DOI: http://dx.doi.org/10.26483/ijarcs.v10i5.6466

Volume 10, No. 5, September-October 2019

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

CASE STUDY

Available Online at www.ijarcs.info

AUTOMATIC DETECTION OF OVERHEAD WATER TANKS FROM SATELLITE IMAGES USING FASTER-RCNN

Ishika Saini
Government Engineering College
Ajmer, Rajasthan, India
ms.ishika1@gmail.com

Dishant Parikh
G.H Patel College of Engineering and Technology
Anand,Gujarat,India
dishant30899@gmail.com

Pranjal Sharma
Government Engineering College
Ajmer, Rajasthan, India
spranjal13@gmail.com

Shweta Khandelwal
Sangam University
Bhilwara, Rajasthan, India
shweta7226@gmail.com

Giribabu Dandabathula
Regional Remote Sensing Centre- West
NRSC/ISRO,
Jodhpur, Rajasthan, India
dgb.isro@gmail.com

Sitiraju Srinivasa Rao
Regional Remote Sensing Centre- West
NRSC/ISRO,
Jodhpur, Rajasthan, India
sitiraju@gmail.com

Abstract: Pattern recognition is pertinent field for detection of urban/man-made features from satellite imagery. Neural networks are best used in object detection for recognising patterns in imageries. Convolutional Neural Networks (CNNs) become way in solving object detection task based on deep learning concepts. This article demonstrates the usability of CNNs for detecting and mapping of small objects from the urban scenes. Identification and mapping of overhead water tanks from satellite imagery is a very important task especially during reconnaissance situation raised due to water contamination. Faster Region based CNN (Faster RCNN) has been used to detect and map the overhead water tanks in the urban scene from satellite imagery. The results from this study indicate that Faster RCNN gives affirmative accuracy towards detection of small objects from satellite imageries.

Keywords: Convolution Neural Network; Transfer learning,; Regional Proposal Network; Small objects in Urban Scene

I. INTRODUCTION

Artificial Neural Networks (ANNs) have proved its efficiency in the problems related to the classification of objects by applying a learning rule [1]. This learning rule is an algorithm which modifies the parameters like weights and threshold of the variables within the network [2]. A network consists of a backbone network and an object detector. A backbone network or base network is a feature extractor from which the detector extracts its discriminative power. Several feature extractors like MobileNet, VGG-16, and Inception will tend to learn the features of the object in the input image based on the size of the object [3-5]. The choice of feature extractor is crucial as the number of parameters and types of layers directly affects memory, speed and performance of the detector [6].

As the size of the network increases, the training becomes slower and requires more and more data. Recent advances have witnessed increased efficiency in using Convolution Neural Network (CNN). Girshick proposed a Fast Region-based CNN (Fast R-CNN) for object detection [7]. This framework portrays two main approaches, first of applying CNN to ground up the region proposals in order to localize objects and secondly, adopting transfer learning when label data is less. In transfer learning, a parent network is trained on a base dataset and a base task, the learned features from this network is then transferred to a second target network to be trained on a target dataset and a task.

Pattern recognition is pertinent field for detection of urban/man-made features from satellite imagery. Identification and mapping of overhead water tanks from satellite imagery is a very important task especially during reconnaissance situation raised due to water contamination. A need has risen to detect overhead water tanks from satellite data for a public health monitoring project. In this article we produce the results on application of Region Proposal Network (RPN) for detecting overhead water tanks from various satellite data captured on various cities and towns. RPNs have helped in predicting the object bounds and objectness score, consequently RPNs are trained end-to-end to generate high-quality region proposals, which are used by Fast R-CNN for detecting the overhead water tanks from a given scene.

