OBJECT DETECTION AND IMAGE CLASSIFICATION USING DEEP LEARNING MODEL

Gudla Swathi.[1], Prof. Kompella Venkata Ramana[2]

[1]PG-Scholar, DEPARTMENT OF COMPUTER SCIENCE AND SYSTEMS ENGINEERING, COLLEGE OF ENGINEERING (A), ANDHRA UNIVERSITY, VISAKHAPATNAM-530003

Email: -Swathigudla27@gmail.com

[2]professor, DEPARTMENT OF COMPUTER SCIENCE AND SYSTEMS ENGINEERING, COLLEGE OF ENGINEERING (A), ANDHRA UNIVERSITY, VISAKHAPATNAM-530003

Email: - Kompellavr@yahoo.com

Abstract: Automatic picture categorization has emerged as the most difficult topic in the field of computer vision in recent years due to the rapid increase in the identification of digital materials. Comparatively speaking to human eyesight, automatic picture understanding and analysis by a system is challenging. Numerous studies have been conducted to resolve issues with the current categorization system, but the results were limited to simple image primitives. Object detection means detect object in image and classify what it is. In this study, we employ a deep learning method to produce the desired outcomes in the field of computer vision. Convolutional Neural Networks (CNN), a machine learning algorithm, are employed by our system to automatically classify the objects in images. As a benchmark for classifying objects in photos, our system uses various labels from a data set of photographs. The enhanced parallel processing capacity offered by graphics processing unit. The reader will comprehend the specifics of the cutting-edge object detection methods including faster region convolutional neural network (Faster RCNN), Common Objects in Context, and others.

Index Terms: CNN, FRCNN, BBOX, GPU, IMAGE CLASSIFICATION, SVM

I Introduction

Automatic object classification has emerged as one of the visual information indexing and retrieval systems' most important difficulties in recent years as a result of the enormous growth of digital content. An interdisciplinary area of artificial intelligence called computer vision tries to give machines the same ability as people do to grasp information from images. There have been numerous research attempts to solve these issues, but these techniques only take into account the low-level properties of image primitives. Processing the photos will not be aided by concentrating on low-level visual attributes. A long-standing issue in computer vision is the identification and classification of objects in a picture. While classifying and comprehending images is a simple process for humans, it may be quite expensive for computers. A group of pixels make up each image in general, and each pixel is represented by a separate value. The computer will now require additional storage space in order to store an image. It needs execute a greater number of calculations in order to classify photos.

Higher configuration systems with more processing power are needed for this. Making judgments based on input in real time is not feasible since it takes more time to do these numerous computations and produce results. it was addressed how to use the convolutional neural network (CNN) deep learning concept to extract features from hyperspectral images (HSI). It leverages the several pooling layers in CNN to extract the nonlinear, invariant features from the HIS that are helpful for accurate image classification and target recognition. Additionally, it deals with the general problems with the HSI image features. Engineering-wise, it aims to automate activities that the human visual system can perform. It is focused on automatically extracting photos, analyzing them, and deriving information from them. Several methods for classifying images were described and compared over the past ten years. However, in general, image classification refers to the process of obtaining data from an image by categorizing its pixels. Unsupervised classification and supervised classification are both viable options discuss the application of the unsupervised learning method in a framework for underwater fish recognition and image classification.

Convolutional layer parameters can be learned, and it is a fundamental building piece. Although the width and height of each filter are tiny, they cover the entire depth of the input volume. Dot products are used to create the 2-dimensional activation map between each filter's input and entries. As a result, the network picks up filters that turn on when it spots a certain feature in a particular location in the input. Without erasing any of the image's information, the pooling layer is utilized to down sample the image. The maximum value from the cluster of neurons from the previous layer is used in max pooling. Every neuron in one layer is linked to every other neuron in the completely connected layer. When compared to conventional classification algorithms, which involve hand-engineered filters, CNNs employ very little pre-processing. One

benefit of CNN is its independence from human interference in learning filters. CNN is a supervised deep learning approach that needs a lot of tagged training data. After training, the model will pick up the weights, improving the classifier's accuracy.

