If the expected and actual bounding boxes match, the IOU is equal to 1. Bounding boxes that are not equal to the actual box are remove by this mechanism.

Basic theory of Algorithms

SSD:

SSD operates as a one-shot detector. It predicts the boundary boxes and the classes directly from feature maps in a single pass and does not have a gave region proposal network. The SSD object detection composes of 2 parts:

1)Extract feature maps.

2)Apply convolution filters to detect objects.

Basic theory of Algorithms

Working procedure of SSD:

Here’s an overview of how an SSD works in deep learning:

- Input Image
- Feature Maps
- Multi-scale Feature Fusion
- Default Anchor Boxes
- Object Detection Head
- Non-Maximum Suppression (NMS)
- Output

Proposed System Model

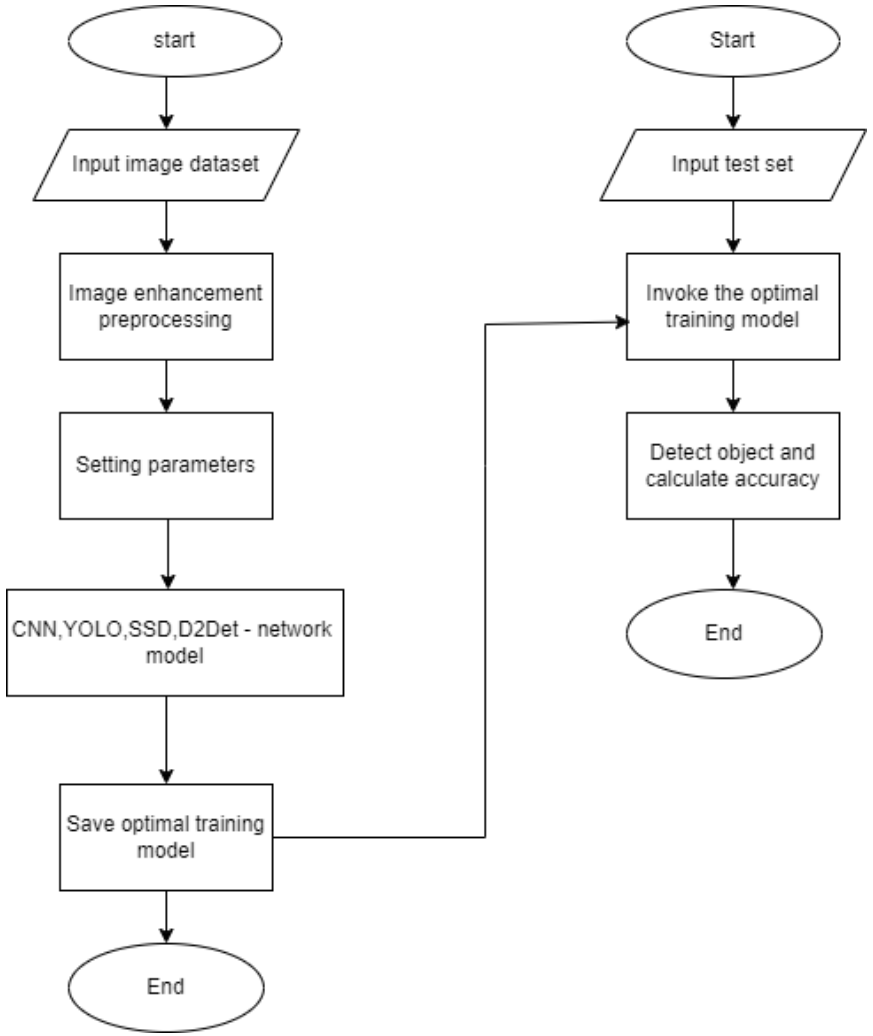

Figure: Proposed System Model

Conclusion and Future Work

Conclusion:

Comparing CNN (Convolutional Neural Networks), YOLO (You Only Look Once), D2Det (Deformable Two-Stage Detector), and SSD (Single Shot MultiBox Detector) algorithms for object detection in deep learning involves considering their strengths, weaknesses, and use cases. Each algorithm has its unique characteristics and is suited for specific scenarios.

Conclusion and Future Work

Future work

Merge Algorithms: We will combine these algorithms and will see whether it gives a better accuracy and result. We will experiment this on various object and conditions. Thus we can make a device that will help others in object detection.

Real images: We will experiment it on real images. We will see the effect of using this on real images. We will observe the past affect.

Disease Detection: Whether a device can be made so that it can say which disease it belongs to.

Leaf names: Will try to make a device that will recognize a leaf names.

Road sign: Weather it can detect road sign from moving objects.


thank you