



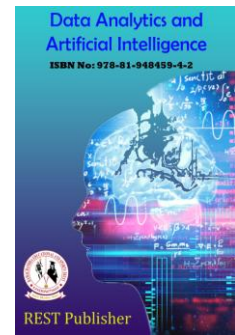
Data Analytics and Artificial Intelligence

Vol: 3(2), 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: <http://restpublisher.com/book-series/daai/>

DOI: <https://doi.org/10.46632/daai/3/2/27>



Object Recognition in Earth Surface Satellite Images Using Digital Image Processing and Machine Learning Techniques with Big Data Technologies

Misba Khan.k

Independent Researcher, Bengaluru, Karnataka, India.

Corresponding Author Email: misbamr711@gmail.com

Abstract. Detection of an object from a satellite image is a difficult process because the presence of objects in a satellite image is unpredictable. Different approaches have been available to detect vehicles, buildings, trees however all these objects were detected individually through machine learning and some other methods. Similarly accuracy in object detection is another major issue. In our proposed work, To analyze the object accurately, Polygon approach is implemented which includes both shape and color as input and processes it with datasets to attain maximum accurate result. Here image parameters have been extracted accurately through feature detection. After segmentation of a particular object from image CNN classification is implemented. Through this, in our proposal we are going to detect roads, trees, buildings, waterway and few other objects accurately with this single approach.

Keywords: Image processing, Object detection, Feature Extraction, Big data technique, Machine Learning.

1. INTRODUCTION

According to rapidly growing urbanization and municipal regions, automatic detection of buildings from remote sensing images is a hot topic and an active field of research. Detection of low-dimensional features (edges, corners) plays an important role in computer vision applications, mostly building detection. There are many competing algorithms for detecting corners and edges in images. Machine learning approaches have been quite successful in solving problems of different domains, be it face recognition, gene sequencing, solar radiation forecasting, image segmentation, information retrieval, handwritten digit classification or fingerprint classification. Here the problem of detection of objects from satellite or aerial imagery is to figure out whether a satellite or aerial image has one or more objects that belong to the class of interest and if present locate their positions in the image. With the availability of Defence Science and Technology Laboratory (Dstl) we are able to recognize different ranges of objects and even can be separately identified than ever before which has opened new possibilities in the field of automatic detection of objects in satellite/aerial imagery.

2. MACHINE LEARNING BASED OBJECT DETECTION

With the advance of machine learning techniques, especially the powerful feature representations and classifiers, many recent approaches regard object detection as a classification problem and have achieved significant improvements. The flowchart of machine learning-based object detection, in which object detection can be performed by learning a classifier that captures the variation in object appearances and views from a set of training data in a supervised or semi-supervised or weakly supervised framework. The input of the classifier is a set of regions (sliding windows or object proposals) with their corresponding feature representations and the output is their corresponding predicted labels, i.e., object or not. As can be seen from Fig. 1, feature extraction, feature fusion and dimension reduction (optional), and classifier training play the most important roles in the performance of object detection and hence we mainly focus on reviewing these three crucial steps.

