An AI & ML Based Detection & Identification in Remote Imagery: State-of-the-Art

Hina Hashmi, Rakesh Dwivedi, Anil Kumar

Abstract:
Remotely sensed images and their allied areas of application have been the charm for a long time among researchers. Remote imagery has a vast area in which it is serving and achieving milestones. From the past, after the advent of AL, ML, and DL-based computing, remote imagery is related techniques for processing and analyzing are continuously growing and offering countless services like traffic surveillance, earth observation, land surveying, and other agricultural areas. As Artificial intelligence has become the charm of researchers, machine learning and deep learning have been proven as the most commonly used and highly effective techniques for object detection. AI & ML-based object segmentation & detection makes this area hot and fond to the researchers again with the opportunities of enhanced accuracy in the same. Several researchers have been proposed their works in the form of research papers to highlight the effectiveness of using remotely sensed imagery for commercial purposes. In this article, we have discussed the concept of remote imagery with some preprocessing techniques to extract hidden and fruitful information from them. Deep learning techniques applied by various researchers along with object detection, object recognition are also discussed here. This literature survey is also included a chronological review of work done related to detection and recognition using deep learning techniques.

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

1. Introduction

In this continuously growing world in the direction and advancement of technology, remote imagery is serving the world in multiple facets. Data nowadays is collected and stored in digital formats that make interpretation and analysis possible on that. Remote sensing images, images are collected through various satellites, Aerial photography, Lidar, Landsat, Spy Satellites, and sentinel images. Remote imagery is nowadays recorded in digital form so that analysis can be done on them for extracting some hidden information from them. But the limitation with the remote imagery is that the images collected from the satellites are not always up to the mark. It usually contains some kind of noise, unclear visuals, and mixed color chan-nels. So, it becomes necessary to process that data before applying any processing on them [31]. Digital image processing involves many procedures and functions to format and correct the data for segmentation and classification. With the help of those processes and techniques, the data can be used refined and it can be used for various commercial purposes like Earth observation, weather forecasting, Forestry, Agricultural usage, Surface changes, and the analysis of Bio-diversity. Remote applications are also useful for interpreting the road conditions in rural areas, identifying crop conditions, etc. For a long time, remote sensing images are being processed and analyzed with the help of deep learning techniques that is part of neural networks. Although, before the development of deep learning techniques, remote sensing imagery was being observed with the help of a support vector machine (SVM) and various ensemble classifiers like Random forest for change detection or image classification. SVM got high demand and attention due to its capability of dealing with high and multi-dimensional data with a limited amount of training data [2].

In the recent era, the emerging trends in DL have renewed the interest of the remote community for neural networks. From the near 2014, the whole remote sensing community has diverted its attention toward DL as DL techniques and algorithms have recorded its achievements in many image analysis tasks like Land cover Land use observations, object detection, scene classification, etc [30] [14] [36] [52] [35] [53] [54]. Going through the vast literature regarding DL is showing that DL is having general approaches related to the development of basic deep learning algorithms [55] and the detailed reviews for various emerging and advanced fields like medical image recognition, speech recognition techniques [56]. Some studies are showing evidence of DL in remote sensing applications [4]. The literature survey by [57] has focused on the applications of DL in remote sensing images classification for major observations. [15] has performed a complete broad review specifically focusing on related uncommon sub-areas of remote data application areas like 3-d modeling. DL algorithms and techniques are having diverse sub-domain within the remote sensing field and the application areas are continuously increasing in getting the more quantitative and systematic analysis of the data [32]. The motivation of this study is to draft a comprehensive review of DL algorithms in various areas and sub-areas of remote sensing applications that include object detection, image segmentation, scene

classification, image registration, image fusion, etc in images as well as videos [33]. By going through a vast review in the area of object detection in remote images through deep learning techniques, we have summarized the findings in various research articles. At last, a critical summary is included followed by major gaps and challenges found.

