iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60544

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

An Analysis: Progress in Identifying Objects, Classification and Segmenting them in Satellite Images

Umang Bagadi, Yash Dahake, Tushar Bhogekar, Vedant Patil, Prof. Minakshi Ramteke
GH Raisoni College of Engineering Nagpur

Abstract: ***Recognizing and differentiating items in satellite pictures is crucial for different goals, such as city planning and disaster response. Recent developments in machine learning, particularly deep learning, and the abundance of high-quality satellite images have significantly boosted progress in this area. CNNs have become essential for precisely identifying and categorising elements depicted in satellite images. These models use convolution and pooling methods on a local scale to identify local patterns and far-reaching connections, ultimately enhancing object detection and segmentation accuracy. This review offers a detailed summary of the advanced methods used to identify and separate objects in satellite images. The article discusses the importance of precise object detection and segmentation in today's dynamic environment, showcasing the progress achieved through deep learning methodologies. The article explores the difficulties of pinpointing key and distinct areas in satellite pictures and shows how deep learning models tackle this issue by detecting the connections between objects and various facets of the surroundings. This research assesses the benefits and drawbacks of various methods in terms of accuracy, computational speed, and scalability by comparing conventional techniques with deep learning strategies. Additionally, it addresses important obstacles like limited data and the necessity for easily understandable models, as well as potential future research paths like incorporating more data sources and improving model generalisation skills. This review paper offers a thorough reference for researchers and practitioners studying object detection and segmentation in satellite imagery, giving information on recent progress, obstacles, and upcoming areas of focus in this quickly developing area. Identifying and separating objects in satellite images is crucial for a range of purposes, such as urban development and emergency aid. Access to high-quality satellite images has enhanced the advancement of machine learning, particularly deep learning.***

Keywords: Deep Learning, Object Detection, Segmentation, Convolutional Neural Network(CNN), Classification

I. INTRODUCTION

High-resolution satellite images are necessary in a wide range of areas, from urban planning to disaster response, in our ever-changing world. The advancement of machine learning, particularly deep learning, has significantly propelled advancements in the field of detecting and delineating objects in satellite images. CNNs are crucial technologies for precisely identifying and categorising objects in intricate satellite images. This review article aims to provide a thorough examination of the most recent methods employed in object detection and segmentation on satellite images. The importance of these methods is highlighted by their application in addressing contemporary issues like urban growth, environmental surveillance, and emergency preparedness. Having access to high-resolution satellite images has been crucial for carrying out more precise research on the Earth's terrain. The examination delves into the intricacies of CNNs, elucidating how these models make use of local operations like convolution and pooling. These functions improve the precision of object detection and segmentation by capturing both local patterns and long-range dependencies more accurately. The article outlines the rapid advancements in deep learning techniques and their specific contributions to improving the analysis of satellite images. The review focuses primarily on the difficulties of extracting distinct and significant regions from satellite images. As previously mentioned, deep learning models tackle these issues by efficiently capturing the relationships between objects and other elements in the scenes. The paper examines traditional methods versus deep learning approaches, assessing their strengths and weaknesses in terms of factors like accuracy, computational efficiency, and scalability.
Furthermore, the evaluation acknowledges and explores major obstacles facing this area, such as limited data and the need for models that can be easily understood. It also examines possible future research paths, highlighting the combination of satellite image analysis with other data sources and the improvement of model generalisation abilities.

©IJRASET: All Rights are Reserved

Essentially, this review provides an extensive guide for researchers and practitioners involved in object detection and segmentation on satellite imagery. It offers information on recent developments and tackles obstacles and future trends in the quickly changing field, promoting a better grasp of current methods' strengths and weaknesses.

II. RELATED WORK

A. *DeepSat – An Educational Platform for Analysing Satellite Images*

The article introduces DeepSat, a custom learning system created for identifying categories of satellite pictures. The text highlights the significance of Deep Learning, particularly the effectiveness of Deep Belief Networks (DBNs) and Convolutional Neural Networks (CNNs) in tasks such as object recognition and acoustic signal modelling. Due to the large and diverse set of data in satellite image classification, conventional supervised learning techniques encounter difficulties, highlighting the need for new approaches. DeepSat addresses this issue by combining handcrafted features with a DBNs-based unsupervised learning system that can extract patterns from large amounts of unlabeled data. Limited research in satellite image classification is primarily due to the scarcity of labelled datasets. In order to address this gap, the paper presents two labelled datasets, SAT-4 and SAT-6, covering a wide area to facilitate advanced research in classifying aerial images. The proposed framework shows exceptional accuracy on both datasets, outperforming state-of-the-art algorithms by significant margins. Additionally, statistical analysis offers more proof of the effectiveness of DeepSat's approach in capturing features to represent satellite information. Overall, the contributions involve supplying labelled datasets and developing an effective classification framework for satellite imagery, demonstrating superior performance compared to existing methods..

The DeepSat framework uses SAT-4 and SAT-6 labelled datasets for training and testing purposes. SAT-4 contains 500,000 image patches that have been resized to 28x28 pixels and sorted into four categories based on land cover: barren land, trees, grassland, and a miscellaneous group. 400,000 patches were allocated for training purposes, while 100,000 were reserved for testing, ensuring distinct datasets for each. Similarly, SAT-6 consists of 405,000 image patches that have been resized to 28x28 pixels and represent six distinct land cover types: arid land, forests, meadows, roads, structures, and water masses. 324,000 patches were reserved for training, with the remaining 81,000 designated for testing, selected from different NAIP tiles. Both data sets were randomized before being used for training and testing to maintain data integrity and ensure unbiased evaluation.

