# Deep Learning Approaches for Water-body Detection from Satellite Imagery

**N.Surve, Suchitra Patil, Abhijeet Pasi, Nishant More**

***Abstract— Agriculture department of Government of India provides funds for digging wells and making farm ponds for improving farmers livelihood. The proper execution of the policies not been worked out due to the manual process. For the automation of the process, we are using the object detection algorithm Faster Region-based Convolutional Neural Network (F-RCNN) to detect the location of the wells and Farm ponds. The faster RCNN method introduced region of interest to improve the model speed. We have tried to address water body detection on satellite imagery with an AI approach to cross verify the utilization of funds is the challenge. The traditional object detection algorithms are low in accuracy, also uses coco image data set. Detecting water body has not been addressed by anybody. We have extended the algorithms for water bodies detection for the complete geographic region of agricultural land. Improved Faster R-CNN algorithm used in the detection of wells and farm ponds. More than 1000 Satellite imagery are used to train the model to detect the water body. The experimental results demonstrated promising accuracy on well and farm ponds detection. The accuracy of the faster RCNN algorithm is 90%, we have tested the accuracy of the results by GIS locations, and ground-truthing is achieved by our designed Android App.This app gives the latitude and longitude of well or farm ponds entered by the person in Jalgaon.***

***Keywords** : About four key words or phrases in alphabetical order, separated by commas.*

## I. INTRODUCTION

The major challenge is collecting the Well and farm pond data from the satellite view of google maps. The satellite view of Well data is currently not present. Thus manually collecting the latitude, longitude for the positive and negative image was the major challenge. For training we need to annotate the wells present in an image using some tool, thus the more number of images more the number of annotations will be.

Once we are done with collecting the data and training, then one crucial task which is left is finding the results. Properly selecting the dimensions of the image for training is also a major challenge. When downloading the images using google maps static API the major problem was that we were able to download only approx. 500 images after that it showed an error that means in a day were able to download a limited amount of data.

In Maharashtra state, a large number of places encounter a deficiency of water very frequently. Agricultural and industrial activities are severely affected. But even more importantly, there are places where even the drinking water demand is not met resulting in distress, mass migration and deaths. Generally, the stress period starts in November and goes on increasing till June.

The geology of most part of Maharashtra is not favorable to underground aquifers. Aquifers in most of the region don't have enough water to meet the water demand throughout the year. Conditions are even worse for the villages where these local aquifers are not present. These villages then have to resort to the tanker from nearby villages. Sometimes these tankers are brought from places as far as 50 km. With the beginning of summer, most of the surface water sources go dry which makes the conditions very harsh.

In the regions with high rainfall, scarcity is mainly attributed to the geology of the region, which affects it in two aspects: most of the area lies in Deccan Basalt rock belt which makes it unfavorable for effective recharging of aquifers. The second one is the terrain. Slopes of terrain make it further difficult for percolation. The major part of the rainwater runs off because of the slope.

The local authorities are faced with managing scarce water resources in a drought situation. They have to allocate water based on available information. Typically the information is dated, incomplete and in a form that requires a lot of judgment, if not guess.

Due to the availability of rainwater at the upper side of farmland, it is said to be the pale grasslands digging up to be available for the crop at an emergency. These ponds are taken in the pavement area on the edge of the tunnel.

Planting in the field is to store overflow water and use it for safe irrigation is the main objective of this government program. When there is stress due to lack of rainfall due to rain irregularities in this palace, if there is any other crop of water released from the water stored in this fodder, then there is a chance. Whereas it is not possible to dig easily, water harvesting and water harvesting programs are organized.

Benefits of farm ponds and well digging are Water in the catchment area is filled with water. Water can be available to provide the crop with water for emergencies. Due to supplemental irrigation, there is a significant increase in crop production. Farming is good for improving the fluctuating and submerging land. It is used for fisheries culture.

The land which has land, under which land is surrendered is selected for building farm ponds or wells. The land where clay is high is suitable for the farm ponds. Also, in the Western Ghats division, it will be beneficial to cultivate land inland under the upper part of the group for rice cultivation. Government of Maharashtra provides funds for digging farm ponds and wells. For the proper execution of the program, we need to check the actual location of farm ponds.

Manually doing a survey is very time consuming and difficult task.

Given the availability of Deep learning approaches and GIS technology, we can easily verify the work done. It will reduce the time and effort needed in manual processing and at 90% accuracy.

