International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IJIEMR Transactions, online available on 28th Jul 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 06

10.48047/IJIEMR/V12/ISSUE 06/09

Title **ABNet: Adaptive Balanced Network for Multiscale Object Detection in Remote Sensing Imagery**

Volume 12, ISSUE 06, Pages: 52-59

Paper Authors **Joshi Padma N,R. Dinesh, K. Lakshmi Siva Kanth Reddy, U. Mahesh,M. Keertana Rao**

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

ABNet: Adaptive Balanced Network for Multiscale Object Detection in Remote Sensing Imagery

1. Joshi Padma N, Associate Professor, Department of CSE, Sreyas Institute of Engineering and Technology, Telangana, India, padmajoshi@sreyas.ac.in
2. R. Dinesh, Department of CSE, Sreyas Institute of Engineering and Technology, Telangana, India, ramagiridinesh2001@gmail.com
3. K. Lakshmi Siva Kanth Reddy, Department of CSE, Sreyas Institute of Engineering and Technology, Telangana, India, shivakanth134@gmail.com
4. U. Mahesh, Department of CSE, Sreyas Institute of Engineering and Technology, Telangana, India, uppunuthulamahesh312@gmail.com
5. M. Keertana Rao, Department of CSE, Sreyas Institute of Engineering and Technology, Telangana, India, mogiligiddakeertanarao@gmail.com

ABSTRACT: Many effective techniques for object identification have been proposed as a result of the development of convolutional neural networks (CNNs). Remote sensing object detection (RSOD) is a difficult problem due to the following factors: 1) a complex backdrop of remote sensing images (RSIs) and 2) an excessively uneven size and sparsity distribution of remote sensing objects. Existing approaches are incapable of solving these challenges with high detection accuracy and speed. In this study, we suggest an adaptive balanced network (ABNet) to overcome these challenges. First, we create an improved effective channel attention (EECA) technique to boost the backbone's feature representation capabilities, which may help overcome the challenges of complicated backdrop on foreground items. Then, to capture additional discriminative information, an adaptive feature pyramid network (AFPN) is constructed to aggregate multiscale data adaptively in multiple channels and geographical locations. Furthermore, since the original FPN overlooks rich deep-level characteristics, a context enhancement module (CEM) is suggested to take use of extensive semantic information for multiscale object recognition. Experiment findings on three public datasets show that our technique outperforms the baseline by adding less than 1.5M new parameters.

Keywords *– Adaptive feature pyramid, context exploitation, local cross-channel attention, multiscale object detection, remote sensing image (RSI).*

1. INTRODUCTION

Acquisitions and uses of remote sensing images (RSIs) have gotten increasingly diversified as airborne technology has advanced. One of the hottest study issues in the field of RSIs analysis is remote sensing object detection (RSOD). It not only locates object areas of interest in RSIs but also categorises multiobject classes, which has been extensively employed in danger response, urban monitoring, traffic control, and other applications. Despite the fact that several methods have been presented for RSOD, particularly for largescale RSIs, this job remains difficult because to complex sceneries and multiscale objects. RSIs, unlike natural scene photos, are often obtained from satellites with broad vistas, resulting in large-scale photographs with background clutter. Furthermore, objects at different RSIs have varied scales because to differences in picture capture altitudes. Furthermore, some types of things, like as ships and vehicles, are often spread thickly in RSIs. The aforementioned concerns are the primary challenges to object recognition in RSIs, which causes most natural image algorithms to be

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

poorly fitted to RSIs. Most RSOD techniques based on convolutional neural networks (CNNs) are inspired by natural image processing approaches. There are two sorts of mainstream object detection approaches: two-stage and one-stage. The former specifies the job as a step-by-step refining process (extraction of areas and categorization of bounding boxes), while the latter executes a one-step approach. Faster RCNN is a two-stage representative approach for implementing the first end-to-end network for generic object identification. Its key innovation is the creation of a region proposal network (RPN) to collect suggestions rather than a sliding window. Typical one-stage approaches include YOLO, RetinaNet, and others. For example, YOLO splits the input picture into many cells using a single network. The projected bounding boxes (b-boxes) and category probabilities of each area are then output directly. These methods, however, are ineffective when dealing with multiscale objects.