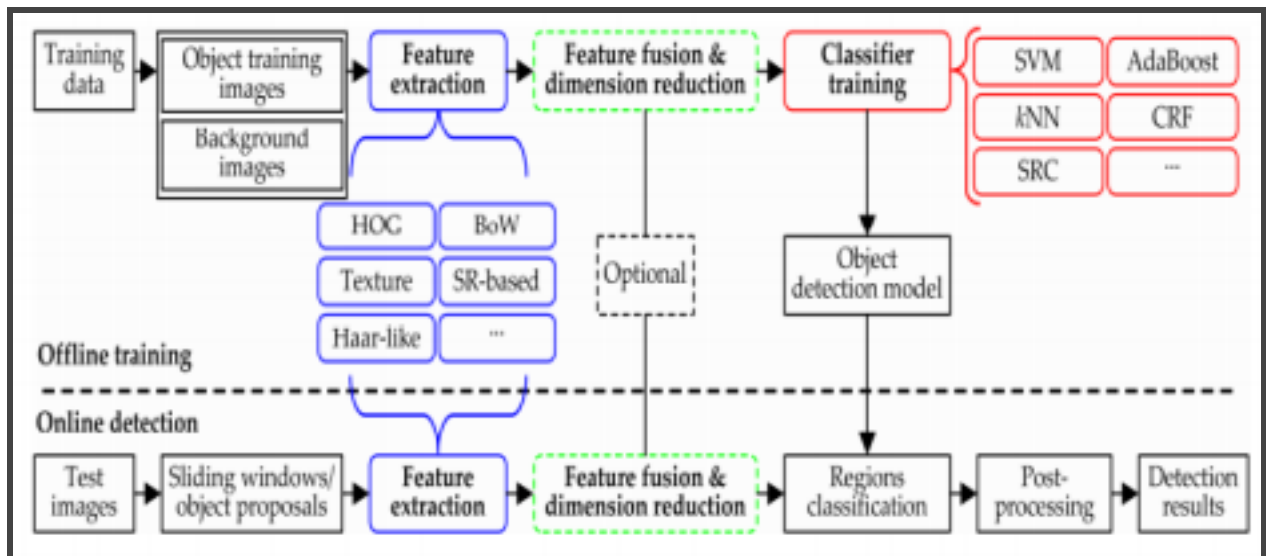


FIGURE 1. The flowchart of machine learning-based object detection

3.SUPERVISED LEARNING

In supervised learning, an algorithm builds a mathematical model from a set of data that contains both the input and desired outputs. These algorithms are trained using labeled examples i.e. input and desired outputs are known. In this learning, the algorithm receives a set of inputs along with corresponding correct outputs. Algorithms learn by comparing its actual output with correct outputs to find out errors. Then, the model is modified accordingly. Some examples of supervised machine learning are nearest neighbor, Naïve Bayes, Decision Tree, Regression Tree etc.

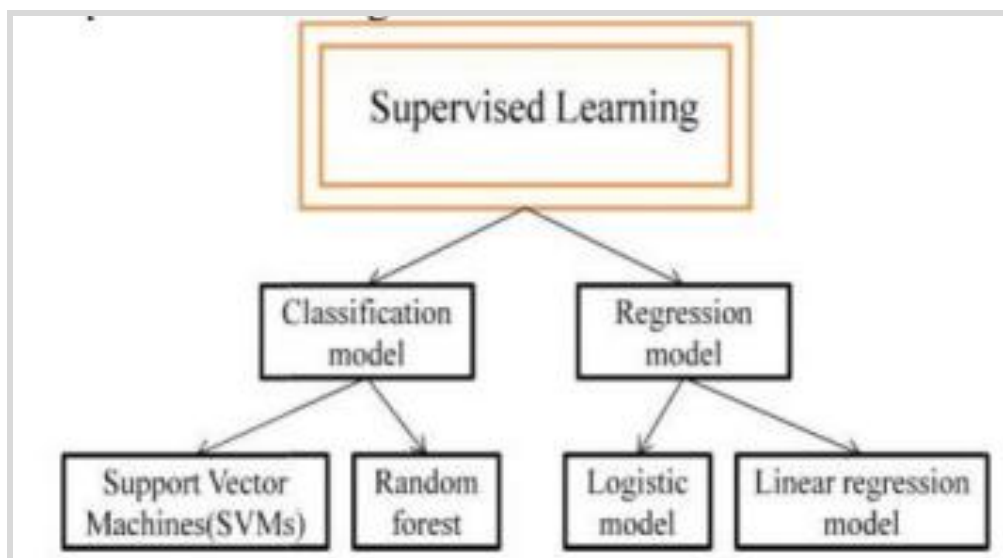


FIGURE 2. Supervised learning

Various supervised learning algorithms for object recognition were identified during background study. Of these, the most commonly used ones are listed below.

- Support Vector Machines (SVM) – Support vector machine is vector based method where the boundaries between the classes in a feature space are identified using the training data in order to perform classification.
- Convolutional Neural Networks (CNN) – Convolutional neural networks are feed forward networks which perform the pooling and convolution calculations alternatively for multiple times. They provide the classification as probability of each class.
- Logistic Regression – Logistic regression is a statistical method for binary classification that uses a sigmoid function. The coefficients for the function are generated from the training data.

- k Nearest Neighbors (k-NN) – k-Nearest neighbors is an instance based learning method where the classification is based on the closeness to the majority of the instances.

4. UNSUPERVISED LEARNING

In unsupervised learning, a mathematical model is to be built from a set of data which contains only inputs. Desired output labels are not present in this type of learning. Unsupervised learning is used against that data which doesn't consist of a historical label. K means Association Rules are an example of such algorithms. Figure describes different methods of unsupervised learning.