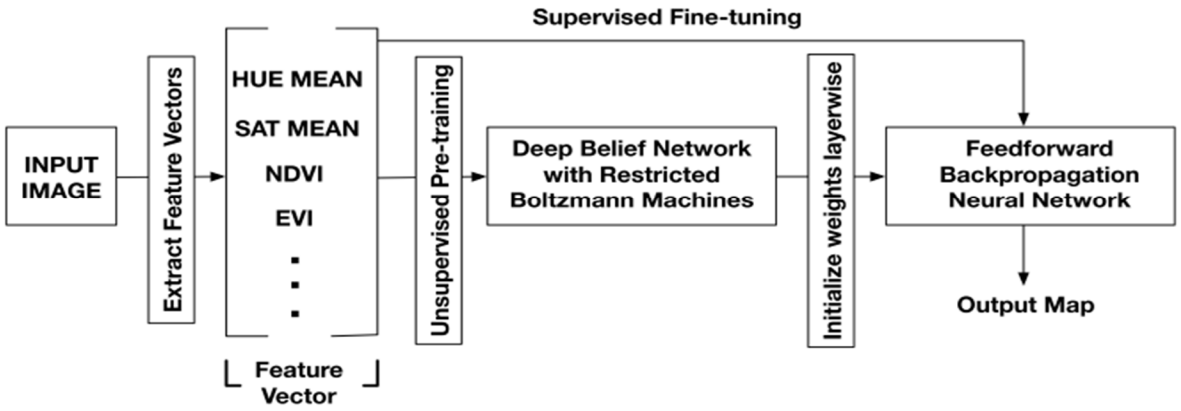


Fig 1. Schematic of the DeepSat classification framework

In conclusion, DeepSat presents a unique method for classifying satellite photos by employing advanced deep learning techniques, specifically Deep Belief Networks (DBNs), to achieve high levels of accuracy. With smaller network sizes, DeepSat allows for faster attainment of high accuracy by utilising feature vectors with fewer dimensions, significantly speeding up the training process. DeepSat outperforms traditional deep learning techniques such as CNNs, with a 11% increase in accuracy for the SAT-4 dataset and a 15% increase for the SAT-6 dataset.

Furthermore, when compared to a Random Forest classifier, the benefits of unsupervised pre-training in DBNs are emphasised over standard supervised learning techniques, as DeepSat consistently achieves superior classification accuracy. These results highlight DeepSat's efficiency in classifying satellite images and its ability to enhance research in this area, providing valuable insights and practical implications for real-life use.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. A Strategy For Identifying Constructions From High-Quality Satellite Images Using Cognitive Techniques

The paper introduces a cognitive approach to identifying buildings in high-resolution satellite (HRS) images, focusing on the active research area of computer vision and its use in change detection and urban monitoring. Different methods for detecting buildings have been examined, classifying them according to data sources like SAR, multispectral, and LiDAR images. Approaches include using shadow data, supervised classification algorithms, support vector machines, graph theory tools, and fuzzy logic.

The suggested cognitive approach integrates building identification with cognitive task analysis (CTA), a hierarchical technique employed in psychological studies to comprehend decision-making processes. CTA entails examining the cognitive processes and expertise utilised by an analyst in carrying out intricate tasks, simplifying the task of extracting information by utilising human cognitive functions. The research explains how CTA is used in identifying buildings, detailing the various input parameters and cognitive abilities required for the task.

CTA's main benefit lies in its capability to produce detailed and accurate data on cognitive processes, offering a methodical way to comprehend intricate tasks. Nevertheless, analysing and confirming gathered data demands time. The methodology consists of identifying the task flow, gaining insight into the knowledge domain, and obtaining knowledge using methods like observation, document analysis, and interviews. The result of this stage is utilised to recognize the necessary structures and types of knowledge needed for constructing detection.

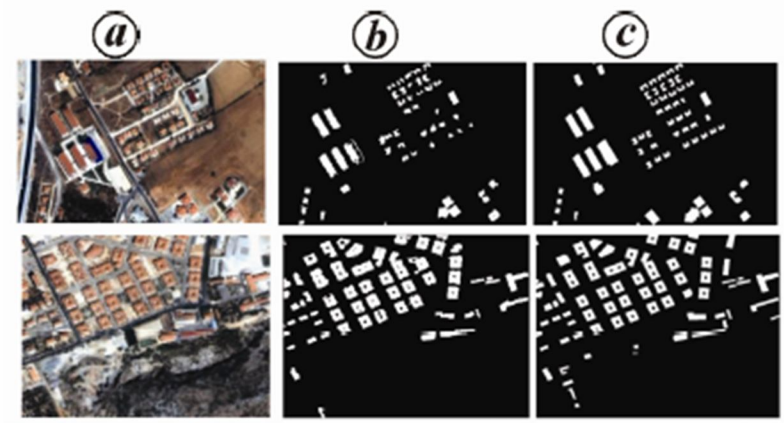


Fig. 2 Results from the building detection technique include: a set of test images, a collection of identified buildings, and their respective ground truth data.

During this stage of the research, a thorough investigation is carried out on all tasks needed for analysing satellite images cognitively, in order to pinpoint the essential knowledge and sub-tasks. In contrast to prior research using semantic networks or flow charts, this study uses a rule-based method for knowledge representation. In particular, a training dataset is carefully created where human interpreters manually label objects in satellite images as either buildings or non-buildings. These explanations are then used as guidelines for future object detection procedures.

After that, the collected data is consolidated through various methods like simulation, prototyping, and observation. These techniques facilitate the collection of crucial data necessary for tackling difficult problems. CTA is a major advancement in cognitive psychology that enables analysts to create descriptive data from task performance. Through the use of CTA, we acquire crucial understanding about the cognitive processes necessary for task performance.

Moreover, the study introduces a method for identifying structures that deals with concerns associated with variations in building characteristics such as dimensions, form, hue, and compactness, which are frequently observed in city areas. However, it is acknowledged that this method may not differentiate effectively between regions lacking structures and regions containing buildings with comparable spectral features. Future studies should seek to confirm this cognitive method on a bigger geographic scope and with a more diverse range of test images to enhance precision.

C. Comparison of Various Deep Learning Models for Detecting Objects in Satellite Images

Object detection, also referred to as automated target recognition (ATR), plays a vital role in extracting intelligence from multi-spectral satellite imagery, assisting in a range of applications like broad area search, persistent site monitoring, disaster relief, and beyond.

Deep learning, specifically convolutional neural networks (CNNs), has become a crucial technology for improving the precision of object detection in both electro-optical (EO) and synthetic aperture radar (SAR) images. Historically, two-stage methods have been preferred for their strong precision, which includes identifying potential objects first and then enhancing classification and localization through a CNN. Nevertheless, there is an increasing interest in single-stage models that can conduct detection and classification at the same time, thanks to their efficiency and similar level of accuracy when compared to two-stage models. Moreover, object detection is achieved through the combination of different methods in multi-stage models. This study compares various advanced deep learning models for object detection, such as RetinaNet, GHM, Faster R-CNN, Grid R-CNN, Double-Head R-CNN, and Cascade R-CNN. Various CNN architectures, like ResNet-50, ResNet-101, and ResNeXt-101, are utilised to extract features. The research assesses how precise and quick these models are when using commercial EO satellite images, by comparing their speed with the sliding window algorithm for baseline evaluation. The results help in choosing appropriate models for practical uses and support the development of automated target recognition systems, like the Lockheed Martin Globally-Scalable Automated Target Recognition (GATR) framework.

