A STATE-OF-THE-ART AI & ML BASED DETECTION & IDENTIFICATION IN REMOTE IMAGERY

Sameer Kumar Panda[1*], Soma Dalbehera[2]
[1*]Assistant Professor, Department of Mechanical Engineering, Nalanda Institute of Technology, Bhubaneswar, Odisha, India
[2] Associate Professor, Department of Mechanical Engineering, Nalanda Institute of Technology, Bhubaneswar, Odisha, India
*Corresponding author e-mail: nabnitpanigrahi@thenalanda.com

Abstract:
Researchers have long been drawn to remotely sensed images and their related fields of application. There is a huge area where remote imaging is being used and making progress. Since the introduction of AL, ML, and DL-based computing, approaches for processing and analysing remote images have grown significantly and now provide a wide range of services, including traffic monitoring, earth observation, land surveying, and other agricultural fields. Machine learning and deep learning have been demonstrated as the most often utilised and highly successful strategies for object detection as artificial intelligence has grown in popularity among researchers. With the possibility of increased accuracy in the same, AI & ML-based object segmentation & identification makes this topic hot and appealing to researchers once more. The efficiency of exploiting remotely sensed imagery for business reasons has been highlighted by a number of researchers who have offered their efforts in the form of research papers. In order to extract hidden and useful information from remote photography, some preprocessing approaches have been explored in this article. This article also discusses object recognition and object detection using deep learning approaches used by many scholars. A chronological evaluation of the research related to detection and recognition utilising deep learning techniques is also included in this literature study.

Keywords: *Convolutional Neural Network, Remote Sensed Imagery, Object Detection, Artificial Intelligence,Feature Extraction, Deep Learning, Machine Learning*

1. Introduction

Remote imaging is a valuable resource for the world in many ways in this ever-evolving technological world. These days, data is gathered and stored in digital formats, allowing for its interpretation and analysis. Images for remote sensing are gathered using a variety of satellites, aerial photography, Lidar, Landsat, spy satellites, and sentinel photos. In order to analyse remote photography and extract any hidden information from it, it is now saved in digital form. Yet, one drawback of remote imaging is that the quality of the photos obtained from satellites isn't always up to par. The sights are typically hazy, noisy, and include a variety of colours channels. As a result, processing those data is required before applying any processing to them [31]. Manyprocesses and functions are used in digital image processing to format and rectify the data for segmentation and classification. The data can be refined and used for a variety of commercial objectives, such as Earth observation, weather forecasting, forestry, agricultural use, surface changes, and the analysis of bio-diversity, with the aid of those methods and techniques. Moreover, remote applications can be used to identify crop conditions, assess road conditions in rural locations, and more. For a very long time, deep learning methods, which are a component of neural networks, have been used to process and evaluate remote sensing picture data. However, prior to the invention of deep learning techniques, remote sensing imagery was studied using a support vector machine (SVM) and other ensemble classifiers, such as Random forest, for change detection or image classification. Due to its capacity to handle large and multi-dimensional data with a small amount of training data, SVM has attracted a lot of interest and demand [2].
Recent developments in DL have reignited interest in neural networks among the remote

community. The entire remote sensing community has shifted its focus to deep learning (DL) since 2014 as DL techniques and algorithms have demonstrated their success in a variety of image analysis tasks, including object detection, scene classification, and Land cover and Land use observations [30] [14] [36] [52] [35] [53] [54]. Reading through the extensive DL literature reveals that DL has general approaches connected to the creation of fundamental deep learning algorithms [55] and in-depth reviews for numerous developing and cutting-edge fields including speech recognition and medical image identification [56]. In several studies [4], DL in remote sensing applications is demonstrated. The applications of DL in the classification of remote sensing images for significant observations were the subject of the literature review by [57]. [15] has carried out a thorough, comprehensive review, particularly concentrating on related, unusual sub-domains of distant data application areas, such as 3-d modelling. Within the realm of remote sensing, DL algorithms and techniques have a variety of sub-do- mains, and the application fields are constantly expanding to obtain a more quantitative and systematic examination of the data [32]. The goal of this work is to create a thorough analysis of deep learning (DL) methods in various remote sensing applications, such as object detection, image segmentation, in both photos and videos, we may perform classification, image registration, image fusion, etc. [33]. We compiled the results from numerous research articles after doing a thorough analysis in the field of object detection in remote images using deep learning algorithms. A critical summary is presented at the end, followed by the biggest gaps and issues discovered.

Remote Imagery and Preprocessing

Remotely sensed images are not simple images, they contain multiple formats and resolution chal- lenges. They can be single channel or multi-band images having variations in their resolutions too. The spatial resolution of remote imagery is the most important aspect that is directly related to the accuracy of objects. To generate land cover maps for various reasons like environment planning, change detection, transport, and traffic planning temporal resolution is used. Medium resolution re- mote sensed imagery is used for data integration, analysis of urban areas, also to differentiate various zones like residential, industrial, and commercial. By reviewing a huge number of databases related to various articles on remote sensing data, features, and parameters, the information is summarized and shown in Table 1.

Tab. 1. Attributes used for Remote Sensing and DL

Attributes	Categories
Remote Sensing Data	Hyper-spectral, Lidar, SAR, etc.
Target Study Area	Urban, Agriculture, Rural, Water
DL Model	CNN, RNN, AE, DBN, other
Target	Scene Classification, Image Fusion, Object Detection, Segmentation, LULC Classification, and

ISSN: 2278-4632
Vol-10 Issue-3 No.01 March 2020

	other
Processing Parameters	Object, Pixel
Samples for Training	Value
Accuracy	Value
Study Site	Value
Paper Category	Conference, Journal
Image Resolution	Value (high resolution, coarse, moderate)

Satellite data includes various other resolutions and types of images. Spatial Resolution is to mea- sure the closed lines in an image. Spatial resolution is dependent on the device from which the image is captured. It is not only to measure the capacity of ppi (pixel per unit). The spatial resolution of any image decides its quality in the form of clar- ity. It generally refers to the count of independent pixels per unit in any image. Spatial resolution is limited by aberrations, diffraction, imperfect focus, and atmospheric distortion. Figure 1 (a), 1(b), 1(c) is showing the difference between multiple range spatial resolutions.

Spectral Resolution is to resolve spectral features as separate components, spectral resolution is used. Color images involve multiple and distinguished light effects on different spectra as Fig. 2 is showing. Multi- band images can resolve finer differences of wave- lengths or spectrum by storing or measuring common RGB images.