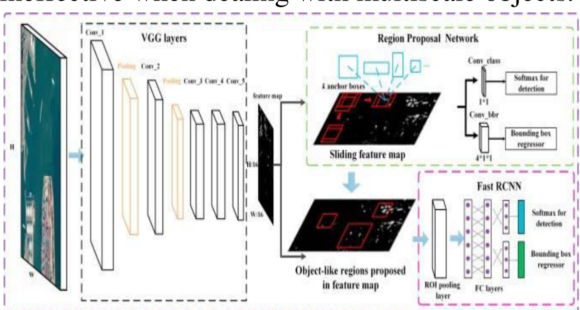

Fig.1: Example figure

For example, Faster RCNN [10] and YOLOv1, v2 solely forecast on the last layer of features. Based on this limitation, feature pyramid networks (FPNs) are used to recognise multiscale features. Following then, a lot of enhanced FPNs were extensively researched. However, FPNs only handle multiscale imbalance at the feature level and cannot resolve other imbalance issues. As a result, Pang et al. suggest balanced sampling and balanced smooth L1 loss to limit sample and objective level imbalance. Furthermore, Chen et al. offer an overlap sampler to choose instances and allow training to correct the sampling imbalance. During the distillation process, a neoteric loss function is created to attract positive pixels and minimise area imbalance between foreground and background. These models for natural scene object recognition help to advance the area of remote sensing.

2. LITERATURE REVIEW

Embedding structured contour and location prior in siamesed fully convolutional networks for road detection:

Road detection from moving vehicles' perspectives is a difficult problem in autonomous driving. Many deep learning algorithms, including as Convolutional Neural Networks (CNN) and Fully Convolutional Networks (FCNN), have recently emerged for this purpose because they can extract high-level local characteristics to detect road sections from raw RGB data (FCN). However, detecting the road's edge properly remains a difficult task. In this study, we present a siamesed fully convolutional network (dubbed "s-FCN-loc") that can use RGB-channel pictures, semantic contours, and location priors all at once to intricately segment a road section. To be more explicit, the s-FCN-loc contains two streams for processing the raw RGB photos and contour maps. Simultaneously, the position prior is immediately appended to the siamesed FCN to improve detection performance. We make three contributions: (1) An s-FCN-loc model is proposed that learns more discriminative features of road boundaries than the original FCN to detect more accurate road regions; (2) The location prior is viewed as a type of feature map and is directly appended to the final feature map in s-FCN-loc to effectively promote detection performance, which is easier than other traditional methods, namely different

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

priors for different inputs (image patches); (3) The convergent speed of training the s-FCN-loc model is The suggested method is tested on the KITTI Road Detection Benchmark and the One-Class Road Detection Dataset, and it obtains a competitive result when compared to the state of the art.

Dual feature extraction network for hyperspectral image analysis:

Hyperspectral anomaly detection (HAD) is a high-impact research project in remote sensing scene interpretation. In this paper, we present an unsupervised technique for HAD, dual feature extraction network (DFEN), to progressively improve discriminating between original data and background. We apply an end-to-end discriminative learning loss on two networks in particular. Adversarial learning, for example, seeks to preserve the original spectrum, while Gaussian restricted learning seeks to learn the background distribution in the potential space. We generate spatial and spectral anomaly scores based on mean squared error (MSE) spatial distance and orthogonal projection divergence (OPD) spectral distance between two latent feature matrices to extract the anomaly. Finally, to lower the false alarm rate even more, the complete detection result is achieved via a simple dot product between two domains. Experiments on eight genuine hyperspectral data sets acquired by various sensors across varied scenarios reveal that the proposed DFEN approach outperforms other related methods in terms of detection accuracy and false alarm rate.