1.1 Remote Imagery and Preprocessing

Remotely sensed images are not simple images, they contain multiple formats and resolution challenges. 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 remote 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 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 measure 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 clarity. 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 wavelengths or spectrum by storing or measuring common RGB images.
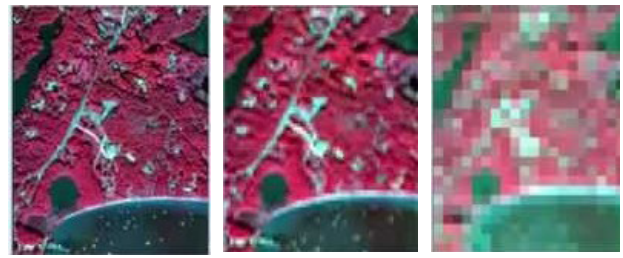
Fig. 1. (a) Spatial Resolution of 1 meter (b) 10 meter (c) 30 meter
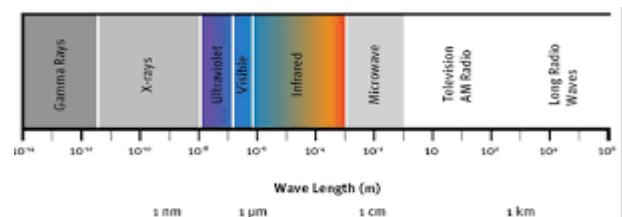
Fig. 2. Spectral Resolution of Remote Sensed Imagery

A temporal resolution is a measurement unit of any area concerning time as a movie and high-speed cameras 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-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 generally 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 instead 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 sometimes beyond the visibility range of light like- ultraviolet and infrared. Hyperspectral Resolution is generally 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 image for detection of objects, material identification, hyperspectral images are used.

As remote sensing imagery is having various resolutions 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 eliminating scene illusions, sensor noise, etc. these types of pre-processing are required for remote imagery. Var-

ious 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 image histogram is a basic concept to understand contrast enhancement.

Geometric Corrections. For correcting the geometric distortion that occurs because of Earth-sensor geometrical variations and to convert that into real-world earth-surface coordinates, geometric corrections are used. Distortion can arise due to reasons that include altitude variations, the velocity of the sensor platform, earth convexity, atmospheric diversion, 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 directional detection filters are used to identify the field boundaries or roads in an image.

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

Piecewise Linear Stretch for contrast enhancement. To utilize the complete range of value in brightness component, the maximum and minimum parameters 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 represented without enhancement, the values from 0 to 44 and the values from 206 up to 255 will not be displayed. By stretching the values from 0 to 45 and up to 205, the important features can be accessed.
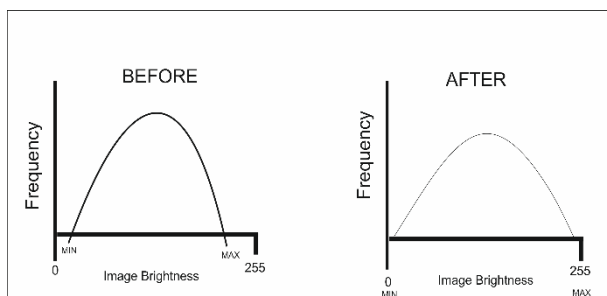

Fig. 3. Piecewise Linear Contrast Stretch

1.2 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 assistants, 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 classification [36]. If comparing deep learning with other traditional 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 labeled 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 categories like Supervised and Unsupervised. Feature extraction is one of the most important aspects of deep learning that uses an algorithm to construct meaningful features automatically for training, understanding, or learning.

From 1943 till now deep learning is being used and improving its applications day by day. Image processing through deep learning can be tracked since early 1943. In [58] created a neural network like the human 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 basic 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 processing. In 1973, Neocognitron that was an artificial neural 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 handwritten digits through the computer. In 1991, a model is created to identify the problems related to vanishing gradient. In [60] a paper is presented on deep belief networks for learning the images in a faster manner. For speech recognition, deep learning algorithms 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 science, art, and fashion.

1.3 Advancement in Image Processing Computing Vision

Over recent years, Computer Vision tasks, such as object detection, image classification especially in digital images and videos grasp a lot of attention from researchers. Computer vision is a part of Artificial Intelligence that makes the computer enables everything 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 detection is the major challenge to solve with the help of the Computer. Object detection is an emerging technology nowadays as it consists of various applications in various domains like Face Recognition, Video- Surveillance, 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 increasing its computational power and results in calculation abilities with more accuracy and reliabilities. In [37] highlighted and focused on the role of computer vision in his work. Several deep learning methods are adopted to develop deep learning models for remote sensing images; some of them include transfer learning, 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 learning, and deep learning techniques for remote sensing imagery segmentation, object detection, and classification.