Precise and prompt identification of objects in satellite images is essential for a variety of Earth science disciplines. Because of the fast growth in data volume, there is a growing need for automated detection algorithms to enhance speed and precision. Quickly identifying and tracking objects in high-definition video is crucial for autonomous vehicles to react promptly and prevent accidents. This has resulted in thorough research and development activities, mainly utilizing deep learning methods, to improve machine understanding at the surface. Nevertheless, employing these methods on satellite images comes with its own set of obstacles. Satellite images usually show objects from above, but off-nadir angles can cause variations unlike the ground-level views of autonomous vehicles. Furthermore, satellite imagery has a larger size and data volume when compared to ground-based systems. The WorldView-3 satellite from DigitalGlobe has the ability to take in panchromatic (31cm) images covering 680,000 km2 every day, with each image being over 10 Gigapixels. Furthermore, the sizes of objects in satellite images may differ greatly and could be hidden by clouds or affected by weather conditions. Utilizing specialized techniques is crucial in successfully overcoming the challenges of identifying items in satellite images due to their complexity.

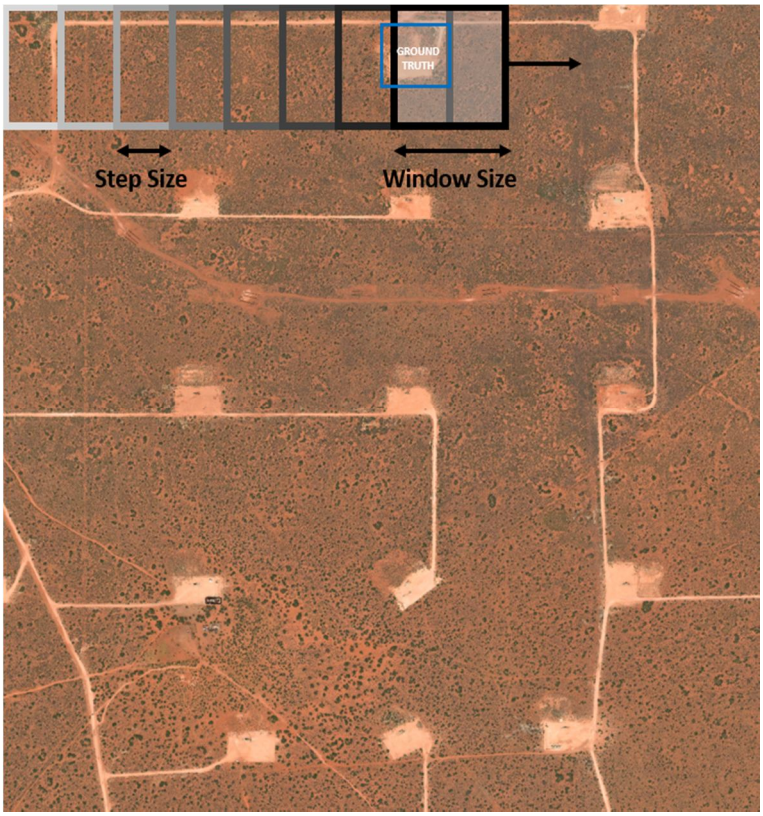


Fig. 3 Diagram illustrating the sliding window method for object detection. Box that is blue represents an assumed true boundary box. Picture: DigitalGlobe, Inc

The sliding window algorithm depicted in Fig. 2 is seen as a simple method for automated target recognition (ATR), but its efficacy is accompanied by drawbacks. Choosing the correct window size for the target is essential for the best performance, even though it may seem simpler to implement at first. Furthermore, the algorithm faces difficulties when there is a notable difference in object sizes, which necessitates the training and utilization of multiple sliding window models simultaneously for precise detection. Not meeting these criteria can result in lower performance on tasks involving detecting objects.

This research evaluated how quickly and effectively current algorithms can detect oil and gas fracking wells and small cars in satellite image datasets used for commercial Earth observation. For fracking wells, most models performed well, with single-stage models proving more effective and efficient. However, two-stage and multi-stage models outperformed single-stage models in detecting small cars, possibly due to different levels of training data accessibility. The performance metrics were barely impacted by the choice of CNN backbone, however, there was a noticeable change in inference speed. Further studies will expand the examination by including additional targets and detection models like the YOLO family, and explore multi-scale approaches for object detection to enhance area coverage rates while maintaining performance. Additionally, the study aims to investigate how accurately and quickly multiple classes of objects can be detected, despite challenges posed by variations in target sizes and characteristics across the classes.

D. Improving Images and Identifying Characteristics in Low-Quality Satellite Images

The research presented a structure that includes image enhancements such as linear and decorrelation stretching to enhance the visual examination of images and streamline the process. They utilised supervised learning, particularly support vector machines (SVMs), with a low-resolution reference layer to decrease ambiguity and improve the accuracy of the reference layer. The process included training a Support Vector Machine, determining the best hyperplane for classifying built-up/non built-up areas, adjusting class labels, and using histogram adjustments for computing textural/morphological features based on contrast.

One important factor was the necessity of a uniform and automated method for standardising uncalibrated images. The research evaluated different image enhancement procedures and examined their effectiveness using Receiver Operating Characteristic (ROC) analysis with existing building footprints. The findings showed that grayscale image contrast was significantly affected by the combination of spectral bands, and adjusting contrast was important for extracting features, particularly in images with low contrast. The project's goal was to discover efficient methods to maintain image contrast sensitivity across various preprocessing conditions.

The study in Torino, Italy, utilised SVMs, Haralick's PANTEX for textural measurements, and the morphological building index (MBI) for morphological features in the experimental setup. The research investigated various methods for creating alternative reference layers and adjusting contrast. The findings showed that changes in contrast had a stronger effect on textural measurements compared to morphological characteristics. Moreover, the selection of band combinations had a notable impact on the quality of feature extraction.