Hyperspectral pansharpening with deep priors:

Although hyperspectral (HS) images may explain small changes in material spectral fingerprints, they have inadequate spatial resolution due to present technological and economical limits. We provide a potential HS pansharpening approach with deep priors (HPDP) for fusing a low-resolution (LR) HS picture with a high-resolution (HR) panchromatic (PAN) image in this study. Unlike previous techniques, we redefine the spectral response function (SRF) for the first time using the bigger eigenvalue of the structural tensor (ST) matrix, which is more in accordance with the properties of HS imaging. Then, on a band-by-band basis, we use HFNet to capture deep residual mapping of high frequency across the upsampled HS picture and the PAN image. The learnt residual mapping of high frequency is specifically injected into the structural transformed HS pictures, which are the extracted deep priors used as an additional constraint in a Sylvester equation to estimate the final HR HS image. Comparative evaluations show that the proposed HPDP approach outperforms other pansharpening methods by assuring greater quality in both the spatial and spectral domains for all kinds of data sets. Furthermore, the HFNet is trained in the high-frequency domain using multispectral (MS) pictures, which overcomes deep neural network (DNN) sensitivity to data sets recorded by multiple sensors as well as the challenge of inadequate training samples for HS pansharpening.

Weakly supervised low-rank representation for hyperspectral anomaly detection:

In this paper, we propose a weakly supervised low-rank representation (WSLRR) method for hyperspectral anomaly detection (HAD), which transforms deep learning-based HAD into a low-lank optimization problem capable of characterising the complex and diverse background in real HSIs while also obtaining relatively strong supervision information. Unlike previous unsupervised and supervised approaches, we begin by modelling the backdrop in a weakly supervised way, which produces higher performance without prior knowledge and is not constrained by densely accurate annotation. LRR is a useful strategy for further probing the complicated background structures when reconstruction biases produced by poorly supervised estimates are taken into account. Instead of directly using traditional LRR methodologies, a dictionary-based LRR is provided, which includes both observable training data and concealed learning data generated by the background estimation model. Finally, the resulting low-rank and sparse parts, as well as the outcome of the first detection, work together to find anomalies. Comparative studies show that the suggested WSLRR approach outperforms state-of-the-art methods in terms of detection performance.

A joint convolutional neural networks and context transfer for street scenes labeling:

Understanding the street landscape is a critical problem for autonomous driving. Scene tagging, which annotates each pixel in the photos with the appropriate class name, is an essential step in this approach. Despite the fact that various ways have been created, there are still certain flaws. For starters, many techniques rely on hand-crafted features with limited visual representation capability. Second, owing to dataset bias, they are unable to reliably categorise foreground items. Third, the classic Markov Random Filed (MRF) inference is prone to excessive smoothness during the refining step. This research presents a combination strategy using priori convolutional neural networks at the superpixel level (referred to as "priori s-CNNs") and soft constrained context transfer to address the aforementioned issues. We make three contributions: (1) A priori s-CNNs model is proposed to discriminately describe various objects; (2) A hierarchical data augmentation method is presented to alleviate dataset bias in the priori s-CNNs training stage, which significantly improves foreground object labelling; and (3) A soft restricted MRF energy function is defined to improve the priori s-CNNs model's labelling performance while reducing over smoothness. The suggested technique outperforms competitors on the CamVid dataset (11 classes) and the SIFT Flow Street dataset (16 classes).

3. METHODOLOGY

Nonetheless, most of these techniques are unable of performing effectively in complex scenarios, particularly for multiscale and dense objects in large-scale RSIs. Attention methods, for example, that are mainly built with fully linked layers, cannot be effectively used to RSIs since they are inefficient to incorporate into CNNs. On the one hand, these approaches lengthen the running time by requiring expensive calculation. On the other hand, they make fine-tuning the networks harder. Meanwhile, although several feature pyramids have improved the detection performance of RSIs to some amount, their structures remain convoluted and need a lot of processing. Furthermore, effectively detecting clustered items in RSIs is difficult. As a result, further research is required to address these issues with RSOD.