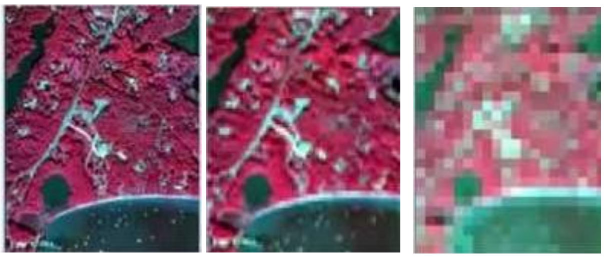


Fig. 1.(a) Spatial Resolution of 1 meter (b) 10 meter (c) 30 meter

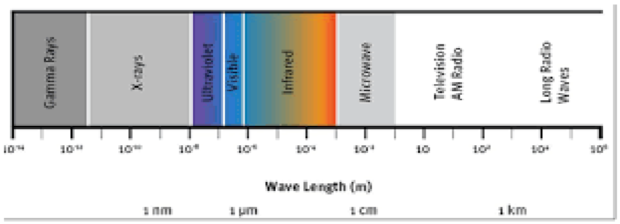


Fig. 2. Spectral Resolution of Remote Sensed Imagery

A temporal resolution is a measurement unit of any area concerning time as a movie and high-speed cam- eras can capture the scenes at many different points in time. Time resolution that a general movie camera captures generally at a rate of 24-48 frames/ second. Although, high-

ISSN: 2278-4632
Vol-10 Issue-3 No.01 March 2020

speed cameras can capture the scenes at a very high speed up to 50-300 frames/ second or even more. Radio-metric Resolution is to determine the fine representation and differentiation among intensity, radiometric resolution is used. It is gener- ally represented as the number of bits or levels. For ex- digital image is having 256 levels i.e. 28 bits. The reflected intensity will be better and finer than most as the bits are higher. While working practically, noise levels are used to limit radiometric resolutions in- stead of bits representation. Multi-Spectral Resolution is performed on the multi-spectral image the image data is captured across the range of electromagnetic spectrum at a specific wavelength. The wavelength can be captured with the help of supported devices that can separate to capture or detect by filters some- times beyond the visibility range of light like- ultravi- olet and infrared. Hyperspectral Resolution is gener- ally applied on hyperspectral images is also a kind of multi-spectral image that captures the image data at several different wavelengths of the electromagnetic spectrum. To extract the data for each pixel in an im- age for detection of objects, material identification, hyperspectral images are used.

As remote sensing imagery is having various reso- lutions and dimensions of the data, rectification, and restoration of data became the important aspect for getting desired information or extracting hidden data by analyzing the image. The pre-processing operation is generally applied to correct and refine platform or sensor-specific radiometric data. It also includes the geometric distortion of the data. For eliminat- ing scene illusions, sensor noise, etc. these types of pre-processing are required for remote imagery. Various pre-processing methods are used to rectify the data collected by the sensor are included further. Each of these methods is different by their working nature or by having a different sensor or platform used for data acquisition.

Radiometric Corrections. Correction of data that includes the issues related to unwanted sensor data, irregularities of sensor data is the prime function of radiometric corrections. After the corrections, the data is converted to accurately measure the reflected light by the sensor.

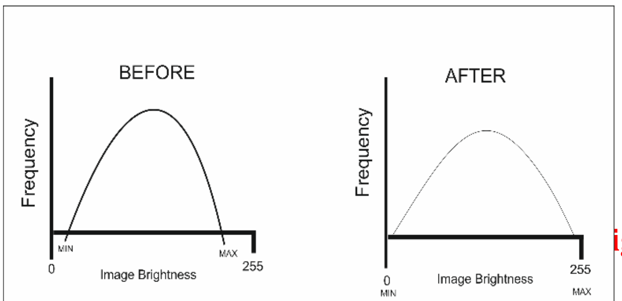
Contrast Enhancement. It involves increasing in contrast value among background and target. An im- age histogram is a basic concept to understand con- trast enhancement.

Geometric Corrections. For correcting the ge- ometric distortion that occurs because of Earth-sen- sor geometrical variations and to convert that into real-world earth-surface coordinates, geometric cor- rections are used. Distortion can arise due to reasons that include altitude variations, the velocity of the sensor platform, earth convexity, atmospheric diver- sion, length of the displaced object.

Spatial Filtering. To reduce the smaller details in an image, Low pass spatial filters are designed to focus more on homogeneous large pixels of the same tone. This makes the smooth appearance of an image. These kinds of spatial filters are very much useful for reducing random noise from the image. Median and Average filters are an example of low pass spatial fitters [34]. On contrary, High pass filters work just opposite the operation that a low pass filter does. It operates to sharpen the fine details in the image to fine-tune the appearance. Some filters like edge or di-rectional detection filters are used to identify the field boundaries or roads in an image.

Band Rationing. One of the most commonly ap- plied transformations is spectral rationing or band rationing. It serves to highlight and focus the spectral variations of surface covers.

Piecewise Linear Stretch for contrast enhance- ment. To utilize the complete range of value in bright- ness component, the maximum and minimum param- eters of data allocated to new applied data. For e.g.: an image is having a minimum brightness value of 45 and a maximum of 205. If that image (Fig. 3) is rep- resented without enhancement, the values from 0 to 44 and the



values from 206 up to 255 will not be dis- played. By stretching the values from 0 to 45 and up to 205, the important features can be accessed.

Fig. 3. Piecewise Linear Contrast Stretch

Satellite Image Processing Advancements Deep learning is a subset of Artificial intelligence that helps in creating automated applications and services for performing analytical physical tasks without human involvement. Now deep learning is behind each everyday product including digital as- sistants, self-driven cars, credit card fraud detection, and many more emerging techniques. Deep learning involves a neural network with a minimum of three layers. This network is developed to simulate human behavior through machines by learning them with a huge amount of data. Deep learning is an organized structure of multiple hidden layers. A single layer in a neural network is capable of predicting results; still, additional layers can be used for enhancing accuracy and efficiency [35].

Recent work in this field is showing that deep learning has achieved a lot in the field of replicating human behavior either in a simple task or complex operations like object detection, and image classifica- tion [36]. If comparing deep learning with other tra- ditional approaches it is performing outstandingly in result predictions with good accuracy. Deep learning was firstly introduced in the 1980s and it has become the most emerging technology for serving the world in various domains. It requires a large amount of la- beled data with the highest computational power to train the model to get more accurate results. Deep learning models learn features directly from data other outside feature extraction techniques are not required while working with a deep neural network. Deep learning can be classified into various catego- ries like Supervised and Unsupervised. Feature ex- traction is one of the most important aspects of deep learning that uses an algorithm to construct meaning- ful features automatically for training, understanding, or learning.