In conclusion, the project demonstrated the importance of contrast adjustment for precise feature extraction from high-resolution satellite images. The proposed approach, combining machine learning with image processing, showed promise in improving image contrast sensitivity. Future studies could explore adaptive histogram equalization, different reference layers, and radiometric feature selection to enhance the methodology.

E. Utilising LVQ Neural Network for Automated Extraction of Features from Satellite Images

The research is centered on extracting features automatically from satellite images, with a specific emphasis on the importance of feature extraction in spatial data management for updating GIS databases. Manual interpretation and automatic classification methods have constraints, prompting an increased focus on semi-automatic techniques. Nevertheless, there is a gap in investigating the efficiency of completely automated techniques because of the intricate nature of satellite images and configuring neural networks. The article introduces a neural network system designed to efficiently extract features from high-resolution satellite images, with a specific focus on obtaining data related to land use and land cover. The method chosen involves using a self-organising supervised learning Artificial Neural Network (ANN) with a Learning Vector Quantization (LVQ) approach. The study classifies four main types of land use or land cover: Buildings, water bodies, vegetation, and roads. The data is segmented into like-shaped polygons and transformed into a typical GIS structure for potential updates to the database.

The model is prepared and evaluated using high-resolution images taken by the QuickBird satellite above the urban region of Colombo. Evaluating accuracy involves quality assessments based on visual inspection, revealing an overall precision level of 78%. The study demonstrates how the LVQ method is efficiently utilized to extract characteristics from intricate remote sensing pictures, highlighting the significance of high-quality training samples and the need for filtering to enhance accuracy.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The importance of feature extraction in map production, particularly in fast-changing urban areas, is emphasised in the introduction, which also highlights the challenges of the slow updating process of GIS databases. The historical background emphasises the focus on extracting land use and land cover from satellite data since the late 1960s, with different methods being tested. The article talks about the drawbacks of automatic object detection techniques and the movement towards more sophisticated classification methods such as Artificial Neural Networks (ANNs). Artificial neural networks are recognized for being capable of managing intricate, non-linear patterns in high spatial resolution remote sensing data, providing benefits like parallel computation and rapid generalisation ability. The importance of updating GIS databases in Sri Lanka is highlighted, pointing out the lack of automated extraction systems in the country.

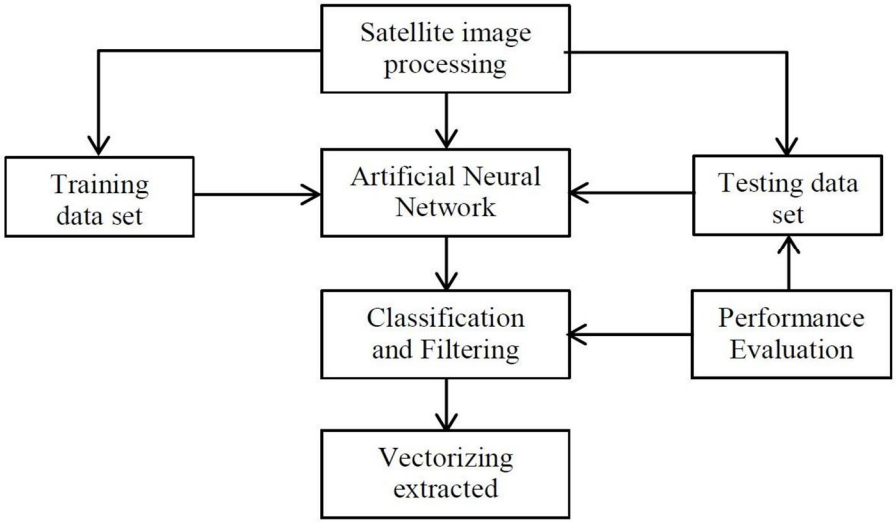


Fig. 4 Approach for the retrieval of polygonal characteristics

The materials and methods section states that the data source is QuickBird satellite images with a resolution of 0.6 m for Panchromatic and 2.4 m for Multispectral. Both training and testing data are selected from the Colombo urban area, which consists of four primary land cover or land use categories. The procedure involves transforming Digital Number (DN) values into spectral reflectance and employing MATLAB's Neural Network Toolbox to implement the ANN. The LVQ technique is specifically utilised for classifying the specified traits.

Metrics for evaluating performance, such as Detection Percentage, Branching Factor, and Quality Percentage, are used to gauge the precision of the created model. The training dataset's classification accuracy is highlighted in the results and discussion section, demonstrating high accuracy for vegetation, water, buildings, and roads.

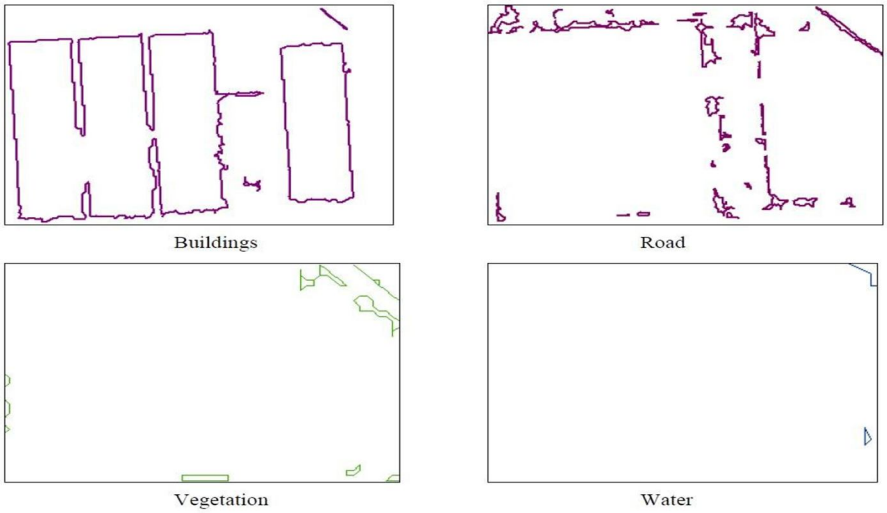


Fig. 5 Features that have been taken out

Examples demonstrate how the extraction process utilises the Tasseled Cap Transformed image space for filtering objects and enhancing accuracy. The extracted features at the end include buildings, roads, vegetation, and water, showcasing the potential for GIS database updates.