Disadvantages:

1. remote sensing photos with complex backgrounds (RSIs)
2. Extremely skewed size and sparsity distribution of distant sensing objects.
3. Existing approaches are incapable of solving these challenges with high detection accuracy and speed.

To address the concerns raised above, we offer an adaptive balanced network (ABNet) composed of a number of components that increase detection accuracy while maintaining better operating performance. To begin, we construct a portable mechanism called enhanced effective channel attention (EECA) to capture local cross-channel correlation in order to alleviate the difficult backdrop of large-scale

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

RSIs. Then, using adaptive pooling, an adaptive feature pyramid network (AFPN) is suggested, which first combines multiscale feature maps. Following that, we introduce a unique selected refined module (SRM) for AFPN reconstruction. A context enhancement module (CEM) is also included to compensate for the absence of contextual information in integrated features and to build a multiscale pyramid network for detection. The multilevel RPN is used to generate and select proposal candidates based on the multiscale pyramid characteristics. Finally, balanced L1 loss is used to train the detector consistently and precisely.

Advantages:

1. The suggested detector's efficacy.
2. Our technique outperforms the baseline in terms of performance.

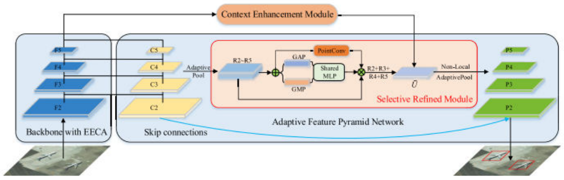

Fig.2: System architecture

MODULES:

To carry out the aforementioned project, we created the modules listed below.

- Data exploration: we will put data into the system using this module. Processing: we will read data for processing using this module.
- Using this module, data will be separated into train and test.
- Model generation: Building the model - yolov5, faster RCNN, AlexNet Backbone of RICNN, ZFNet Backbone of RICADO, ResNet50 Backbone of ABNet, ResNet101, and VGG16. Calculated algorithm accuracy
- User signup and login: Using this module will result in registration and login. User input: Using this module will result in prediction input.
- Prediction: the final predicted value will be presented.

4. IMPLEMENTATION

ALGORITHMS:

YOLOV5:

YOLO, which stands for "You Only Look Once," is an object identification technique that splits photos into grids. Each grid cell is in charge of detecting items inside itself. Because of its speed and precision, YOLO is one of the most well-known object detection techniques. The YOLOv5 Architecture as a Convolutional Neural Network Scheme (CNN). The BackBone, Neck, and Head are the main components. CSPNet is used in the BackBone to extract features from the photos used as input images. The Neck is used to create the pyramid feature.

Faster RCNN:

Faster R-CNN is a single-stage model that is trained from start to finish. It generates region suggestions using a new region proposal network (RPN), which saves time over classic methods like Selective Search. It extracts a fixed-length feature vector from each area suggestion using the ROI Pooling layer. Faster R-CNN is a deep convolutional network that appears to the user as a single, end-to-end, unified network when used for object identification. The network can predict the positions of various items correctly and fast.

AlexNet Backbone of RICNN:

AlexNet is an 8-layer deep convolutional neural network. A pretrained version of the network trained on over a million photos from the ImageNet database may be loaded. The pretrained network can categorise photos into 1000 different item categories, including keyboards, mice, pencils, and other animals.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

ZFNet Backbone of RICADO:

ZFNet is an example of a conventional convolutional neural network. Visualizing intermediate feature layers and the functioning of the classifier inspired the design. When compared to AlexNet, the filter widths and stride of the convolutions are lowered.

ResNet50 Backbone of ABNet:

ResNet-50 is a 50-layer deep convolutional neural network. A pretrained version of the network trained on over a million photos from the ImageNet database may be loaded. The pretrained network can categorise photos into 1000 different item categories, including keyboards, mice, pencils, and other animals.