From 1943 till now deep learning is being used and improving its applications day by day. Image process- ing through deep learning can be tracked since early 1943. In [58] created a neural network like the hu- man brain by using a combination of algorithms and threshold logic. In [59], researchers have developed a continuous backpropagation model by improvising the basic neural networks. In 1962, a simple neural network model for image classification using the ba- sic chain rule is developed. In 1965, a deep learning model for group data handling is developed. During the 1970s, the very first AI winter came into existence that uses some AI techniques for basic image process- ing. In 1973, Neocognitron that was an artificial neu- ral network that used multilayered and hierarchical designs is created. The proposed system was able to recognize visual patterns through the computer. In [50] demonstrated backpropagation by combining CNN along with backpropagation for reading hand- written digits through the computer. In 1991, a mod- el is created to identify the problems related to van- ishing gradient. In [60] a paper is presented on deep belief networks for learning the images in a faster manner. For speech recognition, deep learning algo- rithms are developed. In 2009, ImageNet is launched to serve deep learning researchers. In 2012, AlexNet is developed which is constructed with multiple GPUs concept. In [61] Generative Adversarial Networks is developed to enhance deep learning abilities in sci- ence, art, and fashion.

Advancement in Image Processing Computing Vision

Over recent years, Computer Vision tasks, such as ob- ject detection, image classification especially in digi- tal images and videos grasp a lot of attention from researchers. Computer vision is a part of Artificial Intelligence that makes the computer enables every- thing that a human mind can perform from simple calculations to analyzing or reading minds [3]. Object detection is a technology related to Computer Vision that directly deals with the detection of objects from various classes [37]. Among all the problems and challenges available in image recognition, object de- tection is the major challenge to solve with the help of the Computer. Object detection is an emerging tech- nology nowadays as it consists of various applications in various domains like Face Recognition, Video- Sur- veillance, Crowd Counting, Transport Management, Image annotation, Video co-segmentation, tracking the locations of a ball during a cricket match, etc. Figure 4 is showing the milestones achieved in the field of object detection through traditional and deep learning methods.

Computer vision in favor of AI is touching the benchmarks in the world of the computer by increas- ing its computational power and results in calculation abilities with more accuracy and reliabilities. In [37] highlighted and focused on the role of computer vi- sion in his work. Several deep learning methods are adopted to develop deep learning models for remote sensing images; some of them include transfer learn- ing, learning rate decay, training from scratch, and dropout. Nowadays Artificial Intelligence and the various ways of computing of achieving the same are also in practice and evaluation. So, the researchers are looking the AI ways like Machine Learning and Deep Learning to attain new and amazing results in remote sensing imagery. Hence, during this study, we will be exploring various facets of remote sensing imagery, advanced artificial intelligence-based machine learn- ing, and deep learning techniques for remote sensing imagery segmentation, object detection, and classifi- cation.

In [38] presented various examples of measuring space clustering processes for a range of three mul- tispectral images captured over Ariz, Phoenix. In [51], NASA proposed the study at the Centre for Research, the University of Kansas with the cooperation and support of government research agencies and other universities have shown the applicability of satellite imagery in various fields within agriculture, ocean- ography, and earth sciences. This paper shows how characteristics and features of radar are used to get geo-science information.

In [39] [40], the authors described a process for spatial registration of multi-temporal and digital multi-spectral imagery. Experimental results are also defined here as the result of correlation analysis be-

tween digital satellite photographs and multi-spectral imagery. The registration process of space photogra- phy and multi-spectral airborne line-scanner digital imagery is also described here.

In [41], authors attempted to identify between welters of results to get unbeatable achievements also liked hurdles, also to assess the contribution to get expected image-processing operations in exper- imental and operational usage of the upcoming tor- rent of raw data.

Recent trends in deep learning techniques are emerging with powerful methods for automated sys- tems that involve automatic feature learning through raw data. Most particularly, these methods achieved benchmarks in object detection; this area has become the most interesting area for new researchers [24].

Along with computer vision is touching the trends of the computer world in computational pow- er with a tremendous speed, and result producing capacities with more reliability than human minds. In [37] highlighted the role of vision tasks with the help of certain projects to

ISSN: 2278-4632
Vol-10 Issue-3 No.01 March 2020

prove the tested approach for various research works like virtual skinning, vir- tual painting, human-computer interaction, depth recovery, etc.

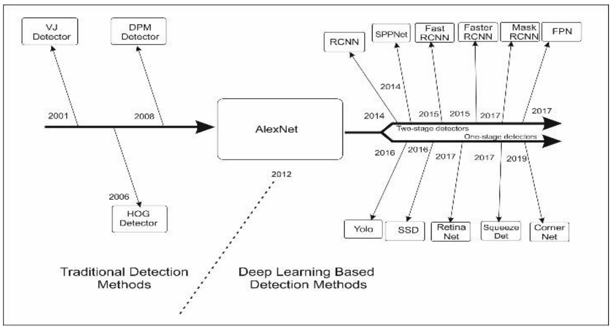


Fig. 4. Object detection milestones [21] [24] DPM,RCNN

Object Detection & Identification

In [42] detected motion of objects using cellular neu- ral network. In [43] provided a model to detect and focus objects in an unfocused background. In [45] de- veloped an object-oriented based segmentation algo- rithm for edge extraction. In [46] developed a model to detect patterns in spectrograms using cellular neu- ral network. In [47] developed a robust model based on deep learning methods to detect pedestrian using low- and high-level features. In [48] developed a mod- el using CNN for logo detection of vehicles. In [53] pro- vided their review for object detection algorithms for optical remote sensing imagery. They focused to re- view the brief history of available literature for deep learning techniques and the datasets for the detection of objects from remote imagery. Table 2 is represent- ing the work (done in the period 2012-19) by many researchers on various datasets and extracted results for object detection in earth observation [2]. The re- view is given for dataset with the category of images, image count including width and the detected objects annotation style. A number of methods are developed for detection and recognition of objects. Many de- tector algorithms are proposed like VJ (Viola Jones),

HOG through SVM, Feature dependent detectors [45] through DL techniques.

Tab. 2. Differences among various open datasets (from2012-2019) [15]

Year	**Datasets**	**Instan-ces**	**Cate-gories**	**Image Width**	**Im-ages**	**Annotation Style**
2008	TAS	1319	1	792	30	Horizontal Box
2012	SZTAKI-INRIA	665	1	800	9	Oriented Box
2014	NWPU VHR-10	3775	10	1000	800	Horizontal Box

2015	VEDAI	3640	9	1024	1210	Oriented Box
2015	UCAS-AOD	6029	2	1280	910	Horizontal Box
2015	DLR 3K Vehicle	14235	2	5616	20	Oriented Box
2016	HRSC2016	2976	1	1000	1070	Oriented Box
2017	RSOD	6950	4	1000	976	Horizontal Box
2017	DOTA	188282	15	800-4000	2806	Oriented Box
2018	DIOR	192472	20	800	23463	Horizontal Box

Deep CNN Architecture

Unlike traditional neural networks, CNNs uses convo- lution operation in their layer [50]. CNN's convolution operation includes multiple stages consisting of four main components, kernel, convolution layer, activa- tion function, and a pooling layer. Each stage repre- sents a feature map in the form of an array [34]. Fig- ure 5 is showing the detailed workflow of CNN having several convolution layers with one or more fully con-Here, represents the trainable bias parameter ma- trix, which is a filter that connects the jth feature map of the (l-1) layer with ith feature map of layer (l). (*) is a 2-D discrete convolution operator.