Extracted Feature Category	Detection Percentage (%)	Branching Factor	Quality Percentage (%)
Vegetation	83.1461	0.0270	81.3187
Water	94.4444	0.2491	73.9130
Road	90.6125	0.5451	60.6528
Buildings	97.2460	0.0124	96.1030
Overall	**91.3625**	**0.2197**	**77.9969**

Table 1:Performance Enhanced through LVQ Neural Network

The last key aspect emphasises the effective development of a vector map through the utilisation of supervised artificial neural network classification, merging remote sensing and GIS technologies. The meticulous construction of the ANN played a role in attaining an overall accuracy of 78%, despite challenges such as atmospheric conditions, mixed pixels, and limited training data. The focus of the research is on extracting features from satellite images using a self-organising supervised learning artificial neural network with emphasis on land use and land cover data utilising an LVQ method. The study provides insight into the challenges, strategies, and outcomes of the proposed approach, emphasising its potential for updating GIS databases in urban areas.

F. Detection of Water Bodies' Location from High-Resolution Satellite Images

Water resources are crucial for environmental and regional planning, natural disaster management, industrial and agricultural production, and various other significant purposes. Precisely examining bodies of water and describing their features is an important initial stage, especially in regions such as India where water bodies are abundant in the surroundings. While satellite imagery is helpful, it may not always provide precise depictions of water features due to challenges like reflections and shadows from buildings. This article investigates techniques for detecting water bodies using satellite remote sensing. The review discusses methodological issues, empirical investigations, and identifies important factors for future research.

1) Methods that are Supervised and Unsupervised

Improvements in Earth observation sensor technology allow for gathering multispectral data with increased dimensionality. Moreover, the incorporation of data from multiple sources leads to greater complexity. The increased dimensionality presents multiple challenges for processing technology :

- The capacity to categorise a larger variety of groups;
- The need for increased computing power to manage data with many dimensions.
- Substantial increase in processing time as a result of the larger dimensionality and growing number of classes.

Typically, automated techniques for identifying patterns are employed to analyse remotely gathered data. One common approach in classification is to use maximum likelihood (ML) assuming Gaussian class distributions. However, Gaussian ML classification leads to longer processing times, leading to increased computational costs in the end. This computational task is particularly crucial for analysing remote sensing data across a large region or when the processing hardware has limited capabilities.

a) Utilising Support Vector Machine (SVM) for Land Cover Mapping (1A1 and 1AA Techniques)

Utilised SVM with One-Versus-One (1V1) and One-Versus-All (1VA) methods for mapping land cover in the source area of the Nile River in Jinja, Uganda. Utilised Landsat's optical bands (1, 2, 3, 4, 5, and 7) to categorise areas as built-up, vegetated, or water bodies. Emphasising unclassified and mixed pixels, it was found that 1AA had a greater prevalence of these types compared to 1A1. Examined SVM classifiers showing differences in classification accuracy across various categories.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Classifier Type	1A1 Unclassified Pixels	1A1 Mixed Pixels	1AA Unclassified Pixels	1AA Mixed Pixels
Linear	16	0	700	9048
Quadratic	142	0	5952	537
Polynomial	69	0	336	2172
RBF	103	0	4645	0

Table 2 Summary of unassigned and combined pixels

b) Utilising Learning Vector Quantization (LVQ) Neural Network for Water Bodies Extraction

Recommended LVQ method for automated identification of water features from Landsat 4 satellite imagery in the Mississippi River region (1986). Comparison was made between Tasseled Cap Transformation (TCT) and conventional rule-based methods to determine the results. The LVQ method was discovered to be automatic but not as effective as rule-based and TCC methods, which are guided by human input.

c) Algorithm for classifying satellite images using Bayesian supervised Learning

Utilised a Bayesian supervised algorithm incorporating the mean intracluster distance in the fuzzy Gustafson-Kessel (GK) and Bayesian algorithm. Fuzzy GK algorithm is modified for FCM to accommodate different data structures by incorporating covariance matrix. Utilised for categorising satellite images from remote sensing using multi-dimensional data, with iteration continuing until minimal alteration in membership value is achieved. Classification outcomes from IKONOS satellite images were exhibited.

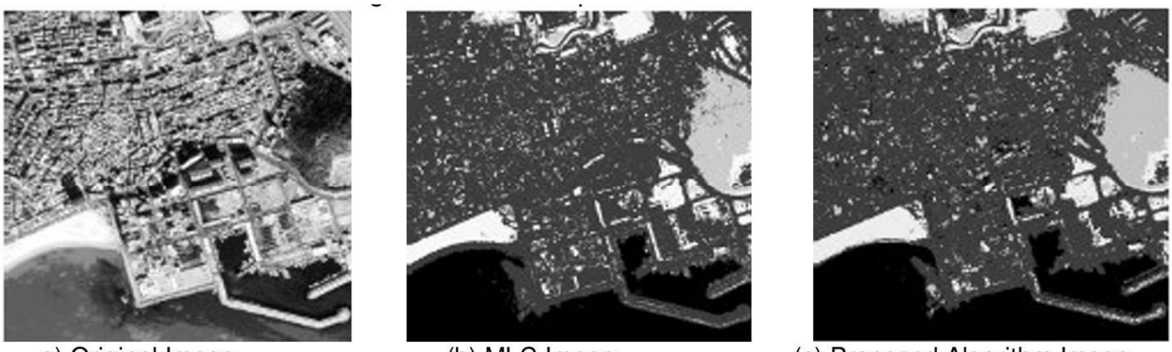


Fig. 6 Results of categorising images with maximum likelihood algorithm (MLC) and a suggested algorithm are compared using IKONOS satellite image.

d) Classification of High-resolution satellite images using texture attributes based on Gabor Filter

Utilised Gabor Filter for textural attributes in supervised classification of high-resolution satellite images. Texture analysis for identifying water bodies using wavelet transform and Gabor filters is suggested. Proposed pixel-wise categorization and image analysis based on objects. Also suggested the use of mathematical morphological analysis and chromaticity analysis to eliminate atmospheric variations.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