II. LITERATURE REVIEW

Shin et al. have employed deep CNN to computer aided detection from medical images to detect thoraco-abdonminal lymph node and interstitial lung disease [8]. Erhan et al. proposed a method to localize objects in an image, which predicts multiple bounding boxes at a time by using the concept of DeepMultiBox [9]. Girshick introduced the Fast-RCNN, an update to RCNN and SPPnet [7]. Ren et al. proposed a unified, deep learning based object detection system based on Faster R-CNN and RPN [10]. Lee et al. has shown the trade-offs between different backbone networks in the terms of speed and accuracy with Faster R-CNN as the classifier [11]. Wang and

Zang applied Faster RCNN on a different number of datasets for accuracy in detecting building areas [12]. Huang et al. performed an experimental comparison between different object detectors with respect to the factors affecting the speed and accuracy [13]. Szegedy et al. proposed a method for significant quality gain with minimum computational cost compared to shallow networks and used their inception architecture for small object detection [14,15].

III. METHODOLOGY

In our work, we have used inception-v2 as the feature extractor and Faster R-CNN for detection purpose.

A. Inception Network as a Feature Extractor

In networks with repetitive max pooling layers, the chances of loss of accurate spatial information is much higher with respect to the scale of the image and size of the target object in the image. Therefore a network has been created with three different size of filters (1*1, 3*3, 5*5) to perform the convolution followed by max pooling at a single layer. The collective output is then transferred to next inception module.

B. Inception v1

To reduce the computation cost, an extra 1*1 convolution is added before the 3* and 5*5 convolutions. This 1*1 convolution is added after the max pooling layer.

C. Inception v2

In this module, the 5*5 convolution has been factorized into two 3*3 convolutions to improve computational speed. Further the factorization of n x n filters is done in the combination of 1 x n and n x 1 to boost up the performance. Figure 1 shows the architectural view of Inception v2.

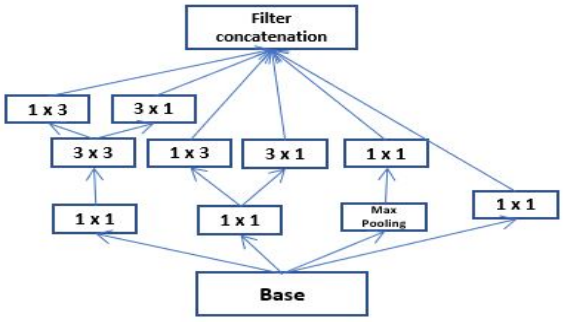

Figure 1. Archictural view of Inception v2

D. Faster RCNN

The state of the art model, Faster RCNN is a combination of two networks. RPN is used to generation of regional proposals and the second network which used these proposals for detecting objects. The RPN ranks the region boxes and nominates the ones with most likely containing objects.

E. RPN

CNNs proved to be important object detection classifiers which take an input, assign learnable weights and bias to various aspects in the image and differentiate features from one another. It constitutes two parts, the feature learning part comprises of convolution layers, activation function and pooling layers. The classification part is a fully connected layer. The last stage of convolution generates a feature map.

To generate region proposals, a small network is transferred over the convolution feature map output. This small network takes input as an n * n spatial window of the input convolution feature map. These feature maps are then passed to two convolution layers in which one layer is for classification and another one is for regression. Each pixel in the feature map generates region candidate boxes which are then fed to classification layer (cls) and regression layer (reg) to get proposals.

F. Object Detection

Now that we have the proposals, the next step is to get the labels and the position of each proposal. The detection network contains two fully connected layers and two dropout layers. The two output layers gives N+1 as one output (N object classes and 1 background) and second output as N * 4 bounding box regress for each candidate box.

CNNs proved to be important object detection classifiers which take an input, assign learnable weights and bias to various aspects in the image and differentiate features from one another. It constitutes two parts, the feature learning part comprises of convolution layers, activation function and pooling layers. The classification part is a fully connected layer. The last stage of convolution generates a feature map.

Figure 2 shows the methodology of detecting overhead water tank from satellite images using Faster RCNN process.