In [38] presented various examples of measuring space clustering processes for a range of three multispectral 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, oceanography, 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 between digital satellite photographs and multi-spectral imagery. The registration process of space photography 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 experimental and operational usage of the upcoming torrent of raw data.

Recent trends in deep learning techniques are emerging with powerful methods for automated systems 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 power 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 prove the tested approach for various research works like virtual skinning, virtual painting, human-computer interaction, depth recovery, etc.
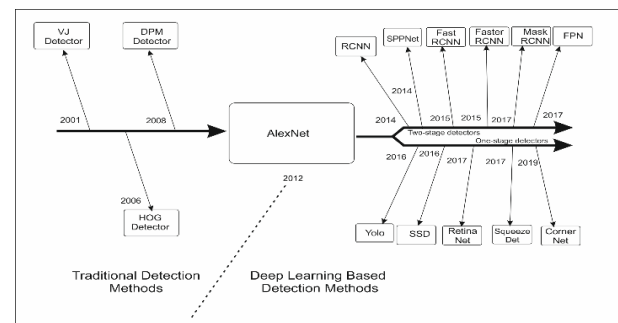

Fig. 4. Object detection milestones [21] [24] DPM, RCNN

Object Detection & Identification

In [42] detected motion of objects using cellular neural network. In [43] provided a model to detect and focus objects in an unfocused background. In [45] developed an object-oriented based segmentation algorithm for edge extraction. In [46] developed a model to detect patterns in spectrograms using cellular neural 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 model using CNN for logo detection of vehicles. In [53] provided their review for object detection algorithms for optical remote sensing imagery. They focused to review the brief history of available literature for deep learning techniques and the datasets for the detection of objects from remote imagery. Table 2 is representing 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 review 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 detector algorithms are proposed like VJ (Viola Jones),

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

Tab. 2. Differences among various open datasets (from 2012-2019) [15]

Year	Datasets	Instances	Categories	Image Width	Images	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 convolution operation in their layer [50]. CNN's convolution operation includes multiple stages consisting of four main components, kernel, convolution layer, activation function, and a pooling layer. Each stage represents a feature map in the form of an array [34]. Figure 5 is showing the detailed workflow of CNN having several convolution layers with one or more fully connected layers, for producing the final output as a classification module. The main components of CNN are described as follows.
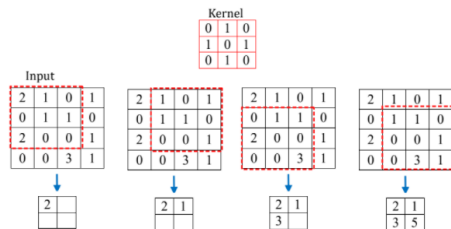
Fig. 5. Convolutional Operation with image matrix (4×4) with kernel matrix (3×3)

Kernels. At each location, each kernel is aimed to detect a particular characteristic. There exists a bank of m1 filters in every convolutional layer, the output of the lth layer consisting of feature maps of size *. The ith feature map can be computed as follows-

$$Y_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{m_1^{(l-1)}} K_{ij}^{(l)} * Y_j^{(l-1)} \tag{1}$$

Here, represents the trainable bias parameter matrix, 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 collection of multiple layers that can calculate the convolution 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 hyper-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-

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

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 input 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 applied in this process. As Eq. 3 is showing, the calculation functions of activation maps carried only the activated features to the next layer. Then, the activation function can be formed as follows

$$\emptyset(Y_i^{(l)}) = \left(B_i^{(l)} + \sum_{j=1}^{m_1^{(l-1)}} K_{ij}^{(l)} * Y_j^{(l-1)} \right) \tag{3}$$

Some activation functions that can be used in the convolution process are as follows:

Rectified Linear Unit function (ReLU). Most commonly used activation function. It is called a half-rectified function. Figure 6 (a) is showing ReLU the graphical representation of the ReLU activation function. Mathematically it can be defined as

$$f(x) = max(0, x) = \begin{cases} 0 \ if x < 0 \\ x \ if x \geq 0 \end{cases} \tag{4}$$

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 between 0 and 1. It can be defined as

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

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.

$$f(x) = tanh(x) = \frac{2}{1+e^{-2x}} - 1 \qquad (6)$$
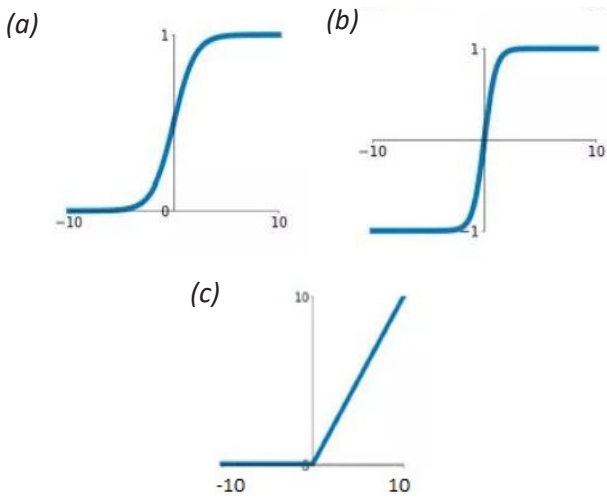
Fig. 6. Representation of various activation functions: a- ReLU b- Sigmoid, c- tanh

Normalization Layers. To implement inhibition paradigms, the normalization layer is used for the observation 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 parametric 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

$$f_{max}(A) = max_{m,n}(A_{nxm}) \qquad (7)$$

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

$$f_{ave}(A) = \frac{1}{n+m}\sum_{i=1}^{n}\sum_{k=1}^{m}(A_{i,k}) \qquad (8)$$
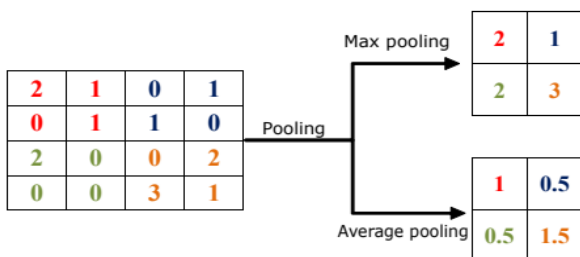
Fig. 7. Operation of Pooling using 2 x 2 filters with stride of two

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

Fully Connected Layer. At this layer, the final output is represented in the form of a 2-D array by connected 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 classification, the softmax activation function is used having a range of probability between [0, 1] with a sum equal to 1. The fully connected layer is situated at the last layer in a neural network. The fully connected layer is having the neurons that have a full connection to all the neurons of previous layers [63]. The calculations of this connection can be evaluated by matrix multiplication function later followed by bias offset [62].

The remainder of this paper is being organized as section-2 is showing the various existing pre-trained models on ImageNet and CIFAR-10 datasets, section-3 is giving the inceptive contributions at beginning of remote senses image processing, section-4 is the detailed literature review as a tabulated formation categorized by various technological benchmarks used, section-5 shows the challenges mapping with reviewed literature and section 6 involves the complete summary.

2. Existing Pre-trained Deep Leaning Models and 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 existing 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 classification 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 dataset

Model	Depth	Size (MB)	Parameters (Million)	Top-1 Accuracy	Top-5 Accuracy
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%

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 interpretation and evaluation of scenes for image analysis.

In paper [5] authors proposed a blackboard model 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 automated 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 detection model named CADCM to target hidden objects in hyperspectral imagery. The process is accomplished in three phases. Initially, a band selection process is applied, 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 image 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 analysis, the effects of image inscrutability along with objects complexity are examined. Edge fragmentation is also shown here in this research. The usage of rigorous photogrammetric space modeling is also demonstrated.

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

In [13] an automatic approach for building footprint 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, geometric 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 comparing the results with manually digitized building reference 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 features of the building. GIS data is used to provide prior building knowledge. Data pre-processing, Object segmentation, 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-cover types. A knowledge-based approach is used that find out the spatial arrangement of the bridge and its surroundings. Testing is done on the IRS-1C/1-D satellite has a spatial resolution of 23.5m.