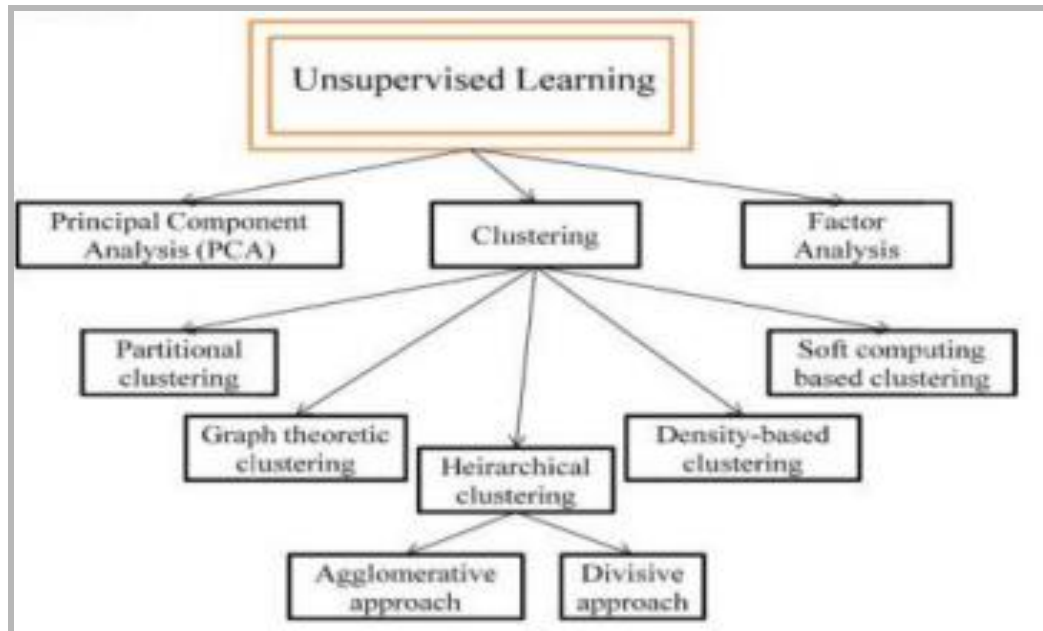


FIGURE 3. Unsupervised learning

K-mean, Clustering approaches, genetic algorithms are some of the unsupervised machine learning techniques. In this learning technique input data is not la-labelled and doesn't have a known result. Models is prepared by deducing structures present in the input data and extract general rules. The main goal of the model is to determine Data patterns/grouping. Clustering is a technique for finding similarity groups in data called clusters. It is often called unsupervised learning as no class values denoting and a derivable grouping of the data instances are given. It refers to a problem of finding hidden structures in the labeled data. It has no measurements of outcome, to guide the learning process.

5. CHALLENGES IN SATELLITE IMAGE PROCESSING

Modern machine learning techniques, and deep learning, in particular, have made tasks like object detection, object counting, semantic segmentation, and generic image classification much more straightforward to create. The field of deep learning was initially focused on the kinds of images in datasets like Image Net (which contains over 1 million individually annotated images scraped from the internet). These images tend to be relatively small files with three “channels” in the visible spectrum (red, green, and blue) and are stored in common file formats like PNG or JPEG. However, in recent years, analogs to ImageNet focused specifically on satellite imagery have filled the void.

6. FORMER RESEARCH

Machine learning is a broad area which encompasses various techniques. Each of these techniques is designed and applicable for specific tasks. In recent years, Convolutional Neural Networks (CNNs), have shown excellent performance in many visual tasks by learning effective feature extraction. This work has been divided into four domains which are Resolution, Segmentation, Object detection and Identification and Registration of unknown class objects. First two domains are considered as pre-processing steps before extraction of objects from satellite images, quality of satellite images increased and segment images into small parts. Other two domains detect the object by using many extraction and structure discovery techniques and pattern matching is also performed in that domain. The previously used techniques in image processing were K-Nearest Neighbor algorithm (KNN), Support Vector Machine (SVM), Genetic Algorithm (GA), Naive-Bayes Classifier.

Objective of the Thesis:

- To detect objects in satellite images efficiently and accurately.
- Result should be shown in bounded box for clear visualization.
- Perform object recognition using the polygon approach, CNN classification on a generated data set
- Satellite images often contain several targets and correct localization of each target is required.