The classifier then displays the class to which an image belongs after receiving it as input. The Google self-driving car is a cutting-edge deep learning experiment from the Google firm and an illustration of the most recent advancement in artificial intelligence. For this project, real-world image data is used as an input, and decisions are made based on the knowledge gathered from the images. CNN is a supervised deep learning approach that needs a lot of tagged training data. After training, the model will pick up the weights, improving the classifier's accuracy.

2. **Literature survey**

A summary of the literature review is provided in this chapter. Some of the researchers' pertinent work is represented in this chapter. Researchers have investigated a variety of currently available strategies; a few of them are covered below.

A common research topic in the fields of deep learning, pattern recognition, and human computer interaction, image categorization has drawn a lot of interest from research experts. Classifies photos by taking the image's features and analyzing them. Here, they underline that the mechanism behind the picture composition also has a significant impact on the extraction of features from the image, whereas most midlevel feature learning methods often concentrate on the process of coding or pooling. Hierarchical image decomposition is a useful method for exploring image content during feature extraction. Each image in this instance has been broken down into a number of semantic components, such as the structure and texture images. Various feature extraction techniques can be used to match the semantic image content (structure and texture) with other images. The two following schemas are used to describe various picture property-related characteristics, including hand-crafted features used in single staged networks and the second one that automatically learns features from raw pixels through multistage networks. has discussed utilising a biologically stimulated model to classify natural photos in [2]. It makes use of well-known similar developments in the visual information system and inference methods of the human brain. This approach is mostly used for natural categorization and image analysis. This system is made up of three crucial components: a clustering of visual information unit, a unit for knowledge organisation, and a unit for biologically inspired visual selective attention. It automatically extracts significant associations between photos using low-level information in the images. In order to classify images more accurately, the system mimics the limitations of the human visual system. The biologically inspired system consists of two parts: a top-down selective attention module that makes decisions on intriguing items through human interaction, and a bottom-up saliency map module that creates a salient area from the low-level properties retrieved from natural images. These two elements closely mimic the workings of the human brain's visual what pathway and where pathway. The grouping of visual information is accomplished by employing output from the knowledge clustering unit and is based on a high-level top-down visual information perception and classification model. These components have been merged into the knowledge structuring unit. The goal of biometrics is to automatically discriminate between subjects in a trustworthy manner and as per target applications based on one or more signals derived from traits like face, fingerprint, iris, voice, or written signature. In [3], spoof finger print detection using convolutional neural network (CNN) is discussed. Biometric technology offers greater benefits than conventional security measures based on something we physically possess or can physically recall, such as a key, card, or PIN. Many fingerprint detection algorithms have been presented in recent years, and they can be categorized into two groups: hardware and software. To detect a human's living attribute, such as blood pressure or heart rate, a specialized device is attached to any hardware sensor in hardware techniques. To differentiate between authentic and false fingerprints, fingerprint image features are used. Two feature extractors—Convolution Networks and Local Binary Patterns—have been applied to this model. Support vector machine classifier (SVM) is a temporary tool that may be used in conjunction with both methodologies to classify real and false finger prints. To train the model, this system uses a dataset made up of photos of actual and false fingerprints that were collected from various sensors. The following sensors—Biometrika FX2000, Digital 4000B, Italdata ET10—are used to collect fingerprint data. Faux prints were made using a variety of materials, including gelatin, wood glue, eco-flex, and silgum. The classifier was trained using the four stages listed below, including

1. Employing picture reduction, contrast equalization, filtering, and region of interest extraction to pre-process data.
2. LBP and Convolutional Network are two approaches used for feature extraction.
3. Dimensionality reduction and data normalization.
4. Utilizing SVM to classify the photos.

3 Implementation Study

The object detection using regional convolution layer is to detect the objects present in image by training and through RCNN model. To circumvent the problem of selecting a huge

number of regions, Ross Girshick et al. proposed a method where we use the selective search for extract just 2000 regions from the image and he called them region proposals. Therefore, instead of trying to classify the huge number of regions, you can just work with 2000 regions. These 2000 region proposals are generated by using the selective search algorithm which is written below.

Selective Search:

- Generate the initial sub-segmentation, we generate many candidate regions
- Use the greedy algorithm to recursively combine similar regions into larger ones
- Use generated regions to produce the final candidate region proposals

Convolutional networks used across the dataset to extract patterns from local portions of the input images. Typically, feature extraction in texture descriptors uses local binary patterns (LBP). Both methods are used to extract features, which are then applied separately in the pipeline.