Convolution Layer. Convolutional layer is the col- lection of multiple layers that can calculate the convo- lution of the inputted image by network weight. The first layer consists of the neuron that can view small images and learn some basic and limited features through it. As the network goes deep inside, layers can view a large portion of the image and can learn more expressive and detailed features by merging previous levels. Each layer is characterized by hy- per-parameters to train through spatial features, zero padding, and stride values between various windows that work to control the output layer. The expression can be given as follows-

$$S_{i,j} = (I * K)_{i,j} = \sum_m \sum_n I_{i,j} * K_{i-m,j-n} \quad (2)$$

In the process of convolution, a small window slides across the image from left to right and top to bottom. At each sliding window location, the sum of the product is calculated with each kernel element with the input element. The process continues with different kernels to form various kinds of feature maps.

The feature map is having a lesser size than the in- put image. However, we can pad the values to keep the size the same. The stride shows the gap size among two successive positions. Commonly, a stride size of 1 is chosen, but a greater size can be chosen to reduce the resolution of feature maps.

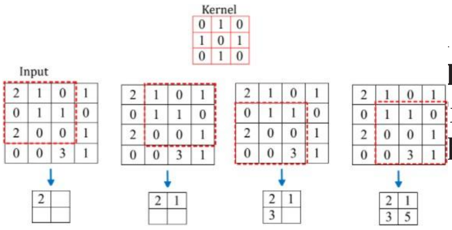
Nonlinear activation function. By getting the feature map, a nonlinear activation function is ap- plied in this process. As Eq. 3 is showing, the calcula- tion functions of activation maps carried only the ac- tivated features to the next layer. Then, the activation function can be

ISSN: 2278-4632
Vol-10 Issue-3 No.01 March 2020

formed as follows

nected layers, for producing the final output as a clas-$\emptyset(Y^{(l)}) = (B^{(l)} + \sum^{m1}$ $^{(l-1)}$
$(l) * (l-1))$ (3)

sification module. The main components of CNN are described as follows.
i i j=1 ij j



be used in the convolution process are as follows:

is **ReLU).** Most commonly used activation function. It of igure 6 (a) is showing ReLU the graphical representation hematically it can be defined as

$$f(x) = max(0, x) = \begin{cases} 0 \; if x < 0 \\ x if x \geq 0 \end{cases}$$

(4)

Fig. 5. Convolutional Operation with image matrix (4×4) with kernel matrix (3×3)

Kernels. At each location, each kernel is aimed to de-tect a particular characteristic. There exists a bank of

Sigmoid Function. As figure 6 (b) is showing, it can be represented by a curve like "S". It is generally used to predict the results as the function varies be- tween 0 and 1. It can be defined as

m1 filters in every convolutional layer, the output ofthe lth layer consisting of feature maps of size *. The

$f(x) = \frac{1}{1+e^{-x}}$ (5)

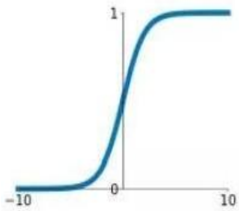
ith feature map can be computed as follows-

$Y_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{m1} K_{ij}^{(l)} * Y_j^{(l-1)}$ (1)

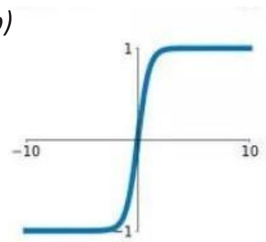
Hyperbolic Tangent (tanh) function. It is very much similar to the sigmoid function as it can be seen in figure 6 (c). The difference is in range, the range is here is [-1, 1]. The benefit of using the tanh function is that here negative numbers are mapped as strongly negative and zero values mapped near zero.

Dropout Layer. This layer is used to help in over-fit- ting issues also improves the network's performance. The dropout layer can be applied in any of the layers.

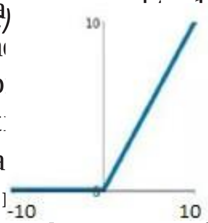
$f(x) = tanh(x) = \frac{2}{1+e^{-2x}} - 1$



(b)



Fully Connected Layer. At this layer, the final out- put is represented in the form of a 2-D array by con- nected with a fully connected layer. By the result of the Convolution process, this layer classifies an image into various classes. The activation function at the last layer computes the probability of results belonging to each class [16]. Commonly, for multi-class classifica- tion, the softmax activation function is used having a range of probability between [0, 1] with a ... The fully connected layer is situated at the last layer in a neural network. Th... layer is having the neurons that have a full connection to all the neurons of p... 3]. The calculations of this connection can be evaluated by matrix multi- plicat... followed by bias offset [62].



The rema... er is being organized as section-2 is showing the various existing pre-trained ... eNet and CIFAR-10 datasets, section-3 is giving the inceptive contributions at beginning of re- mote senses image processing, section-4 is the detailed literature review as a tabulated formation categorized

Fig. 6. Representation of various activation functions: a- ReLU b- Sigmoid, c- tanh

Normalization Layers. To implement inhibition par- adigms, the normalization layer is used for the obser- vation of the biological brain [62].

Pooling Layer. After each successive convolution layer, the pooling layer is presented that can reduce input layer size via some non-linear functions. They also help in reducing the computational and para- metric amount in the network. It also helps to control over-fitting [63]. Figure 7 is representing the pooling operation with [2 x 2] filter. Following are some main pooling operations:

Max pooling. The maximum value is calculated for each input patch. It preserves the maximum value of each stride while sliding over the feature map. Its mathematical representation is as follows

$$f(A) = max_{m,n}(A_{nxm}) \quad (7)$$

Average pooling. For each input, it calculates the av- erage value. This layer divides the input various pool- ing regions by computing their average values.

by various technological benchmarks used, section-5 shows the challenges mapping with reviewed litera- ture and section 6 involves the complete summary.

2. Existing Pre-trained Deep Leaning Modelsand Datasets

Deep neural networks, such as Convolutional neural networks [64], Recurrent neural networks [65], Graph neural networks [66], Attention neural networks [67], have been applied for various AI tasks at a wide range. The pre-trained models can be considered as the models that are developed based on any exist- ing techniques (As mentioned various deep learning techniques). The models that have been already got trained on some dataset like ImageNet, CifarNet-10 are known as pre-trained models. Nowadays pre- trained models are widely used for image classifica- tion and object detection. Table 3, 4are showing the summary and comparison of various pre-trained models trained on ImageNet and CIFAR-10 dataset respectively.