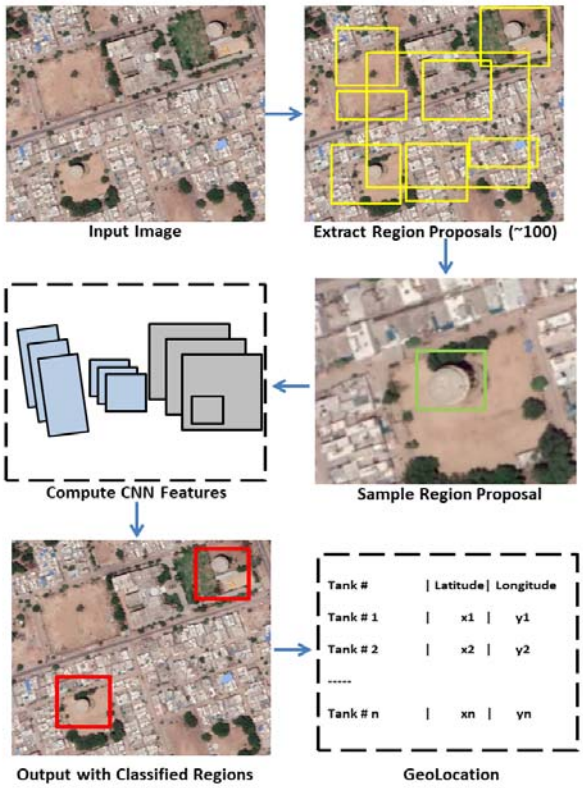

Figure 2. Overhead water tank detection from Satellite image using Faster RCNN process.

G. Establishment of Image Dataset

In this experiment, our dataset is a collection of 175 satellite images (spatial resolution of 0.5 m) containing overhead water tanks for various citites and towns in India. Certain images without overhead water tank are also part of the collection.

Each image is then annotated and the bounding boxes are made such that each box includes the top circular pattern of the overhead water tank with the shadow as an extra feature to differentiate the circular tank with other circular patterns in the background. The coordinates of the bounding boxes are saved in a separate XML file for each image. These XML files are then converted into a combined CSV file which is further converted into a *tfrecord*. *tfrecord* file format is Tensorflow's own binary storage format which has significant impact on the performance and the training time of our model. Moreover, binary data takes less space on disk, takes less time to copy, and can be read much more efficiently from the disk.

H. Training

Training of RPN is done by assigning a binary class label (of being an object or not) to each proposals. A positive label is assign to two kinds of proposals:

1. the proposal with the highest Intersection-over-Union (IoU) overlaps with the ground-truth box or
2. a proposal that has an IoU overlap higher than 0.7 with any ground-truth box

Proposals that are neither positive nor negative do not contribute in training. With these definitions, our loss function for an image is defined as

$$L(\{p_i\},\{t_i\}) = \frac{1}{Ncls}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{Nreg}\sum_i L_{reg} p_i^* (t_i, t_i^*)$$

Where i is the index of the proposal and pi is the probability of a proposal belong to an object, pi* $\epsilon\{0,1\}$ is ground-truth label, ti* is the ground-truth of the box's position containing values(the coordinates of upper left corner, width, height of bounding box), ti = {tx,ty,tw,th} is the predicted bounding box.

Tx = (x – xa)/wa , ty = (y – ya)/ha
Tw= log(w/wa) , th = log(h/ha)
Lcls(pi, pi*) are two categories i.e, target and non-target logarithmic loss,
Lcls(pi, pi*) = -log[pi*pi + (1- pi*)(1-pi)]
Lreg(ti, ti*) is regression loss,
Lreg(ti, ti*) = (ti – ti*)

Where R is smooth L1 function.

λ=10, is parameter for normalization.

Here, we will be using stochastic gradient decent (SGD) optimizer. Through transfer learning, a pre trained model (inception v2) trained on coco dataset is used for classification to initialize our base network. In this paper, we will be using several consecutive steps for training our model. In the first step, the input images are fed to several convolution and pooling layers in each iteration to extract feature map. These feature maps are then passed through the RPN to obtain the proposals. The proposals are then forwarded to the Faster-RCNN detection network.