ResNet101:

ResNet-101 is a 101-layer deep convolutional neural network. A pretrained version of the network trained on over a million photos from the ImageNet database may be loaded. The pretrained network can categorise photos into 1000 different item categories, including keyboards, mice, pencils, and other animals.

VGG16:

VGG16 is an object identification and classification method that can classify 1000 photos from 1000 distinct categories with 92.7% accuracy. It is a common picture classification technique that is simple to employ with transfer learning. The number 16 in the term VGG alludes to the fact that it is a 16-layer deep neural network (VGGnet). This suggests that VGG16 is a really large network with around 138 million parameters. It is a massive network even by current standards.

5. EXPERIMENTAL RESULTS

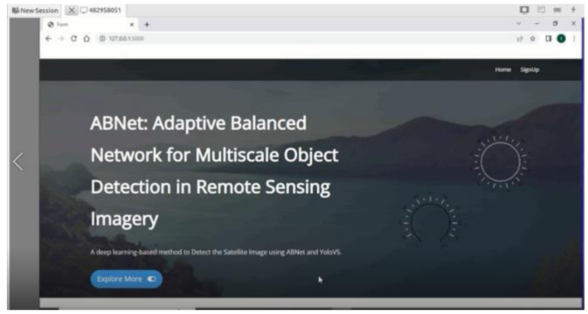

Fig.3: Home screen

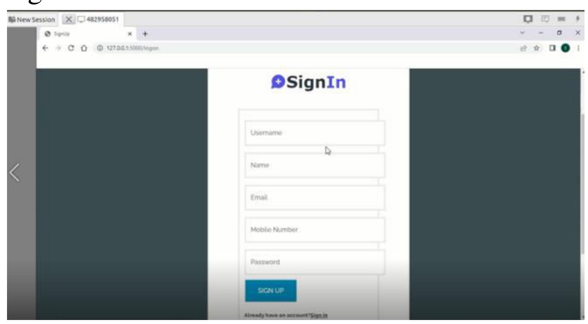

Fig.4: User signup

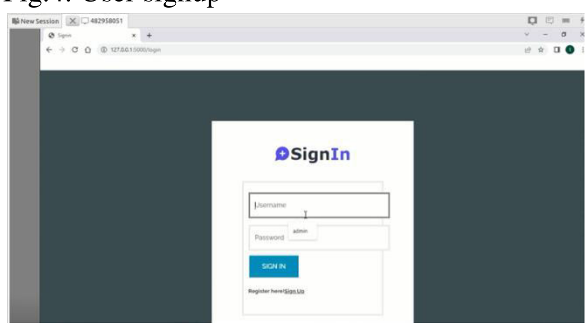

Fig.5: User signin

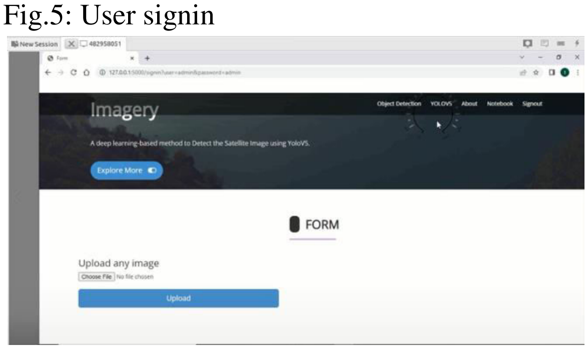

Fig.6: Main screen

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

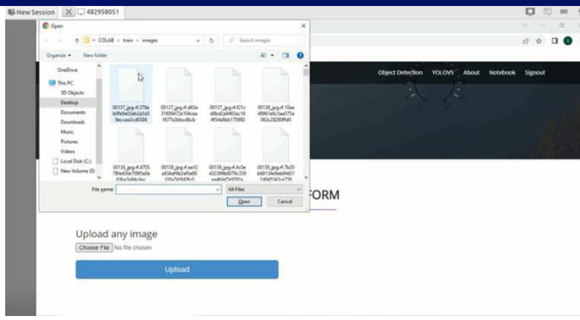

Fig.7: User input

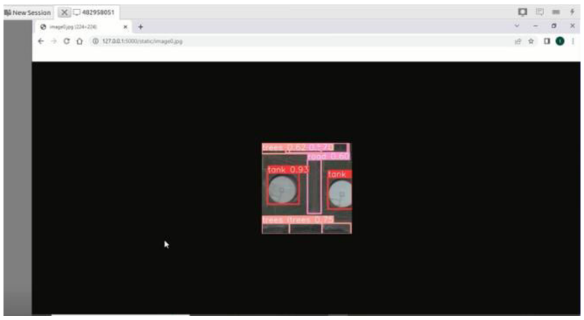
Fig.8: Prediction result

6. CONCLUSION

For RSIs, an upgraded detector ABNet with three enhancements based on Faster RCNN is suggested in this paper. To begin, the EECA method is developed to produce more effective channel feature extraction capabilities for ResNet in order to investigate correlations between local cross-channels. The EECA mechanism emphasises big items in RSIs while suppressing negative information from the intricate backdrop. Second, AFPN is created, which merely uses an MLP, a pointwise convolution, and a nonlocal block to effectively integrate feature maps of different sizes. Third, CEM is used to merge the backbone's deepest-level characteristics into AFPN and coalesce appropriate contextual information. Experiments on three publicly available benchmarks show that ABNet beats numerous state-of-the-art algorithms. Our solution adds fewer than 1.5M new parameters