Numerous methods about regional face features have been given and contrasted over the past ten years

And picture categorization using a geometric model of the human face. Other methods concentrate on techniques for finding the local sub feature of images via template matching. By classifying an image as a face image or a non-facial image, this system can identify semi-frontal human faces in large datasets of complicated images. Additionally, CNN used to automatically extract the key details from the photos. CNN processes the images using three different types of layers. They are classification layers with sigmoid neurons, convolution layers, and sub sampling layers. There are a predetermined number of planes in the convolutional layers. Every plane is taken into account as a feature detector. Locate the photos after locating the image feature. Subsampling layer is then used to reduce input dimension while keeping the image's details. The classification procedure is then carried out using the sigmoid neurons' function. In many different fields, it is essential to accurately identify the target object and track it while managing occlusions and other added complexity. Numerous scholars have experimented with different ways to object tracking (Almeida and Guting 2004, Hsiao-Ping Tsai 2011, Nicolas Papadakis, and Aurelie Bugeau 2010). The application domain has a significant impact on the approaches' character. The following is a representation of some of the research projects that evolved into suggested work in the area of object tracking.

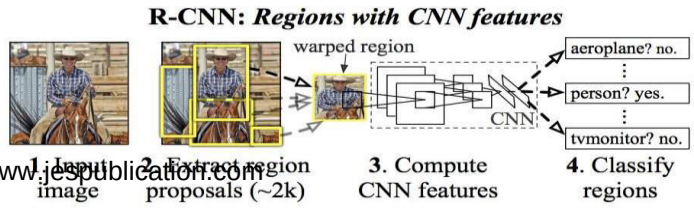

Fig1: R-CNN

3.1 Proposed Methodology

The aim of the proposed system is to detect the object present in image using Faster RCNN approach. After the improvement in architecture of object detection network in R-CNN to Fast R_CNN. The training and detection time of the network decrease considerably, but the network is not fast enough to be used as a real-time system because it takes approximately (2 seconds) to generate output on an input image. The bottleneck of architecture is a selective search algorithm.

The proposed algorithm is consisting of :

- Image collection
- Data set splitting
- Train CNN
- Classify using CNN

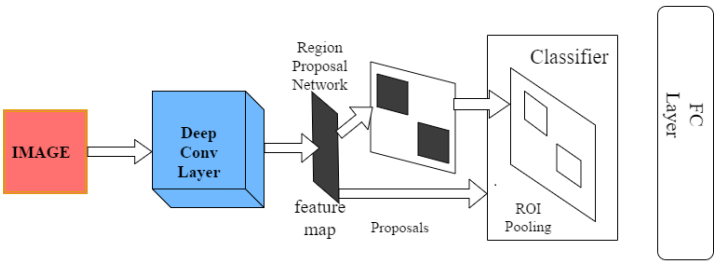

Fig2: System Architecture

4. Methodology

- The below are the steps followed in a Faster RCNN approach: An input image is passed to the convolutional network which then returns the feature map for that image.
- Region proposal network is applied on those feature maps. The result of this step is theset of object proposals.
- A ROI pooling layer is applied on those proposals to make all the proposals same size.
- The object proposals are passed to a fully connected layer that uses a SoftMax layerand a linear regression layer to classify and output the bounding boxes for objects.
- The feature maps are passed to a fully connected layer which has a SoftMax and a linear regression layer. It classifies the object and predicts the bounding boxes for the identifiedobjects.
- The output is the classified image of the CNN networks and the information about the classified image. This shows the importance of the deep learning neuralnetworks

in image classification.