7. PRESENT ANALYSIS

Deep learning has made great progress in solving many problems of pattern recognition and the traditional machine learning algorithm by nonlinear processing. One of the reasons why deep learning can achieve high accuracy and efficiency is because a complicated deep learning architecture can be fitted with a large amount of training data with diversity and variability to form a better deep network structure. In addition, with the rapid development of artificial intelligence (AI), sparse coding and deep learning algorithms are widely used in scene classification because of their high classification accuracy.

Method And Technique: The proliferation of satellite imagery has given us a radically improved understanding of our planet. As these large, complex datasets continue to increase exponentially in number, the Defence Science and Technology Laboratory (Dstl) is seeking novel solutions to alleviate the burden on their image analysts. In this competition, Kagglers are challenged to accurately classify features in overhead imagery. **Jaccard Index:** The Jaccard Index for two regions A and B, also known as the "intersection over union", is defined as:

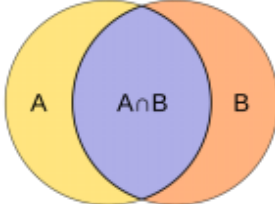
$$Jaccard = \frac{TP}{TP + FP + FN} = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$


FIGURE 4.

Where TP is the true positives area, FP is the false positives area, and FN is the false negatives area. For each object class, of each image, we calculate the TP, FP, and FN areas. Then the Jaccard is calculated for that class using total TP, total FP, and total FN.

For every row in the dataset, submission files should contain the following: ImageId (string), Class Type (int), and Multi polygon WKT (mult ipolygon coordinates in WKT format). You must submit a row for every ImageId and every Class Type. The complete format is in the sample_submission.csv file. If you want to predict an empty polygon for the class, put down MULTIPOLYGON EMPTY. Please submit all the rows in exactly the same order as sample_submission.csv; otherwise you might get a "key not found error". If you have complex polygons, it could take a longer time (4+ mins) to evaluate your submission. Please be patient when the submission is getting evaluated! If you find yourself getting timed out, please consider simplifying (smoothing) your polygons to reduce both the submission file size and the time it takes to evaluate the submission.

Polygon feature extraction: Extracts the cells of a raster based on a polygon by specifying the polygon's vertices.

Classify the Image: Image classification, in its most general form, is the process of sorting or arranging pixels in an image into classes or clusters. When the classes are user-defined, usually by using a series of training areas, the classification is considered "Supervised". When the classes are defined by an automatic process without the user's input, the classification is considered "Unsupervised". The training areas used in a supervised classification can be time consuming to create, but they offer significantly more control over the classification process. Unsupervised classifications don't require the designation of training areas and are therefore quicker and easier to perform, but there is much less control over how the pixel values are grouped.

Separate the Feature from the Surrounding Area: While some classes are now clearly either land or water, many are still found in both regions. This is largely due to the pixel values of the reflections in the water resembling those of the highly reflective rocks on land. The Region Group tool reclassifies the image by grouping adjacent pixels with a common value together. After each area of reflection and each rocky area have unique values, they can be treated separately.

Results of the Region Group tool: Many of these new groups are too small to be integral to defining the shape of the island. Using the Identify tool, I selected the smallest area I thought would be important and used Extract by Attribute to create a new raster containing only the cells with count values equal to or larger than the area I identified. The holes created by removing the smaller pieces were then filled using Nibble. You may want to try Extract by Attributes with different thresholds, run Nibble on the outputs, and compare the results to find which gives enough detail to define an outline that is accurate enough for your purpose.

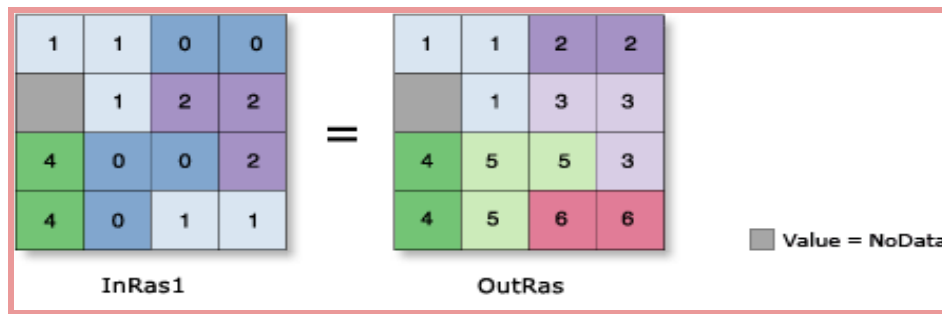


FIGURE 5.