## II. LITERATURE SURVEY

Humans can easily discriminate different object of image and pay attention to the key parts. By contrast, the computer conducts it hardly because images are considered as one grayscale or multiple dimensions(RGB) numerical matrix, the computer needs to recognise the meaning of different numbers. The traditional algorithm is based on computer vision theory such as HOG[1], DPM[2] and so on. With the popularity of deep learning, the method of object detection based on it has improved greatly in the accuracy and speed of detection.

Brilliant object detection strategies influenced by deep learning are originated from RCNN[3]. The faster RCNN[4] technique acquainted area of enthusiasm with improve the model speed. Faster RCNN[5] propelled this stream by learning the attention mechanism with RPN (Region Proposal Network).

As of late, since a crucial achievement is come to by[6], different vision assignments in remote detecting field have accomplished critical advancement with the assistance of deep convolutional neural systems (CNNs) [7]. There are likewise extraordinary advances [4], [5] in object detection from common pictures. Exceptional accomplishments of CNN based article discovery have roused us to structure a profound system engineering for object detection in remote sensed satellite images.

Object detection has been a fundamental and vital topic in the research field of remote sensing images interpretation over the last decade, both in synthetic aperture radar imagery and in optical imagery. Recently, the state-of-the-art object detection for natural images based on the advanced deep neural network have already achieved promising performances [5] [8] [9] [10].
Many research papers have proposed ways of using deep networks for locating class-specific or class-agnostic bounding boxes[11] [12][17]. Selecting suitable size of bounding boxes is also a challenge. In the OverFeat method[12], a fully- connected neural network layer is trained to predict the box coordinates for the localization task that assumes a single object inside a bounding box. The fully connected layer is then turned into a convolution layer for detecting multiple class-specific objects. The Multi-Box methods generate region proposals from a network whose last fully connected layer simultaneously predicts multiple boxes, which are used for R-CNN object detection. Their proposal network is applied to a single image or multiple large image crops (e.g., 224224). Shared computation of convolutions [12] has been attracting increasing attention for efficient, yet accurate, visual recognition. The OverFeat paper[12] computes convolution features from an image pyramid for classification, localization, and detection. Adaptively-sized pooling (SPP) on shared convolution feature maps is proposed for efficient region-based object detection and semantic segmentation. Fast R-CNN enables end-to-end detector training on shared convolution features and shows compelling accuracy and speed.
R-CNN is slow because it performs a Convolution network pass for each object proposal, without sharing computation. Spatial pyramid pooling networks (SPPnets)[13] were proposed to speed up R-CNN by sharing computation. The SPP net method computes a convolutional feature map for the entire input image and then classifies each object proposal using a feature vector extracted from the shared feature map. Features are extracted for a proposal by max-pooling the portion of the feature map inside the proposal into a fixed-size output (e.g., 6 6). Multiple output sizes are pooled and then concatenated as in spatial pyramid pooling[14]. SPPnet

| Name of Algorithm | Characteristics | Drawbacks |
|---|---|---|
| CNN | Applies divide and conquer strategy to classify images | Computational expensive as large number of images required |
| R-CNN | An optimal search criterion is required to create regions. Approximately 2000 regions are extracted from each image | Time consuming as each region is processed by CNN separately. Three models are used for predictions |
| Fast R-CNN | Every image is processed at the max once to the CNN and feature maps are generated. Vigilant search is used on these maps to generate predictions. | Vigilant search is slow and hence computation time is still high. |
| Faster R-CNN | Instead selective search method uses Region proposal network (RPN) method which makes the algorithm much faster. | Region proposal network method takes time. There are systems working sequentially so the performance of system depends on how previous stage system has performed. |
| YOLO (You Only Look Once) | it's incredibly fast and can process 55 frames per second | Speed advantage of YOLO comes at a price for accuracy reduction. R-CNN and its variations are usually more accurate. |

accelerates R-CNN by 10 to 100x at test time. Training time is also reduced by 3x due to faster proposal feature

Following table gives information about different algorithms available for object detection in deep neural network. Mainly CNN, R-CNN, Fast RCNN, Faster RCNN and YOLO algorithms are compared here. The algorithms are compared based on characteristics and certain short comings of those algorithms.