IV. Results and Discussions

Considering Faster RCNN for transfer learning, models are trained on images that are scaled up to M pixels on the shorter edge whereas in SSD, images are always resized to a fixed shape M x M. Setting up the Tensorflow as backend, the target model is trained using the generalized features of parent network Faster RCNN Inception-v2 keeping stride size as 16, IoU threshold as 0.69, batch size as 1 and initial learning rate as 0.00019. Softmax function is used to normalize the input vector into a probability distribution with k probabilities. The mean average precision (mAP) value of Faster RCNN with our tank dataset was noted as 29. The checkpoint with the average loss of about 0.7 is saved and the corresponding inference graph is used for testing purpose. The i and j coordinates of the centroid of each predicted bounding box is calculated:

Ymin= boxes[0][i][0]*H
Xmin= boxes[0][i][1]*W
Ymax= boxes[0][i][2]*H
Xmax= boxes[0][i][3]*W
i = (Ymin+Ymax)/2
j = (Xmin+Xmax)/2

Where, Xmin, Ymin are the coordinates of top left of bounding box and Xmax, Ymax are the coordinates of bottom right. H and W are the height and the width of the image and i, j are the coordinates of the centroid of the predicted bounding box.

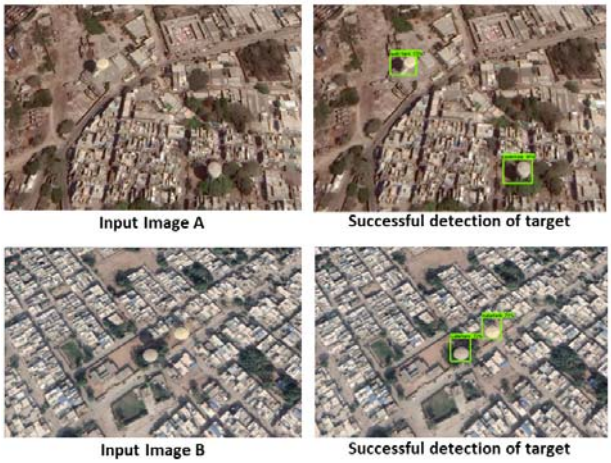

Figure 3. Figures showing successful identificaiton of the targets

V. Conclusion

Faster RCNN is proven effective in extracting the small targets even with varied background. We evaluated the previously studied Faster RCNN model on dataset containing satellite images for detecting over-head water tanks. The circular pattern of the tank with the on-ground tank shadow is considered as the target object in the input image. Through transfer learning, the features of the Faster RCNN Inception-V2 model have been taken as the generalized features for the target model. Further accuracy can be improved in future by adding more images in dataset with various patterns and features of the over-head water tanks.

VI. References

[1] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160, 3-24.

[2] Lippmann, R. P. (1987). Anintroduction to computing with neural nets. IEEE Assp magazine, 4(2), 4-22.

[3] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference

on Computer Vision and Pattern Recognition (pp. 4510-4520).

[4] Yu, W., Yang, K., Bai, Y., Xiao, T., Yao, H., & Rui, Y. (2016, June). Visualizing and comparing AlexNet and VGG using deconvolutional layers. In Proceedings of the 33 rd International Conference on Machine Learning.

[5] Nguyen, L. D., Lin, D., Lin, Z., & Cao, J. (2018, May). Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation. In 2018 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 1-5). IEEE.

[6] Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., ... & Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7310-7311).

[7] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

[8] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging, 35(5), 1285-1298.

[9] Erhan, D., Szegedy, C., Toshev, A., & Anguelov, D. (2014). Scalable object detection using deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2147-2154).

[10] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[11] Lee, C., Kim, H. J., & Oh, K. W. (2016, October). Comparison of faster R-CNN models for object detection. In 2016 16th International Conference on Control, Automation and Systems (ICCAS) (pp. 107-110). IEEE.

[12] Wang, X., & Zhang, Q. (2018, August). The Building Area Recognition in Image Based on Faster-RCNN. In 2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC) (pp. 676-680). IEEE.

[13] Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., ... & Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7310-7311).

[14] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[15] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[16] Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights into imaging, 9(4), 611-629.