In [19] authors investigated ship detection in TerraSAR-X (TSX) ScanSAR images (19-m resolution). Kolmogorov-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 verified.

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

In [21] an automatic content-based analysis is presented to detect arbitrary objects in aerial imagery. In this, the two-stage training model using a convolutional 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 enhanced 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 generated by multiple-scale layers is combined. VHR-10 dataset 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-resolution remote imagery. Algorithms are tested and evaluated through the NWPU-VHR-10 class data set. For performance measurement, execution time and recall rate are used as performance parameters. By the statistical results, authors proved that EdgeBoxes algorithm is showing optimal results over Selective Search in recall 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 extracting 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 connection are used for fusion. After extracting the features 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 training 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 predict 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). Proposed 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 accurate 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 available 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-sampling 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-assembled datasets SDOTA, SDD to show the excellent performance of the proposed method by comparing it with other detectors. The proposed method is achieving 74.3% mAP on the public DIOR dataset.

In [1] authors proposed EFPN to detect small objects like plants, small buildings, etc. It is created to improve 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 object detection using deep learning and machine learning. 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 areas, 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	DataSet	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	DataSet	Parameter	Research Gap
[49]	object relationship reasoning CNN (ORRCNN)	Aerial Image data set (AID), UC Merced Land-Use data set, and WHU-RS19 data set	**accuracy for multiband data**	Need more prior information and parameters of the geometrical shape 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	DataSet	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 Existing Methods

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 detection of objects in remote sensing images. We have divided the found research gap in two categories, accuracy and efficiency. Accuracy can affect due to various reasons like class imbalance, captured image condition or environment, image noise, dual priorities, etc. Table 6 is showing the complete mapping of found research 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.

5.1 Challenges Involved in Achieving High Accuracy

A. **Internal class imbalance.** One of the major challenges 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 detect the same objects having a different shape, color, or pose in multiple images [3][7][13][21][35][26].

B. **Imagery conditions and unconstrained environments.** Factors that include lighting, occlusion, weather conditions, viewpoint, object physical location, 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].

5.2 Challenges Involved in Achieving High Efficiency

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

a challenge for the detector [19] [23] [36]. Sometimes image data especially remote images are having 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 is mainly focusing on a thorough, comprehensive, and chronological study of existing work done in the field of AI including machine learning and deep learning. This paper is performing the role of a pathway to all new researchers who are wishing to work in this field. This review paper is not only focusing on the published papers on object detection in remote imagery. It also involves the review of the papers those are having a modern approach of deep learning like CNN [2] [3] [4], R-CNN [6] [9] [11] [12] [15] and CornerNet [32] [35], etc. The paper is proving a complete outlook on existing ML and DL frameworks, and the existing datasets used in object detection. Some problems and challenges in computing vision are also discussed here like crowd detection, color imbalance, live detection, etc. This paper is reviewing some most relatable topics like Ship detection, building detection, geospatial object detection, cloud detection, small-scale objects with the highest resolution in remote images. Some papers which involve transfer learning concepts are also reviewed in this literature review.

AUTHORS

Hina Hashmi* – CCSIT, Teerthanker Mahaveer University, Moradabad, India, e-mail: hinahashmi170@gmail.com,

Rakesh Dwivedi – CCSIT, Teerthanker Mahaveer University, Moradabad, India,

Anil Kumar – Indian Institute of Remote Sensing, Dehradun, India.

* Corresponding author

REFERENCES

[1] Q. Yao, X. Hu and H. Lei, "Multiscale Convolutional Neural Networks for Geospatial Object Detection in VHR Satellite Images", IEEE Geoscience 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 vector machines in remote sensing: A review", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, 2011, 247–259, 10.1016/j.isprsjprs.2010.11.001.