Tab. 3. Various pre-trained models trained on ImageNet

$f\ (A) = \frac{1}{\quad} \sum^{n} (\sum^{m})$

dataset

ave

ISSN: 2278-4632
Vol-10 Issue-3 No.01 March 2020

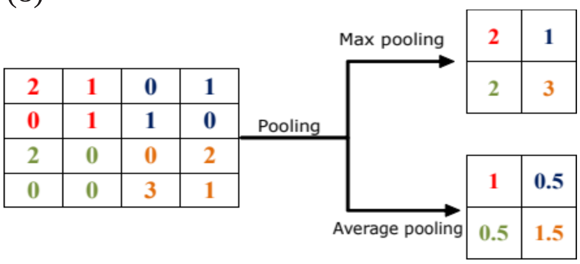
$n + m$

i=1

k=1

i,k

(8)



Model	Depth	Size (MB)	Parame-ters (Million)	Top-1 Accuracy	Top-5 Accura-cy
Xception	81	88	22.9M	79.0%	94.5%
VGG16	16	528	138.4M	71.3%	90.1%
VGG19	19	549	143.7M	71.3%	90.0%
ResNet50	107	98	25.6M	74.9%	92.1%
Inceptionv3	189	92	23.9M	77.9%	93.7%
DenseNet201	402	80	20.2M	77.3%	93.6%
EfficientNetB0	132	29	5.3M	77.1%	93.3%

Fig. 7. Operation of Pooling using 2 x 2 filters with stride of two

Tab. 4. Various pre-trained models trained on CIFAR-10

dataset

Model	Depth	Size (MB)	Parameters (Million)	Top-1 Accuracy	Top-5 Accuracy
DenseNet121	242	33	8.1M	75.0%	92.3%
EfficientNetV2L	479	52	1.19M	85.7%	97.5%
MUXNet-m	289	62	2.1M	98.0%	98.3%
AutoFormer-S 384	384	81	23M	99.1%	99.2%

3.Literature Review

In [6] authors presented a model for the interpreta- tion and evaluation of scenes for image analysis.

In paper [5] authors proposed a blackboard mod- el as a control structure for the detection of objects in aerial images.

In [7] researchers focused their research on feature extraction in SAR images as this process has a higher level of complexity due to the level of noise and quality issues in these images. This paper implemented an au- tomated algorithm for feature extraction of large-scale objects from SAR imagery. For the experiments, images of the Ottawa area are used through SAR imagery.

In [8] authors proposed a model to detect a cover change in the forest on Landsat data.

In [9] authors developed a model for the detection of manmade objects like airports, bridges, industries, etc.

In [10] authors proposed a classification and detec- tion model named CADCM to target hidden objects in hyperspectral imagery. The process is accomplished in three phases. Initially, a band selection process is ap- plied, next band rationing is done and finally, automatic target detection is achieved. Results show the targets hidden by natural background, shades, or objects can be detected finely.

In [11] two matrices namely object space and im- age space are used to refer and monitor the existing monocular building detection system with the help of 83 images collected by 18 different sites. By the anal- ysis, the effects of image inscrutability along with ob- jects complexity are examined. Edge fragmentation is also shown here in this research. The usage of rigorous photogrammetric space modeling is also demonstrat-ed.

In [12] a review article is presented to reduce cloud impacts by analyzing various existing algorithms.

In [13] an automatic approach for building foot- print extraction and its 3-D reconstruction from the imagery of airborne light and ranging (LIDAR) data is represented. Initially, a digital surface model (DSM) is generated to extract objects higher than the ground surface. To separate a building from other objects, ge- ometric characteristics such as size, height, and shape are used. Extracted building footprints are simplified for better quality using an orthogonal algorithm. Roofs are identified by information like ridgelines and slopes. Finally, an accuracy assessment is conducted by com-paring the results with manually digitized building ref-erence data.

In [14] authors represented the image analysis method for the extraction of building features. They have used three consecutive steps to accomplish this task. Initially, the supervised neural network is inputted by RGB multi-band images for roof identification. Next, spatial details are extracted through a hybrid approach of edge and region segmentation. Lastly, the extracted information is used to refine the results.

In [15] authors proposed a building extraction method. GIS data is used as an input in this method. A segmentation algorithm is used to extract the fea- tures of the building. GIS data is used to provide prior building knowledge. Data pre-processing, Object seg- mentation, and result post-processing are the three steps used in this method. Experimental results are also included in this to showcase the efficiency of the algorithms.

In [17] a two-step model method for tree detection is implemented including segmentation followed by classification is proposed here. The results presented show the effectiveness of the approach.

In [18] authors detected bridges in multispectral remote images through their developed model. The multi-seed supervised classification technique is used to classify the multispectral image into eight land-cov- er types. A knowledge-based approach is used that find out the spatial arrangement of the bridge and its sur- roundings. Testing is done on the IRS-1C/1-D satellite has a spatial resolution of 23.5m.

In [19] authors investigated ship detection in Ter- raSAR-X (TSX) ScanSAR images (19-m resolution). Kol- mogorov-Smirnov test is applied for the verification of the goodness of fit for the K-distribution to TSX images. A target detection algorithm is developed and also ver-ified.

In [20] amorphous-shaped objects are detected by their developed model. The model is showing the re- sults of experiments achieving high accuracy rates.

In [21] an automatic content-based analysis is pre- sented to detect arbitrary objects in aerial imagery. In this, the two-stage training model using a convolution- al neural network is implemented also verified over remote imagery. Model is tested for accuracy using UC- Merced data set with an accuracy of 98.6%.

In [22] a deep CNN model is invented with en- hanced functionality for feature extraction along with region classification and region proposal. Their method is based on ResNets that consists of multiple sub-networks (Object detection and Object proposal). To enhance feature map resolution, the output gener- ated by multiple-scale layers is combined. VHR-10 da- taset is used to train the model in the proposed work.

In [23] two algorithms are studied (Edgeboxes and Selective Search) for object detection. The evaluation is also performed on both algorithms using high-reso- lution remote imagery. Algorithms are tested and eval- uated through the NWPU-VHR-10 class data set. For performance measurement, execution time and recall rate are used as performance parameters. By the statis- tical results, authors proved that EdgeBoxes algorithm is showing optimal results over Selective Search in re-call rate.

In [25] a feature fusion method is proposed for extracting the fine-grained features from multiple layers of the remotely sensed image as those images have dense features concerning intra-class differenc-es, high inter-class similarities, and multidirectional objects. They have initially used ResNet50 for extract- ing the features from multiple layers then the channel attention method is used to enhance the features. As a result, cross-layer bilinear pooling and feature con- nection are used for fusion. After extracting the fea- tures through ResNet50, squeezing-and- excitation (SE) method is used in its advanced version. Along with this, the traditional activation functions are replaced with the HandTan function which is simpler and more effective than traditional activation functions. The deep learning framework PyTorch is used to build the train- ing model. The comparison is shown in their research with the existing models.