above the baseline, while yet maintaining a reasonable operating performance. We discover that ABNet's detection performance for tiny objects is not much enhanced. As a result, we will examine how to create a lightweight and better detector for tiny things in our future work. In addition, we will investigate the EECA mechanism's performance in various distant sensing tasks.

REFERENCES

[1] Q. Wang, J. Gao, and Y. Yuan, "Embedding structured contour and location prior in siamesed fully convolutional networks for road detection," IEEE Trans. Intell. Transp. Syst., vol. 19, no. 1, pp. 230–241, Jan. 2017.
[2] W. Xie, J. Lei, S. Fang, Y. Li, X. Jia, and M. Li, "Dual feature extraction network for hyperspectral image analysis," Pattern Recognit., vol. 118, Apr. 2021, Art. no. 107992.
[3] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 5, pp. 1529–1543, May 2020.
[4] G. Ganci, A. Cappello, G. Bilotta, and C. Del Negro, "How the variety of satellite remote sensing data over volcanoes can assist hazard monitoring efforts: The 2011 eruption of nabro volcano," Remote Sens. Environ., vol. 236, Jan. 2020, Art. no. 111426.
[5] W. Xie, X. Zhang, Y. Li, J. Lei, J. Li, and Q. Du, "Weakly supervised low-rank representation for hyperspectral anomaly detection," IEEE Trans. Cybern., vol. 51, no. 8, pp. 3889–3900, Aug. 2021.
[6] Q. Wang, J. Gao, and Y. Yuan, "A joint convolutional neural networks and context transfer for street scenes labeling," IEEE Trans. Intell. Transp. Syst., vol. 19, no. 5, pp. 1457–1470, May 2018.
[7] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," ISPRS J. Photogramm. Remote Sens., vol. 98, pp. 119–132, Dec. 2014.

[8] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, “Object detection in optical remote sensing images: A survey and a new benchmark,” ISPRS J. Photogramm. Remote Sens., vol. 159, pp. 296–307, Jan. 2020.
[9] J. Ding et al., “Object detection in aerial images: A large-scale benchmark and challenges,” IEEE Trans. Pattern Anal. Mach. Intell., early access, Oct. 6, 2021, doi: 10.1109/TPAMI.2021.3117983.
[10] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.