Isolate the Feature: The results of a well-Nibbled raster should be large, clearly defined groups. To create a single polygon outlining the island I converted the raster to a polygon feature class using Raster to Polygon. I assessed each polygon and deleted those which were not part of the island. Which polygons are deleted depends entirely on the goal of the project. I needed to create a polygon defining the perimeter of dry land, so any polygons which fell on the ocean, intertidal zone or ponds were deleted. Once I was satisfied that the remaining polygons were all part of the island, I merged them using the Dissolve tool to create a single polygon representing the boundary of the shape that we are going to identify from the satellite image. Additional attributes from the input raster, if any, will be carried over as-is to the output raster attribute table. Depending on the nature of the property being recorded, some of the attribute values may need to be recalculated. When a multiband raster is specified as the Input Raster (in_raster in Python), all bands will be used. To process a selection of bands from a multiband raster, first create a new raster dataset composed of those particular bands with the Composite Bands tool, and use the result as the Input Raster (in_raster in Python). The default output format is a geodatabase raster. If an Esri Grid stack is specified as the output format, note that the name cannot start with a number, use spaces, or be more than nine characters in length. To extract based on a polygon in a feature class instead of providing a series of x,y pairs, you can use the Extract By Mask tool. The polygon object may consist of just one part, or many parts as a polygon class. In the later case, all polygon parts must be contiguous, such that they could be outlined by one polygon. To extract based on a polygon feature that contains multiple disconnected parts, use the Extract By Mask tool. The center of the cell is used to determine whether a cell is inside or outside a polygon. If the center is within the arcs of the polygon, the cell is considered fully inside, even if portions of the cell fall outside the polygon. The polygon has a limit of 1,000 vertices. Polygon vertices must be entered in a clockwise order. The first and last vertex should be the same to close the polygon. This is especially important if multiple polygons are to be used. In that case, if the last point for each individual polygon is not identical to its starting vertex, polygons may be closed automatically by directly connecting to the first vertex. However, this may produce a result different from what you may have expected, so take care if employing this technique. The arcs of the polygon can cross one another, but convoluted polygons are not recommended.

D convolution: In the digital domain, convolution is performed by multiplying and accumulating the instantaneous values of the overlapping samples corresponding to two input signals, one of which is flipped. This definition of 1D convolution is applicable even for 2D convolution except that, in the latter case, one of the inputs is flipped twice. This kind of operation is extensively used in the field of digital image processing wherein the 2D matrix representing the image will be convolved with a comparatively smaller matrix called 2D kernel.

- **Padding:** In the previous animations the edges essentially get “trimmed off” and the pixels on the edge are never at the center of the kernel, because there is nothing for the kernel to extend beyond the edge. This technique solves padding the edges with extra, “fake” pixels (usually of value 0, hence the oft-used term “zero padding”).
- **Striding:** Often when running a convolution layer, you want an output with a lower size than the input. One way of accomplishing this is by using a pooling layer (eg. taking the average/max of every 2x2 grid to reduce each spatial dimension in half).

RMSprop Optimizer: The RMSprop optimizer is similar to the gradient descent algorithm with momentum. The RMSprop optimizer restricts the oscillations in the vertical direction. Therefore, we can increase our learning rate and our algorithm could take larger steps in the horizontal direction converging faster. The difference between RMSprop and gradient descent is on how the gradients are calculated.

Convolution Neural Network: In this model we have used CNN as a feature extractor and multi-layer perceptron (MLP) as a classifier. The description of the model is given as: The first convolution stage is with a rectified linear unit (ReLU) activation function followed by a max-pooling layer. This layer has 20 convolution filters, each one of them has a size of 5x5.

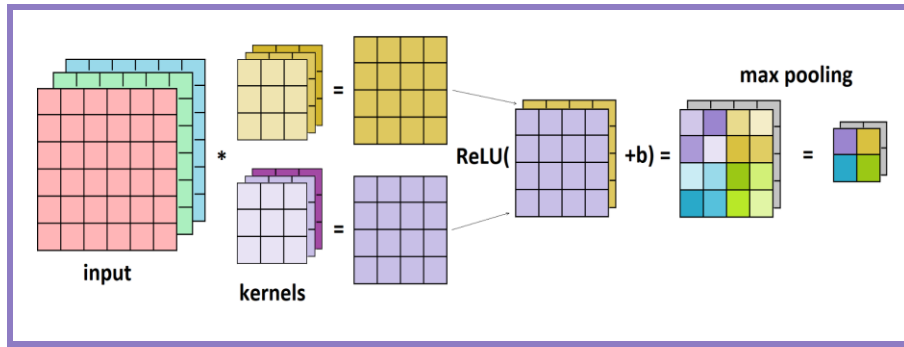


FIGURE 6.