In [27] a branch regression framework is presented in a dual-mode based on remote image observations in detecting targets. This model can independently pre- dict various variables and orientations effectively. To deal with multi-level features through spatial pooling, an advanced smart feature is added to the research. This is an example of an advanced, accurate model that is able in doing parallel operations of localization and classification.

In 2020, Yang et al. [28] suggested a novel cloud detection neural network with an encoder-decoder structure, called CDnetV2 as a sequence of work in the detection of Cloud. If Channel Attention Fusion Model (CAFM), Spatial Attention Fusion Model (SAFM), and Channel Attention Refinement Model (CARM). Pro- posed HFFM is used to extract the linguistic high-level information from HLSIGFs. Here, some experimental results on the ZY-3 satellite thumbnail data set show that the proposed CDnetV2 is achieving exact and ac- curate results and it is outperforming several state-of-the-art methods.

In [29] authors proposed a novel approach to cope with such kinds of variant labels, i.e., class attention module and decomposition-fusion strategy. A class attention module is created to generate multiple class attention modules. Salient detection is proposed which

breaks down semantic segmentation into multi-class major detection and then combines them to produce a semantic map. Some experiments have been done on US3D Dataset. The imbalance label problem is also resolved with more accuracy than the previously avail- able approach.

In [26] PTAN is developed including three-stage strategies for object detection in HD images. The model is achieving a mAP value of 0.7958. On the NWPU VHR- 10 dataset, PTAN is achieving a mAP value of 0.9187.

In [4] authors proposed an up-sampling, down-sam- pling feature pyramid for obtaining the richer context information by bi-directionally involving shallow and deep features, and skipping connections. Experiments are done on DIOR, NWPUVHR-10, and on the self-as- sembled datasets SDOTA, SDD to show the excellent performance of the proposed method by comparing it with other detectors. The proposed method is achiev- ing 74.3% mAP on the public DIOR dataset.

In [1] authors proposed EFPN to detect small ob- jects like plants, small buildings, etc. It is created to im-prove feature extraction capabilities.

In [3] authors recommended a dynamic curriculum procedure that can learn the object detectors with the help of training images.

4. Literature Review Findings

Table-5 is showing the major findings by thoroughly reviewing many articles and research papers on ob- ject detection using deep learning and machine learn- ing. From the past till now many researchers have presented their research work and ideas for feature extraction [7] and object detection including various kinds of objects in remotely sensed images [5] like buildings [13], trees, plants [1], roads, commercial ar- eas, water, quality of soil, ships [4], oil slicks, bridges, industries, airports [9], shadow detection [10] and cloud detection [3].

Tab.5. Literature Review Findings

Reference	**Algorithm/ Model Methodology**	**Data Set**	**Parameter**	**Research Gap**
Benchmark Technique: CNN Deep Learning				
[1]	A feature fusion method based on improvised ResNet50	Open access Remote imagery	Validation of dataset	Need to be more accurate
[3]	MSCNN	Geospatial VHR Satellite images NWPU VHR-10 Challenging data set 650 images with a resolution of 0.5–2.0 m.	Accuracy	Scale and rotation dependent
[26]	CDnetV2	ZY-3 satellite thumbnail	Accuracy with fine grained features	Model is providing various accuracy rations in variant models. Need to create a hybrid model with more accuracy.
[4]	Rotated Region Proposal Network	HRSC2016 dataset.	work on dense dataset	More classes and sub-classes can be included

Reference	**Algorithm/ Model Methodology**	**Data Set**	**Parameter**	**Research Gap**
[49]	object relationship reasoning CNN	Aerial Image data set (AID), UC Merced Land-Use data set, and	**accuracy for multiband data**	Need more prior information and parameters of the geometrical shape

	(ORRCNN)	WHU-RS19 data set		for template designing
[69]	Deep learning algorithms on NVIDIA DGX-1 supercomputer	Pre-trained dataset of SpaceNetfine-tuned on planet database.	Time Complexity & Efficiency	For the training of CNN, a huge set of training data along with more computation powers is needed
[22]	An enhanced deep CNN based	VHR-10 data set	substantial number of densely packed objects	Need to be more enhanced
[21]	Two-stage training model using convolutional neural network	UCMerced Dataset	Arbitrary objects	Sometimes outliers are involved in predictions
[51]	**AASM**	**Open access satellite images**	Ability and Efficiency	Sensitive to shape and viewpoint change
[35]	Convolutional capsule network	Open access Remote imagery	fine grained features from multiple layers	Unable in detection for robust dataset
[26]	hierarchical bilinear pooling (HBP) with hierarchical attention and bilinear fusion net HABFNet	1. UC Merced dataset released in 2010 2. AID dataset released in 2017 3. NWPU-RESISC45 dataset released in 2017	Accuracy with fine grained features	Limit number images are used for the testing and training.
[68]	RTANet	**Publicly available open dataset**	**Speed**	Prediction accuracy is limited in number of images given for testing
[24]	optical remote sensing video (ORSV)	ORSV images	motion-drive	Need to be changed in terms of methodological view
[23]	Selective Search and EdgeBoxes	NWPU VHR-10	Involves high recall rate, faster	Class imbalance issues occurring

Benchmark Technique: R-CNN Deep Learning				
[27]	compatibility loss clustering method (CLCM)	1. DOTA 2.UCAS -AOD 3.NWPU VHR-10 4.RSOD-Dataset	Accuracy & efficiency	More layers can be added to enhance the accuracy of prediction.
[59]	R-CNN algorithm with dialed convolution	HRSC2016 dataset	Accuracy with respect to its feature extraction	Computationally expensive
[49]	CFEM, A context-based feature enhancement module	ISPRS Vaihingen data set	Speed & Accuracy	Traditional Neural Network approach is using, need to be more enhance with respect to methodology
[25]	PTAN (A patch-based three-stage aggregation network)	1.DOTA 2.NWPU VHR-10	**Accuracy & efficiency**	Need to enhance the performance
[29]	Class attention module with multi-class segmentation network	US3D Dataset	Accuracy & efficiency	Results need to be improved on various parameters
Benchmark Technique: Machine Learning				
[5]	**Blackboard model**	Expert systems for image processing	knowledge representation	Real time detection is not possible
[6]	Multi Expert System for Scene Interpretation and Evaluation (MESSIE)	Suburban images	Class of an object from general structure	Multiple spatial scales and aspect ratio
[7]	automated algorithm for feature extraction	Images of Ottawa area are used through SAR imagery	Detection of homogeneous areas	Limited data is used
[8]	A correlation mechanism for bi-temporal band pairs	Multi-temporal Landsat TM data	water reflectance	Traditional approach is used