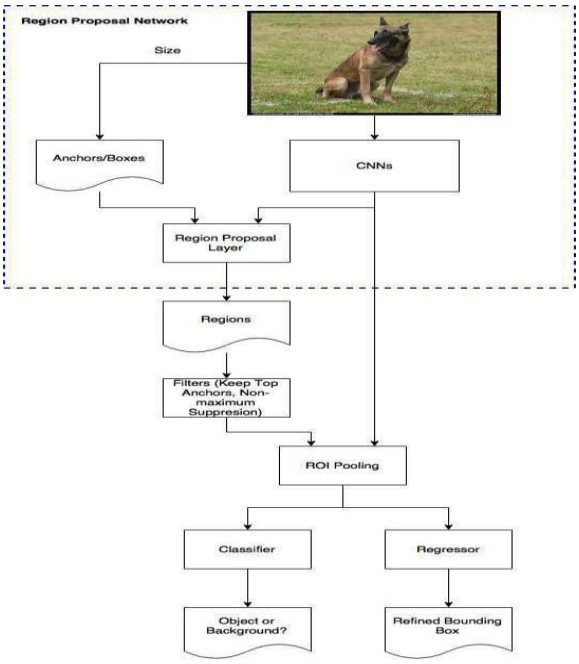

Fig 3:-proposed Methodology

5 Results and Evolution Metrices

Output of bounding boxes values of particular image

```
        names.extend(xml.names)
        boxes.extend(xml.boxes)
        image_id.extend([xml.image_id] * len(xml.names))
        xml_path.extend([xml.xml_file] * len(xml.names))
        img_path.extend([xml.img_path] * len(xml.names))
    a = {"image_id": image_id,
         "names": names,
         "boxes": boxes,
         "xml_path":xml_path,
         "img_path":img_path}

    df = pd.DataFrame.from_dict(a, orient='index')
    df = df.transpose()

    return df

df = xml_files_to_df(XML_FILES)
df.head()
```

	image_id	names	boxes	xml_path	img_path
0	2007_000027.jpg	person	[174, 101, 349, 351]	/content/drive/MyDrive/dataset/VOC2012/Annotat...	/content/drive/MyDrive/dataset/VOC2012/JPEGIma...
1	2007_000032.jpg	aeroplane	[104, 78, 375, 183]	/content/drive/MyDrive/dataset/VOC2012/Annotat...	/content/drive/MyDrive/dataset/VOC2012/JPEGIma...
2	2007_000032.jpg	aeroplane	[133, 88, 197, 123]	/content/drive/MyDrive/dataset/VOC2012/Annotat...	/content/drive/MyDrive/dataset/VOC2012/JPEGIma...
3	2007_000032.jpg	person	[195, 180, 213, 229]	/content/drive/MyDrive/dataset/VOC2012/Annotat...	/content/drive/MyDrive/dataset/VOC2012/JPEGIma...
4	2007_000032.jpg	person	[26, 189, 44, 238]	/content/drive/MyDrive/dataset/VOC2012/Annotat...	/content/drive/MyDrive/dataset/VOC2012/JPEGIma...

Fig4: Boxes columns consists xmin, xmax ,ymin, ymax values of box

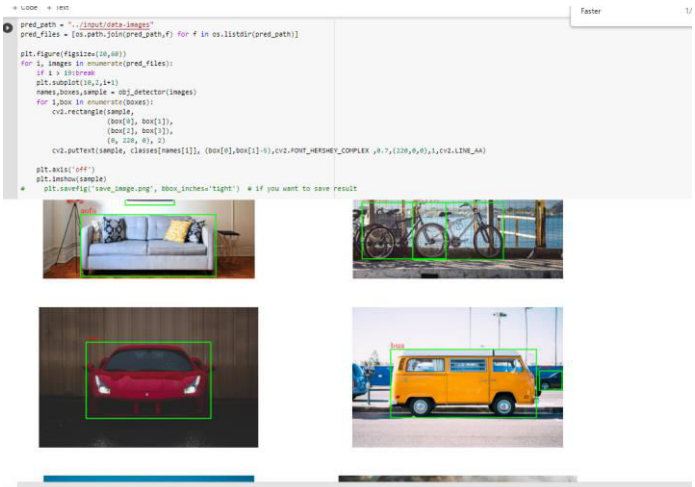

Fig5: after training using FRCNN model and using coco data set the result of output of object detection