[3] X. Yao, X. Feng, J. Han, G. Cheng and L. Guo, "Automatic Weakly Supervised Object Detection From High Spatial Resolution Remote Sensing Images via Dynamic Curriculum Learning", IEEE Transactions on Geoscience and Remote Sensing, 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 Novel CNN-Based Method for Accurate Ship Detection 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 Geoscience 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 imagery 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 optimize extraction of forest cover change features", IEEE Transactions on Geoscience and Remote 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 Environment. 1998 IEEE International Geoscience and Remote Sensing. Symposium Proceedings. 1998, 1016–1018, 10.1109/IGARSS.1998.699658.

[11] J. A. Shufelt, "Performance evaluation and analysis 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 clouds: detection, remediation and simulation approaches for hyperspectral remote sensing systems". 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, "Building footprint extraction and 3-D reconstruction 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 features from high resolution aerial imagery for natural hazards risk assessment". In: IEEE International 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 International Geoscience and Remote Sensing Symposium, 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 resolution multi-spectrum remote sensing imagery". In: Proc. 2005 IEEE International Geoscience 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 Bridge Detection Technique for Multispectral Images", IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 9, 2008, 2720–2727, 10.1109/TGRS.2008.923631.

[19] R. L. Paes, J. A. Lorenzzetti and D. F. M. Gherardi, "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 Detection Enhancement (SIDE)", IEEE Journal of Selected 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 Neural 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 network for densely packed objects detection in remote sensing images". In: 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), 2017, 1–4, 10.1109/ RSIP.2017.7958800.

[23] A. Farooq, J. Hu and X. Jia, "Efficient object proposals extraction for target detection in VHR remote sensing images". In: 2017 IEEE International Geoscience and Remote Sensing Symposium (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 Optical Remote Sensing Video With Motion-Drive Fusion Network". In: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, 5476–5479, 10.1109/ IGARSS.2019.8900412.

[25] B. Sui, M. Xu and F. Gao, "Patch-Based Three-Stage 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 Remote Sensing Imagery With Cloud-Snow Coexistence", 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, "Effective Complex Airport Object Detection in Remote Sensing Images Based on Improved End-to-End Convolutional Neural Network", IEEE Access, vol. 8, 2020, 172652–172663, 10.1109/ ACCESS.2020.3021895.

[28] J. Yang, J. Guo, H. Yue, Z. Liu, H. Hu and K. Li, "CDnet: CNN-Based Cloud Detection for Remote Sensing Imagery", IEEE Transactions on Geo-

science and Remote Sensing, vol. 57, no. 8, 2019, 6195–6211, 10.1109/TGRS.2019.2904868.

[29] Z. Rao, M. He and Y. Dai, "Class Attention Network for Semantic Segmentation of Remote Sensing Images", 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2020, 150–155.

[30] P. Sharma, N. Ding, S. Goodman and R. Soricut, "Conceptual Captions: A Cleaned, Hypernymed, Image Alt-text Dataset for Automatic Image Captioning". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, 2556–2565, 10.18653/v1/P18-1238.

[31] A. Sisodia, Swati and H. Hashmi, "Incorporation of Non-Fictional Applications in Wireless Sensor Networks", International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 11, 2020, 42–49, 10.35940/ijitee. K7673.0991120.

[32] H. Hashmi, R. K. Dwivedi and A. Kumar, "Identification of Objects using AI & ML Approaches: State-of-the-Art". In: 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), 2021, 1–5, 10.1109/ SMART52563.2021.9676273.

[33] A. Kumar, H. Hashmi, S. A. Khan and S. Kazim Naqvi, "SSE: A Smart Framework for Live Video Streaming based Alerting System". In: 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), 2021, 193–197, 10.1109/ SMART52563.2021.9675306.

[34] S. Belongie, J. Malik and J. Puzicha, "Shape matching and object recognition using shape contexts", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, 2002, 509–522, 10.1109/34.993558.

[35] D. Yu, H. Guo, Q. Xu, J. Lu, C. Zhao and Y. Lin, "Hierarchical Attention and Bilinear Fusion for Remote Sensing Image Scene Classification", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, 2020, 6372–6383, 10.1109/JSTARS.2020.3030257.

[36] D. Xue, T. Lei, X. Jia, X. Wang, T. Chen and A. K. Nandi, "Unsupervised Change Detection Using Multiscale and Multiresolution Gaussian-Mixture--Model Guided by Saliency Enhancement", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, 2021, 1796–1809, 10.1109/JSTARS.2020.3046838.