The output dimension is the same as that of the input shape i.e, 64×64. As this is the first layer in our model, we have to define its input shape i.e, (1,64,64).The second convolution layer is also followed by a rectified linear unit activation function and a max-pooling layer.The output layer of the MLP has softmax function as activation function, which can be defined as:

$$P(y = j|X) = \frac{e^{X^T \cdot w}}{\sum_{k=1}^K e^{X^T \cdot w}}$$

Where K is the number of classes and P(y = jjX) is the probability of the input being in class j given its feature vector X.

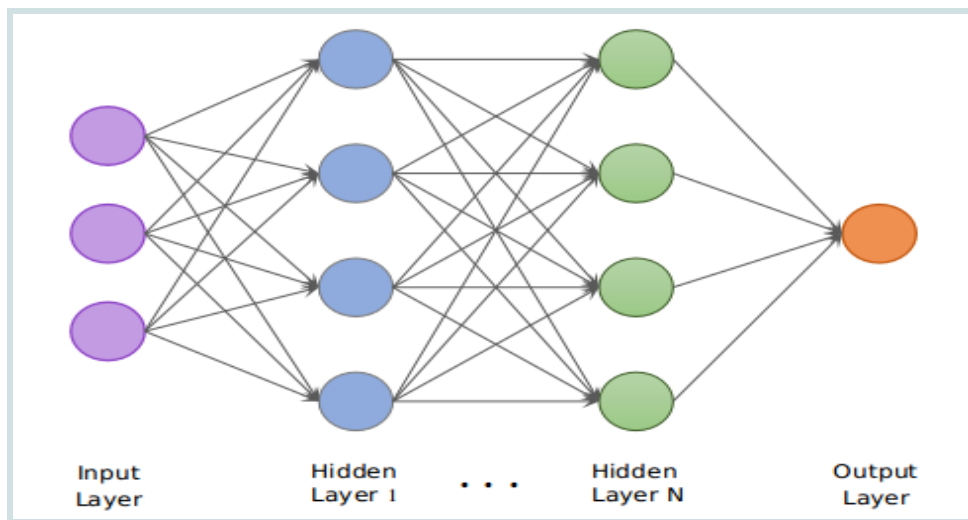


FIGURE 7.

This figure demonstrates the training process at a block level. The input to the network is a 2D image. The network has an input layer which takes the image as the input, output layer from where we get the trained output and the intermediate layers called the hidden layers. As stated earlier, the network has a series of convolutional and sub-sampling layers.

Algorithm Used-Polygon feature extraction: The input raster can have any cell size and must be a valid integer raster dataset. The Field parameter allows you to choose which attribute field of the input raster dataset will become an attribute in the output feature class. If a field is not specified, the cell values of the input raster (the VALUE field) will become a column with the heading Grid_code in the attribute table of the output feature class. The Table-1 represents the Raster to Polygon conversion.

8.APPLICATIONS BIG DATA IN REAL-TIME

Nowadays, satellite images have become an indispensable segment in many applications such as medical imaging, remote sensing, crime prevention, education, multimedia, and data mining...). These applications require satellite images as a source for various processes like segmentation, interest point, edge detection, object recognition, tracking, and

others. Rasters are digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps. Data stored in a **raster format** represents real-world phenomena: Thematic data (also known as discrete) represents features such as land-use or soils data.

The four main types of big data technologies are Data Storage, Data Mining, Data Analysis, and Data Visualizations.

Data storage-It contains databases like Apache, Hadoop, SQL and many more. In our project SQL is used to store data like image’s Locations, Dimensions, Attributes, Variations, Images, diversions and Structures.

Data Mining-Objectives of Mining like Rapidminer and Presto are applicable to perform operations like Extracting information’s like Patterns and Sketches from Raw Data, after which the Structured and Unstructured data into Useful Data.