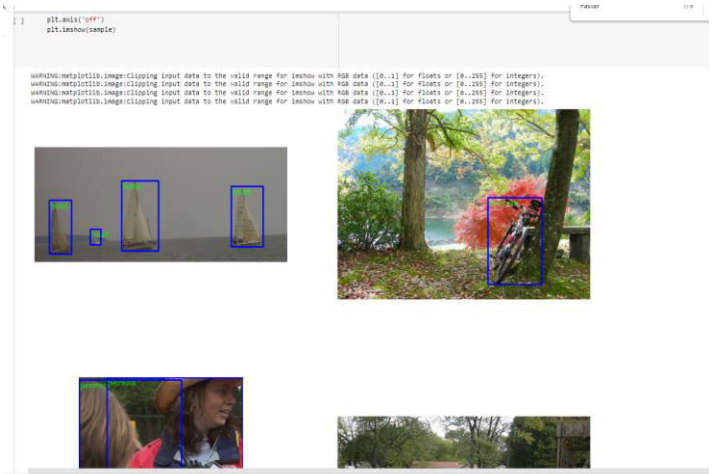

fig 6: -detected object we can see in boundary boxes

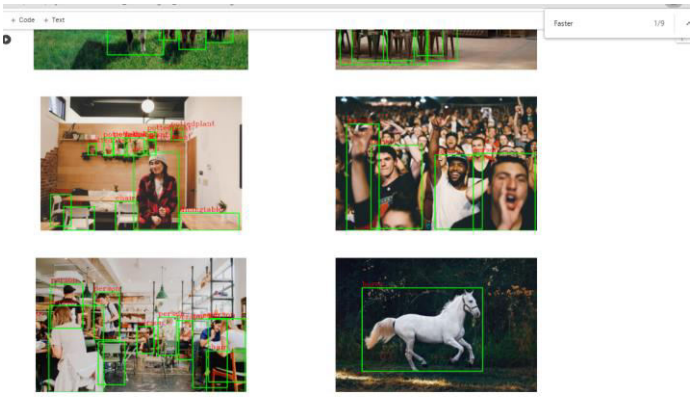

fig 8: -detected multiple objects using boundary boxes

6 Conclusion

In this project, we used Convolutional Neural Networks (CNN) for image classification. This data sets used both and training and testing purpose using CNN. It provides the accuracy rate more than 90%. Images used in the training purpose are small and Grayscale images. The computational time for processing

these images is very high as compare to other normal JPEG images. Stacking the model with more layers and training the network with more image data using clusters of GPUs will provide more accurate results of classification of images. The future enhancement will focus on classifying the colored images of large size and its very useful for image segmentation process.

7 References

1.PARINITA BADRE, SIDDHANT BANDIWADEKAR, PRACHI CHANDANSHIVE, AAKANKSHA CHAUDHARI, SONALI JADHAV. Automatically Identifying AnimalsUsing Deep Learning. In: 2018 International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE). Vol. 6, Issue 3, March **(2018).** ISSN(Online): 2320-9801 ISSN (Print): 2320-9798.

2.GULLAL SINGH CHEEMA, SAKET ANAND. Automatic Detection and Recognition ofIndividuals in Patterned Species. IIIT – Delhi.Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. In November **(2017)**. arXiv:1703.05830v5.

3.TIBOR TRNOVSZKY, PATRIK KAMENCAY, RICHARD ORJESEK, MIROSLAV BENCO, PETER SYKORA. Animal Recognition System Based on Convolutional Neural Network. In September **(2017)**. DOI: 10.15598/aeee.v15i3.2202.

4.SAMER HIJAZI, RISHI KUMAR, CHRIS ROWEN. Using Convolutional Neural Networks for Image Recognition. IP Group, Cadence.

5.SHAOQING REN, KAIMING HE, ROSS GIRSHICK, JIAN SUN. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497v3 [cs.CV] 6 Jan **(2016).**

6.HU W, HUANG Y, WEI L, ZHANG F, LI H (2015) Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors 2015.

7.REN, S., HE, K., GIRSHICK, R., SUN, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: NIPS. pp. 91–99 (2015)

8.SIMONYAN, K., ZISSERMAN, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

9.KRIZHEVSKY, A., I. SUTSKEVER and G. E. HINTON. ImageNet classification with deep convolutional neural networks. Annual Conference on Neural Information Processing Systems (NIPS). Harrah's Lake Tahoe: Curran Associates, 2012, pp. 1097–1105. ISBN 978- 1627480031