G. High Resolution Satellite Images Are Used With Artificial Neural Networks To Automatically Identify Building Features

For many years, the automation of building extraction has been extensively studied in the fields of digital photogrammetry and computer vision, with applications ranging from automatically extracting information in images to updating GIS databases. Numerous techniques and algorithms have been suggested for generating 2D or 3D building models from satellite and aerial images, often utilising edge-based strategies like identifying linear features, organising for isolating parallelogram structures, and confirming building polygons with geometric information and shadow considerations. Acknowledging the complexity involved, more people are realising the importance of using different algorithms, cues, and data sources together to enhance the dependability and strength of extraction procedures. The development of high-tech satellite imaging sensors like IKONOS has provided essential data, allowing accurate spatial resolution for urban areas and helping to detect and extract features such as roads and buildings. Automated techniques have been proposed to accelerate the creation of urban maps in response to the sluggish process of manual building extraction. One popular approach in contemporary building extraction studies using high-resolution satellite images involves supervised methods. These approaches involve either utilising primary training data for classification to generate theories on the locations and dimensions of possible building attributes, or using training datasets and model databases for building categorization and comparison. This article introduces a newly developed automated system designed specifically to recognize and extract buildings from high-resolution satellite images. The system recommendation combines structural and spectral information, utilising Artificial Neural Networks methods to enhance the accuracy and detection rate of building extraction.

The system is segmented into three primary stages, as shown in Figure 1. During the first stage, the system performs image processing and segmentation. In the next stage, features are obtained from segments of the image, and artificial neural networks are used to predict if each segment is a building using the extracted features. The system goes through two separate stages: the Learning Phase involves training the neural network with manually recorded data to reach a desired accuracy level, while the Application Phase involves testing the system on a different dataset.

The first part of image processing consists of different steps like image preprocessing, seed-based image segmentation using the region growing algorithm, and image post-processing. Pre-processing of images in the urban area of Kashan, Iran, includes merging PAN and Multispectral images from IKONOS using ENVI 4.0 software. Then, the seeded region growing algorithm is utilised to initially segment the image, specifically focusing on homogeneous roof areas. The calculation of image intensity (I) involves an equation that includes the RGB values. Afterwards, image segmentation is improved by the post-processing stage using opening and closing operators, refining jagged edges, and filling in small gaps within regions.

Geometric Characteristics

This section outlines the geometric characteristics used and how they are calculated, with a focus on both area and perimeter.

- The region's size is determined by tallying the pixels it contains. If the number of pixels calculated is greater than 10,000, then it will be determined that the space does not represent a building.
- Perimeter Calculation: The perimeter of a region is determined by counting the pixels along its edges.

Features related to the structure

The structural characteristics are essential in differentiating buildings from other items. The following attributes are taken into account:

Roundness: Regardless of the size of the region, roundness is determined by dividing the area by the square of the perimeter. It varies between 0 and 1, as determined by the formula:

$$Roundness = \frac{4\pi . Area}{Perimeter^{2}}$$

Compactness is determined by the number of times opening and closing functions must be repeated to fully eliminate a region.

Particular Angles: Different types of region axes are showcased, with the main axis being the line that joins two contour points that are the most distant from each other. The cross axis is a vertical line that joins two contour points that are the farthest from the main axis. Additionally, two secondary axes can be found as lines at right angles to the main axis and located at the maximum distance from the perimeter. An extra axis is established for every side of the primary axis. The six points created by calculating these axes closely mimic the shape of the region, forming a hexagon.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The Evaluation of Results section examines how well the trained neural network has performed in developing recognition. The system is evaluated on image sections that were not utilised during training, and the results are compared with saved manual results to determine the accuracy of the system. Assessment criteria consist of true positive (TP), true negative (TN), detection percentage (DP), and branch factor (BF).

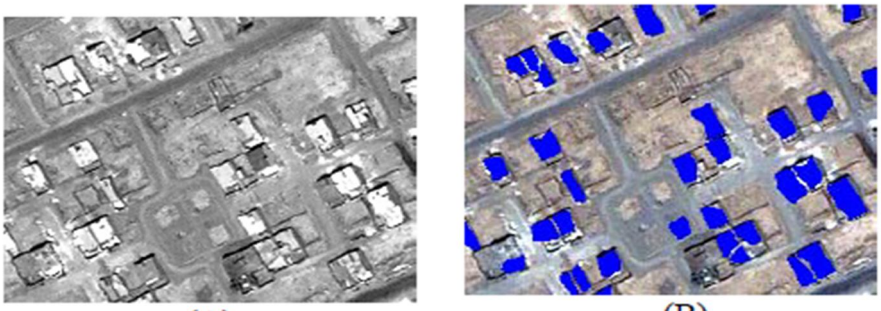


Fig 7. (A) Input of satellite image (B) Output of the system

We use image sections measuring 1000 by 1000 pixels to evaluate the system's performance, encompassing an estimated 1 square kilometer. Two measurements are used to compare the system's results with manually saved data and assess its accuracy.

$$DP = \frac{100.TP}{TP + TN}$$

$$BF = \frac{100.FP}{TP + FP}$$

Parameters like detection percentage (DP) and branch factor (BF) are used to assess the system's performance. Findings from four different areas of the image show high detection rates varying from 80.4% to 91.7%, with corresponding branch percentages ranging from 17.8% to 21.3%. In general, the suggested system demonstrates encouraging outcomes in automatically identifying buildings, offering a significant contribution to the study of satellite image analysis and urban mapping..

H. You Only Look Twice: Fast Detection of Objects at Various Scales in Satellite Images.

This study focuses on the difficulty of modifying advanced object detection frameworks, like YOLO (You Only Look Once), for identifying objects in satellite images. Despite the advancements in computer vision utilizing CNNs and frameworks such as Faster R-CNN, SSD, and YOLO, they are still constrained when it comes to processing the sizable input sizes often associated with satellite imagery. YOLO excels in its speed of inference and performance on commonly used datasets such as PASCAL VOC, Picasso Dataset, and People-Art Dataset, making it a strong choice for detecting objects in satellite imagery.

Nevertheless, utilizing deep learning techniques on satellite images presents distinctive obstacles such as limited object size, full rotational variance, and insufficient training data. Algorithms must be effective at resolutions between 30 cm and 3-4 meters per pixel to accurately identify small and densely clustered objects in satellite imagery. Furthermore, items in satellite images may be oriented in various ways, as opposed to objects in typical datasets such as ImageNet that are typically oriented vertically. Additionally, the lack of annotated training data for satellite imagery and the extremely high resolution of satellite images pose additional computational obstacles.