Reference	Algorithm/ Model Methodology	Data Set	Parameter	Research Gap
[9]	A multi-valued recognition system	IRS satellite Imagery	Multiple class selection	Not showing more accuracy for multiple sub classes data
[10]	CADCM	HYDICE Hyperspectral digital imagery	Hidden targets in hyperspectral images	Only working for hidden targets, also including shadows
[11]	Object space and image space-based matrices	83 images of 18 sites	Performance evaluation	Limited data is tested
[12]	An automatic recognition system	Various open access databases	minimizing the deleterious effects of cloud	Not tested on a valid dataset
[13]	digital surface model (DSM)	Imagery of airborne light and ranging (LIDAR)	building footprint extraction	Not able to fine-grained the results due to limited classes used
[15]	A segmentation algorithm	GIS data	efficiency of the algorithms	Need to enhance the performance
[36]	SE-MGMM	synthetic aperture radar images	change detection	System is not obtaining a good accuracy for large amount of data or on live images.
[16]	Multistage model for road detection	SPOT multispectral images of district near Hongqiao Airport	improved detection probability	Model is showing multiple areas of Class imbalance
[17]	Two-step approach	LIDAR aerial image	Segmentation using weighted features	multiple scales are needed to be applied
[18]	A model to detect bridges over water bodies	IRS-1C/1-D satellite images of 23.5.	multispectral imagery	Dual priorities

[19]	Kolmogorov-Smirnov	TerraSAR-X (TSX) ScanSAR images (19-m resolution)	Ship Detection	Various overfitting problems involved
[20]	neighborhood model	hypothesis imagery and DIRSIG	Amorphously shaped objects	Need to be more enhanced in terms of accuracy
[55]	An algorithm for building shadow detection	panchromatic satellite imagery	Shadow detection	More classes and sub-classes can be included

5. Research Gaps & Challenges in ExistingMethods

The goal of object detection is to achieve the highest accuracy with efficiency by developing an automated robust detection algorithm. By critically reviewing the existing work done in this field it is analyzed that two major challenges still exists in finding or detec- tion of objects in remote sensing images. We have di- vided the found research gap in two categories, accu- racy and efficiency. Accuracy can affect due to various reasons like class imbalance, captured image condi- tion or environment, image noise, dual priorities, etc. Table 6 is showing the complete mapping of found re- search gap or challenges with the reviewed literature in this paper. By critically reviewing many articles it is analyzed that there is still a gap exists in finding out the optimized solution in object detection in remote sensed images.

Challenges Involved in Achieving HighAccuracy

A. **Internal class imbalance.** One of the major chal-lenges that a model faces while dealing with real ob-jects like shapes, sizes, colors, or directions of objects is class imbalance. This issue can occur due to the mentioned reasons a model could not be able to de- tect the same objects having a different shape, color, or pose in multiple images [3][7][13][21][35][26].

B. **Imagery conditions and unconstrained en- vironments.** Factors that include lighting, occlusion, weather conditions, viewpoint, object physical loca- tion, shadow clutter, blur, motion, etc. [8] [11] [41] [22] [9].

C. **Imaging noise.** Imaging noise is one of the challenges in achieving high accuracy. Also factors like compression noise, low-resolution images, and filter distortions [47] [35] [17] [43] [20].

Challenges Involved in Achieving HighEfficiency

Low computational devices like mobile have low memory and less computational speed; it is a bottle- neck in detecting objects with high efficiency [39]. Millions of unstructured and structured real-world existing object categories for distinguishing are

ISSN: 2278-4632
Vol-10 Issue-3 No.01 March 2020

a challenge for the detector [19] [23] [36]. Some- times image data especially remote images are hav- ing a large size this became a challenging situation for object detectors. To find some unseen objects is also a challenge [44] [33].

Tab. 6. Challenges Mapping with Reviewed Literature

Research Gaps & Challenges in Existing methods			
Challenges In Achieving High Accuracy			**Challenges In Achieving High Efficiency**
Internal Class Imbalance	**Imagery Conditions and Unconstrained Environments**	**Imaging Noise**	
Yao et al. [3] Ionescu et al. [7], Haithcoat et al. [13], Sevo et al. [21], Yu et al. [35], Guo et al. [26]	Coppin et al. [8], Shufelt et al. [11], Nagy et al. [41], Deng et al. [22], Mandal et al. [9]	Takarli et al. [47], Yu et al. [35], Secord et al. [17], Yang et al. [43], Grant et al. [20]	Anuta et al. [39], Paes et al. [19], Farooq et al. [23], Xue et al. [36], Tolluoglu et al. [44], Kumar et al. [33]

6. Conclusion

This paper's major objective is to provide a complete, chronological analysis of the work that has already been done in the subject of artificial intelligence, including machine learning and deep learning. This publication serves as a starting point for all aspiring researchers who want to work in this area. The published papers on object detection in remote images are simply one focus of this review paper. It also entails an evaluation of articles using a contemporary deep learning approach, such as CNN [2] [3] [4], R-CNN [6] [9] [11] [12] [15], and CornerNet [32] [35], among others. The paper offers a comprehensive overview of the ML and DL frameworks currently in use as well as the datasets currently being used for object detection. This article also discusses certain issues and difficulties with computer vision, such as crowd detection, colour imbalance, live detection, etc. This study revisits some of the most pertinent subjects, including ship detection, building detection, cloud detection, geographic item detection, and the tiniest things with the highest resolution in remote photos. In this literature study, some studies that address transfer learning ideas are also examined.

REFERENCES

[1] Q. Yao, X. Hu and H. Lei, "Multiscale Convolu- tional Neural Networks for Geospatial Object Detection in VHR Satellite Images", IEEE Geo- science and Remote Sensing Letters, vol. 18, no. 1, 2021, 23–27, 10.1109/LGRS.2020.2967819.

[2] G. Mountrakis, J. Im and C. Ogole, "Support vec- tor machines in remote sensing: A review", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, 2011, 247–259, 10.1016/j.isprsj- prs.2010.11.001.

[3] X. Yao, X. Feng, J. Han, G. Cheng and L. Guo, "Au- tomatic Weakly Supervised Object Detection From High Spatial Resolution Remote Sensing Images via Dynamic Curriculum Learning", IEEE Transactions on Geoscience and Remote Sen- sing, vol. 59, no. 1, 2021, 675–685, 10.1109/TGRS.2020.2991407.

[4] L. Li, Z. Zhou, B. Wang, L. Miao and H. Zong, "A No- vel CNN-Based Method for Accurate Ship Detec- tion in HR Optical Remote Sensing Images via Rotated Bounding Box", IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 1, 2021, 686–699, 10.1109/TGRS.2020.2995477.