[37] P. Peer and B. Batagelj, "Computer vision in contemporary art (introduction to the special session)", 2008 50th International Symposium ELMAR, vol. 2, 2008, 471–474.

[38] R. M. Haralick and G. L. Kelly, "Pattern recognition with measurement space and spatial clustering for multiple images", Proc. of the IEEE, vol. 57, no. 4, 1969, 654–665, 10.1109/PROC.1969.7020.

[39] P. E. Anuta, "Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques", IEEE Transactions on Geoscience Electronics, vol. 8, no. 4, 1970, 353–368, 10.1109/TGE.1970.271435.

[40] W. P. Waite and H. C. MacDonald, ""Vegetation Penetration" with K-Band Imaging Radars", IEEE Transactions on Geoscience Electronics, vol. 9, no. 3, 1971, 147–155, 10.1109/ TGE.1971.271487.

[41] G. Nagy, "Digital image-processing activities in remote sensing for earth resources", Proc. of the IEEE, vol. 60, no. 10, 1972, 1177–1200, 10.1109/ PROC.1972.8879.

[42] T. Roska, T. Boros, P. Thiran and L. O. Chua, "Detecting simple motion using cellular neural networks". In: IEEE International Workshop on Cellular Neural Networks and their Applications, 1990, 127–138, 10.1109/CNNA.1990.207516.

[43] Xiou-Ping Yang, Tao Yang and Lin-Bao Yang, "Extracting focused object from defocused background using cellular neural networks". In: Proc. of the Third IEEE International Workshop on Cellular Neural Networks and Their Applications (CNNA-94), 1994, 451–455, 10.1109/ CNNA.1994.381633.

[44] O. Tolluoglu, O. M. Ucan, S. Kent and S. Kargin, "Edge detection in noise mounted picture". In: *Proc. of the IEEE 12th Signal Processing and Communications Applications Conference,*, 2004, 335–338, 10.1109/SIU.2004.1338328.

[45] G. Grassi, E. Di Sciascio, A. L. Grieco and P. Vecchio, "A New Object-Oriented Segmentation Algorithm based on CNNs - Part I: Edge Extraction". In: *2005 9th International Workshop on Cellular Neural Networks and Their Applications*, 2005, 158–161, 10.1109/CNNA.2005.1543185.

[46] K. Slot, P. Korbel, M. Gozdzik and H. Kim, "Pattern detection in spectrograms by means of Cellular Neural Networks". In: *2006 10th International Workshop on Cellular Neural Networks and Their Applications*, 2006, 1–6, 10.1109/ CNNA.2006.341632.

[47] F. Takarli, A. Aghagolzadeh and H. Seyedarabi, "Robust pedestrian detection using low level and high level features". In: *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, 2013, 1–6, 10.1109/IranianCEE.2013.6599574.

[48] W. Thubsaeng, A. Kawewong and K. Patanukhom, "Vehicle logo detection using convolutional neural network and pyramid of histogram of oriented gradients". In: *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2014, 34–39, 10.1109/JCSSE.2014.6841838.

[49] Z. Li, Q. Wu, B. Cheng, L. Cao and H. Yang, "Remote Sensing Image Scene Classification Based on Object Relationship Reasoning CNN", *IEEE Geoscience and Remote Sensing Letters*, vol. 19, 2022, 1–5, 10.1109/LGRS.2020.3017542.

[50] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", Proc. of the IEEE, vol. 86, 1998, 2278–2324, 10.1109/5.726791.

[51] A. Femin and K. S. Biju, "Accurate Detection of Buildings from Satellite Images using CNN". In: 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 2020, 1-5, 10.1109/ICECCE49384.2020.9179232.

[52] D. Marmanis, J. D. Wegner, S. Galliani, K. Schindler, M. Datcu and U. Stilla, "Semantic Segmentation of Aerial Images With an Ensemble of CNNs", ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. III-3, 2016, 473–480, 10.5194/isprsannals-III-3-473-2016.