Fig. 8 YOLT Dataset for Training

In reaction to the constraints of current object detection frameworks for aerial images, a new framework named You Only Look Twice (YOLT) is presented in the paper. YOLT is designed for examining geospatial images and works with external Python libraries, allowing a larger user base to use it without needing to know C programming. YOLT's network architecture is specifically created to tackle the difficulties in precisely identifying dense objects like cars or buildings in satellite images. By employing a 22-layer structure that decreases by a factor of 16, a dense prediction grid of 26x26 is generated for a 416x416 pixel input image. This design, based on the YOLO network but enhanced for compact objects, features a passthrough layer to enhance the accuracy of small objects by allowing access to more detailed features. All convolutional layers, except the last one, are normalised in batches with a leaky rectified linear activation, while the final layer predicts bounding boxes and classes. In general, YOLT provides a hopeful remedy for enhancing object detection in aerial images, especially in situations with dense objects, like parking lots.

Object Class	F1 Score	Run Time (km^2/min)
Car†	0.90 ± 0.09	32
Airplane*	0.87 ± 0.08	32
Boat *	0.82 ± 0.07	32
Building*	0.61 ± 0.15	32
Airport*	0.91 ± 0.14	6000

† IOU = 0.25
* IOU = 0.5

Table 3: YOLT Performance and Speed

To sum up, this study introduces a new method to tackle the difficulties of detecting objects in satellite images through a fully convolutional neural network system called YOLT (You Only Look Twice). Traditional object detection algorithms work well in regular datasets but face challenges with satellite imagery due to differences in object sizes, orientations, and resolutions. In order to address these restrictions, we created a customised process that utilises two classifiers trained at various levels to enhance accuracy in detection.

The YOLT pipeline shows encouraging results, with object detection F1 scores varying between 0.6 and 0.9 for various categories. Although not reaching the levels of well-known datasets such as ImageNet, the scores achieved signify notable advancements in satellite imagery object detection. Furthermore, our method demonstrates the capacity to be trained on information from one sensor and then utilise the model on a different one, emphasising its flexibility and real-world usability.

I. Real-Time Applications Utilise Deep Learning To Recognize Objects In Multispectral Satellite Imagery

Quantamental hedge funds use satellite imagery in their financial trading algorithms to seek alpha as a valuable data source. The integration of satellite data in real-time and computer vision enables investment managers to analyse ground truth information for making predictions in financial markets. Practical uses include predicting company revenue by analysing the number of cars in parking lots, estimating production from supply chain activity, forecasting agricultural commodity prices using crop yield estimates, and monitoring global oil tank lids to detect oil supply. These observations go further than just finance, helping government agencies and non-profit organisations in humanitarian endeavours by evaluating the economic repercussions of COVID-19, identifying forest fires, tracking world whale populations, and performing urgent flood modelling for disaster response. The Committee on Earth Observation Satellites (CEOS) predicts that the whole planet will be mapped by commercial satellite images in high resolution (< 30 cm per pixel) with almost real-time updates. This forecast predicts an increased need for the advancement of accurate and immediate computer vision methods. Nevertheless, current computer vision systems require considerable processing durations (> 30 min) for around 100 km² of satellite images, achieving accuracy levels similar to or lower than expert human annotators (~ 90%). Moreover, existing academic studies do not have techniques that improve object recognition models specifically designed for real-time tasks such as algorithmic trading, which is hindering the integration of satellite images in these areas.

A schematic workflow diagram illustrates the complete process, starting from obtaining satellite images to generating and delivering end signals to the algorithmic trading system.The focus of advancements, highlighted in blue from components P5 to P10, is discussed in detail in the following subsections:

1) The authors suggest four different U-net models with varying depths to discover the best configuration that balances prediction speed and accuracy. The models are made up of a number of layers that is divisible by two, with one fully-connected layer that uses a Sigmoid activation function to produce per-pixel semantic segmentation as the final output. The models' depth can vary from 21 to 39 layers, with each design adding six layers at a time to improve feature extraction. The encoding process acquires semantic/contextual details, while the decoding process includes up-sampling for spatial information restoration. The choice of these suggested configurations is based on visual inspection and quantitative testing.

2) Analysis of Computational Complexity (P7): The paper concentrates on the computational complexity of neural networks to tackle significant signal latency resulting from slow object recognition models. An in-depth examination is conducted on the quantity of floating-point operations (FLOPs) in relation to matrix operations like matrix-matrix multiplications. The writers adapt the FLOPs method to determine the Giga-FLOPs (G-FLOPs) as an indicator of model complexity. This examination aids in comprehending the model's computational cost and how it affects the speed of object recognition.

$$\text{G - FLOPs} = \left[\underbrace{\sum_{e=1}^{E} \left(2 \times \left(\prod_{d=1}^{D_e} A_{ed} \right) \times F_e \times H_e \times W_e \right)}_{\text{Convolutional } (Conv2D) \text{ layers}} + \underbrace{\sum_{b=1}^{B} \left(\prod_{x=1}^{X_b} P_{bx} \times \prod_{z=1}^{Z_b} O_{bz} \right)}_{\text{Max-pooling}(MaxPool) \text{ layers}} \right] / 10^9$$

3) Pixel Frame Sequencing (P9): The training of the neural network involves breaking down large AOI satellite images into frames to accommodate memory constraints on GPUs/TPUs. In order to deal with the issue of mixed landscapes and cropped objects appearing at the edges of frames, the authors have introduced a conditional approach called "pixel frame selection." This method supplies the U-net with specific pixel frames during training by adhering to rules like randomly selecting frames, avoiding frames from various satellite images, and eliminating duplicate frames. This method is focused on decreasing inaccurate object polygons and background noise in the training data, resulting in enhanced training accuracy.