**Table 1: Comparison of different algorithms used for object classification**

## III. PROPOSED SYSTEM

The major challenge in this research is to collect image dataset for training as Well and farm ponds data is not available. To solve this problem we have used Google static map API to download the image set for positive and negative image set. By generating a key for Google static map API we can download some images from a satellite view of Google maps. As the image dimension plays an important role we have downloaded images with the 112 x 112 pixels image size by keeping the zoom size as maximum i.e. 19. We have collected 1200 Positive images and 2000 negative images from the Google static map API.

Once the dataset is ready our next step is to train the model. In Faster - RCNN we use the bounding boxes for the feature extraction and prediction. On training dataset, we need to annotate the wells and farm ponds so that it can be used for the training dataset and used for generation of weight file used for prediction of wells. We are using the BBox labelling tool for generating bounding boxes and finding the coordinates value of x, y, h and w. Once we have the coordinates value for x, y, h and w, we need to convert it into the Faster - RCNN format. Location of the image, the value of x, y, h, w and then the name of the class. This text file is then used for the training process. A Faster R-CNN network takes as input an entire image and a set of object proposals. The network first processes the whole image with several convolutional and max-pooling layers to produce a convolutional feature map. Then, for each object proposal, a region of interest (RoI) pooling layer extracts a fixed-length feature vector from the feature map. Each feature vector is fed into a sequence of fully connected (FC) layers that finally branch into two sibling output layers: one that produces softmax probability estimates over K object classes plus a catch-all background class and another layer that outputs four real-valued numbers for each of the K object classes. Each set of 4 values encodes refined bounding-box positions for one of the K classes.

### A. Algorithm for faster RCNN for detection of Well and farm ponds from satellite image

We have summarized below the steps followed by a Faster R-CNN algorithm to detect wells and farm ponds from satellite image:

**Step 1:** Pass an input image to the Convolution Network which will in turn return feature maps for the image.

**Step 2:** Use Region Proposal Network (RPN) on output obtained from previous step, the feature maps and get object proposals.

**Step 3:** Use ROI pooling layer to change all the proposals to the equal size.

**Step 4:** Finally, pass these proposals to a fully connected layer for classification and predict the bounding boxes for the image.

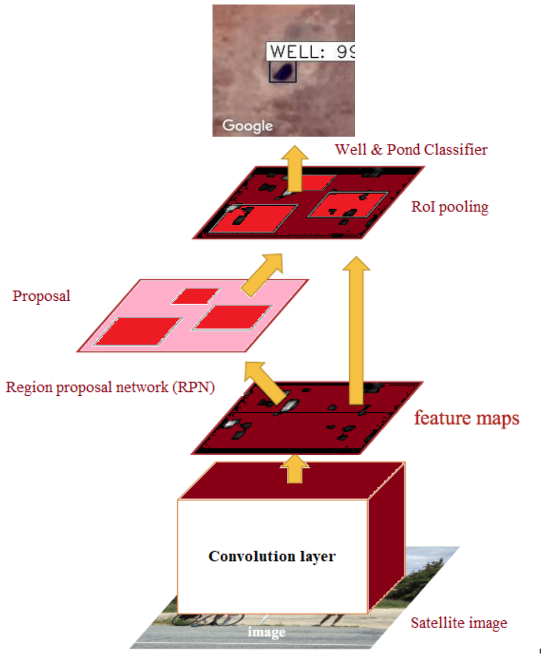


**Fig. 1. Proposed System**

The bounding boxes should not be very large. Otherwise two objects in one frame cannot get detected by the network.

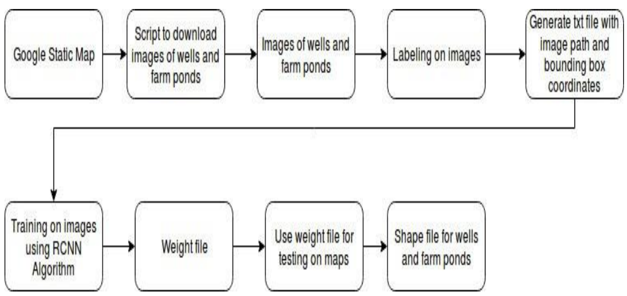


**Fig. 2. Diagram for Faster RCNN**

The RoI pooling layer uses max pooling to convert the features inside any valid region of interest into a small feature map with a fixed spatial extent of H  W (e.g., 7  7), where H and W are layer hyper-parameters that are independent of any particular RoI. In this paper, an RoI is a rectangular window into a convolutional feature map. Each RoI is defined by a four-tuple (x,y, h, w) that specifies its top-left corner (x,y) and its height and width (h, w).