[53] L. Cheng, K. E. Trenberth, M. D. Palmer, J. Zhu and J. P. Abraham, "Observed and simulated full-depth ocean heat-content changes for 1970–2005", Ocean Science, vol. 12, no. 4, 2016, 925–935, 10.5194/os-12-925-2016.

[54] N. Kussul, M. Lavreniuk, S. Skakun and A. Shelestov, "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data", IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, 2017, 778–782, 10.1109/LGRS.2017.2681128.

[55] M. I. Elbakary and K. M. Iftekharuddin, "Shadow Detection of Man-Made Buildings in High-Resolution Panchromatic Satellite Images", IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 9, 2014, 5374–5386, 10.1109/TGRS.2013.2288500.

[56] X. Zhang, S. Noda, R. Himeno and H. Liu, "Cardiovascular disease-induced thermal responses during passive heat stress: an integrated computational study: An integrated bioheat transfer model", International Journal for Numerical Methods in Biomedical Engineering, vol. 32, no. 11, 2016, 5374–5386, 10.1002/cnm.2768.

[57] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu and M. Pietikäinen, "Deep Learning for Generic Object Detection: A Survey", International Journal of Computer Vision, vol. 128, no. 2, 2020, 261–318, 10.1007/s11263-019-01247-4.

[58] W. S. McCulloch and W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity", The Bulletin of Mathematical Biophysics, vol. 5, no. 4, 1943, 115–133, 10.1007/BF02478259.

[59] Y.-C. Hsieh, C.-L. Chin, C.-S. Wei, I.-M. Chen, P.-Y. Yeh and R.-J. Tseng, "Combining VGG16, Mask R-CNN and Inception V3 to identify the benign and malignant of breast microcalcification clusters". In: 2020 International Conference on Fuzzy Theory and Its Applications (iFUZZY), 2020, 2022–01-04, 10.1109/iFUZZY50310.2020.9297809.

[60] G. E. Hinton, S. Osindero and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets", Neural Computation, vol. 18, no. 7, 2006, 1527–1554, 10.1162/neco.2006.18.7.1527.

[61] V. Dumoulin, I. Goodfellow, A. Courville and Y. Bengio, "On the Challenges of Physical Implementations of RBMs", Proceedings of the AAAI Conference on Artificial Intelligence, vol. 28, no. 1, 2014, 10.1609/aaai.v28i1.8924.

[62] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge", International Journal of Computer Vision, vol. 115, no. 3, 2015, 211–252, 10.1007/s11263-015-0816-y.

[63] G. M. Castelluccio and D. L. McDowell, "Microstructure and mesh sensitivities of mesoscale surrogate driving force measures for transgranular fatigue cracks in polycrystals", Materials Science and Engineering: A, vol. 639, 2015, 626–639, 10.1016/j.msea.2015.05.048.

[64] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", Communications of the ACM, vol. 60, no. 6, 2017, 84–90, 10.1145/3065386.

[65] L. Tian and A. Noore, "Software Reliability Prediction Using Recurrent Neural Network with Bayesian Regularization", International Journal of Neural Systems, vol. 14, no. 3, 2004, 165–174, 10.1142/S0129065704001966.

[66] A. Benamira, B. Devillers, E. Lesot, A. K. Ray, M. Saadi and F. D. Malliaros, "Semi-Supervised Learning and Graph Neural Networks for Fake News Detection". In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2019, 568–569, 10.1145/3341161.3342958.

[67] X. Zhou, M. W. Koch and M. W. Roberts, "A Selective Attention Neural Network for Invariant Recognition of Distorted Objects". In: IJCNN--91-Seattle International Joint Conference on Neural Networks, vol. ii, 1991, 10.1109/IJCNN.1991.155543.

[68] L. Fu, D. Zhang and Q. Ye, "Recurrent Thrifty Attention Network for Remote Sensing Scene Recognition", IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 10, 2021, 8257–8268, 10.1109/TGRS.2020.3042507.

[69] R. Larionov, V. Khryashchev and V. Pavlov, "Separation of Closely Located Buildings on Aerial Images Using U-Net Neural Network". In: 2020 26th Conference of Open Innovations Association (FRUCT), 2020, 256–261, 10.23919/FRUCT48808.2020.9087365.