Data Analysis-The Process of Analyzing Data is performed before the preferred algorithm is applied to the data converted. This analysis works well in cleaning the data by removing unwanted blur pixels into clear and useful data which would be then worth Converting.

Data Visualization-This process of big data application creates Stunning Visualization of the information converted from raw data. This helps in the creation of Graphs and Charts for the fluctuations and the rate of data analyzed after Mining and Analysis.

TABLE 1. Raster to Polygon conversion (in raster, out polygon features, {simplify}, {raster field})

Parameter	Explanation	Data type
in_raster	The input raster dataset. The raster must be integer type.	Raster Layer
out_polygon_features	The output feature class that will contain the converted polygons.	Feature Class
simplify (Optional)	Determines if the output polygons will be smoothed into simpler shapes or conform to the input raster's cell edges. <ul style="list-style-type: none"> • SIMPLIFY —The polygons will be smoothed into simpler shapes. This is the default. • NO_SIMPLIFY —The polylines will conform to the input raster's cell edges. 	Boolean
raster_field (Optional)	The field used to assign values from the cells in the input raster to the polygons in the output dataset. It can be an integer or a string field.	Field

TABLE 2. Extract By Polygon (in raster, polygon, {extraction area})

Parameter	Explanation	Data Type
in_raster	The input raster from which cells will be extracted.	Raster Layer
polygon [point,...]	A polygon (or polygons) that defines the area of the input raster to be extracted. Each polygon part is a list of vertices defined by Point classes. The points are specified as x,y coordinate pairs in the same map units as the input raster. The form of the object for a single polygon is as follows: • [point(x1,y1), point(x2,y2), ..., point(xn,yn), point(x1,y1] Optionally, a group of multiple polygons can be specified by	Point
	using a Polygon class to define a list of polygon parts. Note that in order to do this, all parts must be contiguous, and thus be encompassed by one single perimeter shape. The form of the object in this case would be in a contiguous list as follows: <ul style="list-style-type: none"> •[point(x1,y1), point(x2,y2), ..., point(xn,yn), point(x1,y1), point(x'1,y'1), point(x'2,y'2), ..., point(x'n,y'n), point(x'1,y'1), ...] In all cases, the last coordinate for each polygon part should be the same as the first in order to close it.	
extraction_area (Optional)	Identifies whether to extract cells inside or outside the input polygon. <ul style="list-style-type: none"> • INSIDE —A keyword specifying that only the cells inside the input polygon should be selected and written to the output raster. All cells outside the polygon will receive NoData values on the output raster. • OUTSIDE —A keyword specifying that the cells outside the input polygon should be selected and written to the output raster. All cells inside the polygon will receive NoData. 	String

9. CONCLUSION

Due to its powerful learning ability and advantages in dealing with occlusion, scale transformation and background switches, deep learning based object detection has been a research hotspot in recent years. This paper provides a detailed review on deep learning based object detection frameworks which handle different sub-problems, such as occlusion, clutter and low resolution, with different degrees of modifications on CNN. Initially satellite image is preprocessed and contrast of particular image has been enhanced for better object detection. Using polygon feature extraction, the shape of objects in an image has been detected accurately and classified through CNN. Hence our system detects objects like trees, buildings, roads, vehicles and few other objects in an accurate way and results in a bounded box. Compared to other existing methods of object detection in satellite images our method achieved better results.

REFERENCES

- [1]. Pengcheng Wei, Zhen Zhou, (January 2018) "Research on face feature extraction based on K-mean algorithm" EURASIP Journal on Image and Video Processing.
- [2]. Y. Xiao, S.K. Lim ; T.S. Tan (2014), "Feature extraction using very high resolution satellite imagery" IGARSS 2004. 2014 International Geoscience and Remote Sensing Symposium.
- [3]. Zhao, Hongzan Jiao (2017), "Satenn: satellite image dataset classification using agile convolutional neural networks," Remote Sensing Letters, vol. 8, no. 2, pp. 136–145.
- [4]. Zhong, Z., Li, J., Cui, W., & Jiang, H. (2016). "Fully convolutional networks for building and road extraction". In Geoscience and Remote Sensing Symposium (IGARSS), (pp. 1591-1594).IEEE.
- [5]. Kazuki Uehara and Hidenori Sakanashi, (2019) "Object Detection in Satellite Images Based on Active Learning Utilizing Visual Explanation" 11th International Symposium on Image and Signal Processing and Analysis (ISPA).