RoI max-pooling works by dividing the h w RoI window into an H W grid of sub-windows of approximate size h/H w/W and then max-pooling the values in each sub-window into the corresponding output grid cell. Pooling is applied independently to each feature map channel, as in standard max pooling. The RoI layer is simply the special case of the spatial pyramid pooling layer used in SPPnets in which there is only one pyramid level. We use the pooling sub-window calculation given in.

Once a Fast R-CNN network is fine-tuned, detection amounts to little more than running a forward pass (assuming object proposals are pre-computed). The weight file is generated as 'model-frcnn.hdf5'. The network takes as input an image (or an image pyramid, encoded as a list of images) and a list of objects proposals to score.

Once we detect the wells and farm ponds on images we need to find the location of the wells and farm ponds detected. We are using haversine formula to find the distance between two latitudes and longitude, and also the location of the detected well and farm ponds based on the centre point and the distance. Finally, we are generating the shapefile of the detected wells and farm ponds. This shapefile has the points as the attributes having the location of wells and farm ponds.

## IV. DATASET

We have created our own image data-set instead of using existing data-sets. For training and testing, we are dividing the data-set into two parts. The 70% images are used for training and 30% for testing. We have downloaded this image data-set from Google static map API. For training, we are using all the images with well and farm ponds (positive images) as we want to find bounding box coordinates for each image and we are using positive and negative images(images without well and farm ponds) for the testing purposes.

## V. RESULTS AND DISCUSSIONS

Once the training on image dataset is completed we can use the weight file i.e 'model-frcnn.hdf5' file for the testing on other images. As an output, it will generate the image with the bounding box to the detected object with its class name as WELL and the score of detection. In following figure well is detected with the bounding box, generated with the score as 99.

Fig. 3. Well 1


Fig. 4. Well 2

.

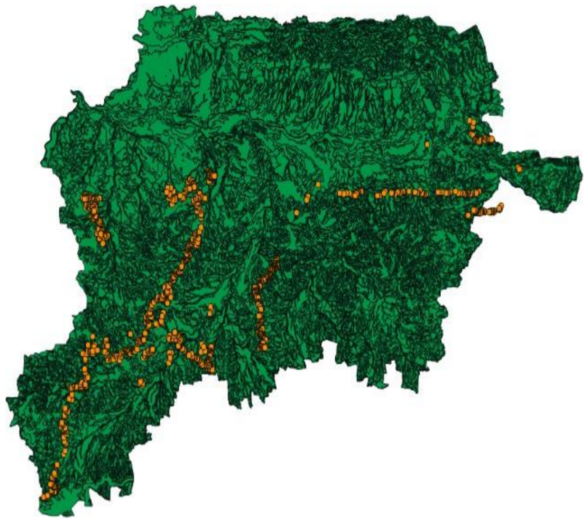

Fig. 5. Sample Object - Well- detected for Jalgaon District Map:

We have found the location in latitude and longitude format Wells and farm ponds are marked with yellow color and placing them in the GIS layer so that we can have the

data for wells and farm ponds. Figure 6 above shows the sample wells detected for the Jalgaon district from Maharashtra, India. The layer is created in QGIS for Jalgaon District and for the detected wells. Green color is the Soil data for Jalgaon district and yellow dots on it are the wells or Farm pond detected from the satellite image using faster RCNN.

For sample we are using 1000 positive images for raining and 470 positive and negative images for testing.

**Table 2: Confusion matrix for well detection**

| | | Actual Class | |
|---|---|---|---|
| | | ***Water bodies*** | ***Non water bodies*** |
| **Predicted class** | ***Water bodies*** | 300 | 20 |
| | ***Non water bodies*** | 30 | 123 |

We have used true positive 300 images, true negative 120 images, false positive selected images are 20 and false negative images are 30.

**TP**= 300 / 330 = 0.90

**FN** = 30 / 330 = 0.10

**FP** = 20 / 140 = 0.14

**TN** = 123 / 140 = 0.86

$$\text{Accuracy} = \frac{(\text{TP+TN})}{(\text{Number of images for testing})}$$

**Accuracy** = (300 + 123) / (470) = 0.9

So we find the accuracy as 90% for the above sample testing.