[5] T. Matsuyama, "Knowledge-Based Aerial Image Understanding Systems and Expert Systems for Image Processing", IEEE Transactions on Geo- science and Remote Sensing, vol. GE-25, no. 3, 1987, 305–316, 10.1109/TGRS.1987.289802.

[6] P. Garnesson, G. Giraudon and P. Montesinos, "An image analysis, application for aerial ima- gery interpretation". In: Proc. 10th International Conference on Pattern Recognition, vol. 1, 1990, 210–212, 10.1109/ICPR.1990.118094.

[7] D. Ionescu and G. Geling, "Automatic detection of large object features from SAR data". In: Proc. IGARSS '93 - IEEE International Geoscience and Remote Sensing Symposium, 1993, 1225–1227, 10.1109/IGARSS.1993.322663.

[8] P. R. Coppin and M. E. Bauer, "Processing of multitemporal Landsat TM imagery to opti- mize extraction of forest cover change featu- res", IEEE Transactions on Geoscience and Re- mote Sensing, vol. 32, no. 4, 1994, 918–927, 10.1109/36.298020.

[9] D. P. Mandal, C. A. Murthy and S. K. Pal, "Analysis of IRS imagery for detecting man-made objects with a multivalued recognition system", IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 26, no. 2, 1996, 241–247, 10.1109/3468.485750.

[10] Hsuan Ren and Chein-I Chang, "A computer- -aided detection and classification method for concealed targets in hyperspectral imagery". In:

IGARSS '98. Sensing and Managing the Environ- ment. 1998 IEEE International

Geoscience and Remote Sensing. Symposium Proceedings. 1998, 1016–1018, 10.1109/IGARSS.1998.699658.

[11] J. A. Shufelt, “Performance evaluation and analy- sis of monocular building extraction from aerial imagery”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 4, 1999, 311–326, 10.1109/34.761262.

[12] J. G. Shanks and B. V. Shetler, “Confronting clo- uds: detection, remediation and simulation ap- proaches for hyperspectral remote sensing sys- tems”. In: Proc. 29th Applied Imagery Pattern Recognition Workshop, 2000, 25–31, 10.1109/ AIPRW.2000.953599.

[13] T. L. Haithcoat, W. Song and J. D. Hipple, “Buil- ding footprint extraction and 3-D reconstruc- tion from LIDAR data”. In: IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, 2001, 74–78, 10.1109/ DFUA.2001.985730.

[14] K. Chen and R. Blong, “Extracting building fe- atures from high resolution aerial imagery for natural hazards risk assessment”. In: IEEE In- ternational Geoscience and Remote Sensing Symposium, vol. 4, 2002, 2039–2041, 10.1109/ IGARSS.2002.1026437.

[15] J. Duan, V. Prinet and H. Lu, “Building extraction in urban areas from satellite images using GIS data as prior information”. In: Proc. IEEE Inter- national Geoscience and Remote Sensing Sym- posium, 2004. IGARSS ‘04., vol. 7, 2004, 4762– 4764, 10.1109/IGARSS.2004.1370223.

[16] Yan Dongmei, Zhao Zhongming and Chen Zhong, “A fused road detection approach in high re- solution multi-spectrum remote sensing ima- gery”. In: Proc. 2005 IEEE International Geo- science and Remote Sensing Symposium, 2005. IGARSS ‘05., vol. 3, 2005, 1557–1560, 10.1109/ IGARSS.2005.1526290.

[17] J. Secord and A. Zakhor, “Tree Detection in Urban Regions Using Aerial Lidar and Image Data”, IEEE Geoscience and Remote Sensing Letters, vol. 4, no. 2, 2007, 196–200, 10.1109/ LGRS.2006.888107.

[18] D. Chaudhuri and A. Samal, “An Automatic Brid- ge Detection Technique for Multispectral Ima- ges”, IEEE Transactions on Geoscience and Re- mote Sensing, vol. 46, no. 9, 2008, 2720–2727, 10.1109/TGRS.2008.923631.

[19] R. L. Paes, J. A. Lorenzzetti and D. F. M. Gherar- di, “Ship Detection Using TerraSAR-X Images in the Campos Basin (Brazil)”, IEEE Geoscience and Remote Sensing Letters, vol. 7, no. 3, 2010, 545–
548, 10.1109/LGRS.2010.2041322.

[20] C. S. Grant, T. K. Moon, J. H. Gunther, M. R. Stites and G. P. Williams, “Detection of Amorphously Shaped Objects Using Spatial Information Detec- tion Enhancement (SIDE)”, IEEE Journal of Se- lected Topics in Applied Earth Observations and Remote Sensing, vol. 5, no. 2, 2012, 478–487, 10.1109/JSTARS.2012.2186284.

[21] I. Sevo and A. Avramovic, “Convolutional Neu- ral Network Based Automatic Object Detection on Aerial Images”, IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 5, 2016, 740–744, 10.1109/LGRS.2016.2542358.

[22] Z. Deng, L. Lei, H. Sun, H. Zou, S. Zhou and J. Zhao, “An enhanced deep convolutional neural ne- twork for densely packed objects detection in remote sensing images”. In: 2017

Internatio- nal Workshop on Remote Sensing with Intel- ligent Processing (RSIP), 2017, 1–4, 10.1109/ RSIP.2017.7958800.

[23] A. Farooq, J. Hu and X. Jia, "Efficient object pro- posals extraction for target detection in VHR remote sensing images". In: 2017 IEEE Inter- national Geoscience and Remote Sensing Sym- posium (IGARSS), 2017, 3337–3340, 10.1109/ IGARSS.2017.8127712.

[24] Y. Li, L. Jiao, X. Tang, X. Zhang, W. Zhang and L. Gao, "Weak Moving Object Detection In Opti- cal Remote Sensing Video With Motion-Drive Fusion Network". In: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sen- sing Symposium, 2019, 5476–5479, 10.1109/IGARSS.2019.8900412.

[25] B. Sui, M. Xu and F. Gao, "Patch-Based Three-Sta- ge Aggregation Network for Object Detection in High Resolution Remote Sensing Images", IEEE Access, vol. 8, 2020, 184934–184944, 10.1109/ ACCESS.2020.3027044.

[26] J. Guo, J. Yang, H. Yue, H. Tan, C. Hou and K. Li, "CDnetV2: CNN-Based Cloud Detection for Re- mote Sensing Imagery With Cloud-Snow Coexi- stence", IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 1, 2021, 700–713, 10.1109/TGRS.2020.2991398.

[27] Y. Han, S. Ma, Y. Xu, L. He, S. Li and M. Zhu, "Ef- fective Complex Airport Object Detection in Re- mote Sensing Images Based on Improved End- -to-End Convolutional Neural Network", IEEE Access, vol. 8, 2020, 172652–172663, 10.1109/ ACCESS.2020.3021895.