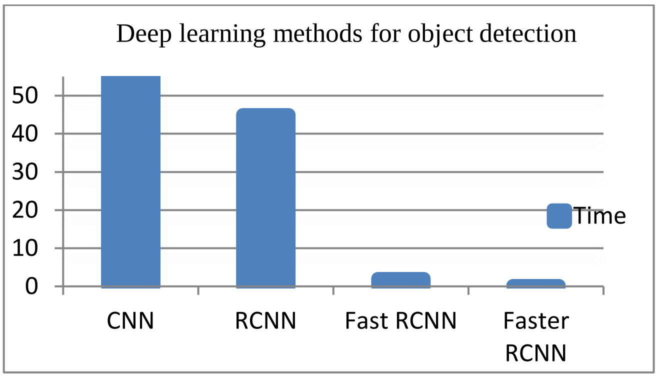


Fig 6 Chart for comparing performance of various object detection techniques

## VI. CONCLUSION

The manual process of collecting information about wells and farm ponds will take much time and efforts of the individual. By making it automated it will reduce time and efforts along with that the accuracy of collected data will increase. By using the Faster RCNN algorithm on Google satellite imagery the object detection will reach on another level as it is not only applicable to real-time image-set like COCO but also on satellite imagery. We have tried to address water body detection on satellite imagery with an AI approach to cross verify the utilization of funds is the challenge. As a result, we can check the result as the core is getting 99 for wells. The experimental results demonstrated accuracy on well and farm ponds detection. The accuracy of the algorithm is 90%.

## REFERENCES

[1]Dalal, Navneet, and Bill Triggs. ”Histograms of oriented gradients for human detection.” international Conference on computer vision & Pattern Recognition (CVPR’05). Vol. 1. IEEE Computer Society, 2005.

[2]Felzenszwalb, Pedro F., David A. McAllester, and Deva Ramanan. ”A discriminatively trained, multiscale, deformable part model.” Cvpr. Vol. 2. No. 6. 2008.

[3]Girshick, Ross, et al. ”Rich feature hierarchies for accurate object detection and semantic segmentation.” Proceedings of the IEEE con- ference on computer vision and pattern recognition. 2014.

[4]Girshick, Ross. ”Fast r-cnn.” Proceedings of the IEEE international conference on computer vision. 2015.

[5]Ren, Shaoqing, et al. ”Faster r-cnn: Towards real-time object detection with region proposal networks.” Advances in neural information processing systems. 2015.

[6] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. ”Imagenet classification with deep convolutional neural networks.” Advances in neural information processing systems. 2012.

[7] Zhu, Xiao Xiang, et al. ”Deep learning in remote sensing: A compre- hensive review and list of resources.” IEEE Geoscience and Remote Sensing Magazine 5.4 (2017): 8-36.

[8] Lichao Huang, Yi Yang, Yafeng Deng, and Yinan Yu, Densebox: Unifying landmark localization with end to end object detection, Computer Science, 2015.

[9]Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander CBerg, Ssd: Single shot multibox detector, in European conference on computer vision. Springer, 2016, pp. 2137.

[10] Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, and Jiajun Liang, East: “An efficient and accurate scene text detector”, arXiv preprint arXiv:1704.03155, 2017.

[11] C. Szegedy, A. Toshev, and D. Erhan. “Deep neural networks for object detection”, In NIPS, 2013.

[12] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. “Overfeat: Integrated recognition, localization and detection using convolutional networks”. In ICLR, 2014.

[13] K. He, X. Zhang, S. Ren, and J. Sun. ”Spatial pyramid pooling in deep convolutional networks for visual recognition”. In ECCV, 2014.

[14] S. Lazebnik, C. Schmid, and J. Ponce. “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories”. In CVPR, 2006.

[15] Li, Qingpeng, et al. "Hierarchical region based convolution neural network for multiscale object detection in remote sensing images." IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018.

[16] Tang Pei, Wu Xiaoyu. "Object proposals detection", 2016 2nd IEEE International Conference on Computer and Communications (ICCC), 2016

[17] Mingjie Li, Weiwei Guo, Zenghui Zhang,Wenxian Yu, Tao Zhang. "Rotated Region Based Fully Convolutional Network for Ship Detection", IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018