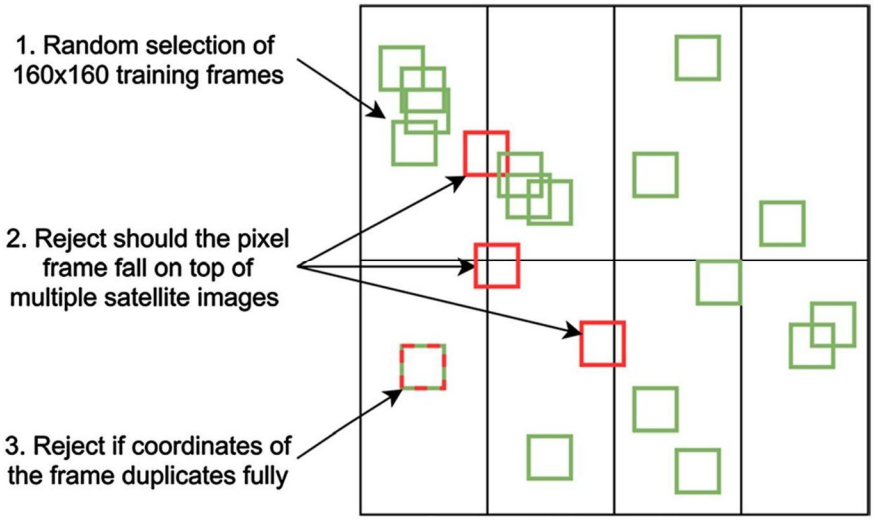


Fig. 9 Approach of selecting pixel frames for training networks. The colour green is used for valid pixel frames, while red is used for rejected pixel frames that were not used in training.

4) Prediction Frame Sequencing (P10): A method known as "prediction frame sequencing" is used to improve prediction by enabling the neural network to take into account a minimum of two distinct backgrounds (prediction frames) for object classification. If there is a discrepancy in classification, the object is deemed as being correctly identified. The effectiveness of this method is evaluated by training two identical networks on the same datasets, with the second network applying the "prediction frame sequencing" technique, leading to an increased accuracy in object recognition.

J. Deep Learning Segmentation and Classification for Urban Village Through U-Net Algorithm with Worldview Satellite Imagery

Unplanned urban areas create major obstacles for urban management and reconstruction planning because of the absence of geospatial data. Identifying specific buildings in these regions using remote sensing images is especially challenging because of their intricate terrains and densely populated areas. Nevertheless, recent developments in deep learning provide possible answers to this issue. In this research, we suggest a framework for urban village mapping using the U-Net deep learning structure. The study is centred on a location in Guangzhou City, China, using a Worldview satellite image containing eight pan-sharpened bands at a 0.5-meter spatial resolution, as well as a vector file outlining building boundaries. The research involves ten urban village locations in the Worldview picture scene, with six sites chosen for training the deep neural network model and four sites for testing. Models for both building segmentation and classification were trained and assessed.

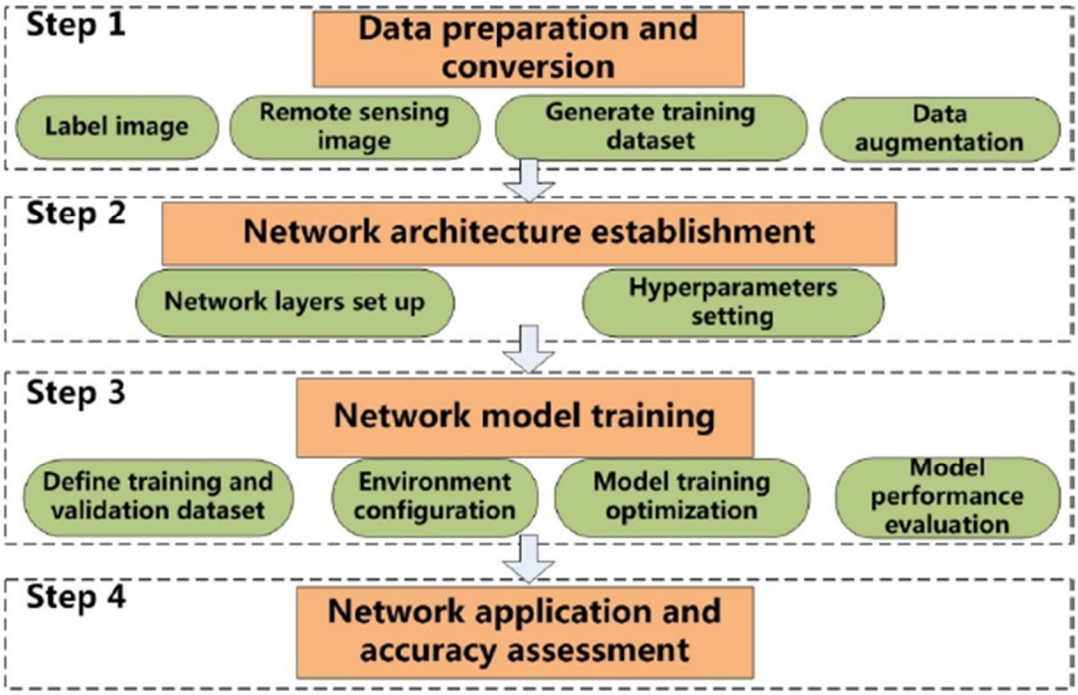


Fig. 10 Process of training and applying deep neural networks

This review brings attention to the difficulties presented by spontaneous settlements caused by quick urban growth and the constraints of conventional mapping techniques like on-site surveys. It highlights the importance of using effective image-based approaches to tackle these issues, considering the laborious process involved in manual remote sensing methods. Recent developments in remote sensing mapping have been dependent on object-based image analysis and machine learning techniques, but they frequently struggle to accurately outline individual buildings in densely populated urban regions. CNNs, especially in deep learning, show potential in tasks involving semantic segmentation, and the U-Net design has become popular for its capability to achieve accurate segmentation even with a small amount of training data.

Nonetheless, it is still uncertain whether deep learning techniques can effectively and accurately distinguish between neighbouring buildings in densely populated urban areas, as well as the potential of CNNs for categorising structures in these environments. This study suggests a deep learning model using U-Net to address the issue of identifying and classifying high-density buildings in urban villages. The study's goal is to show how individual buildings can be mapped and to confirm the accuracy of the approach by using this framework.

III. CONCLUSION

To sum up, this review paper thoroughly explores the most recent methods in identifying and segmenting objects in satellite images, highlighting the crucial impact of advanced machine learning, specifically deep learning techniques such as Convolutional Neural Networks (CNNs). It emphasizes how important accurate object identification is for city planning, dealing with disasters, and monitoring the environment, showcasing how deep learning can greatly improve the accuracy of image processing.

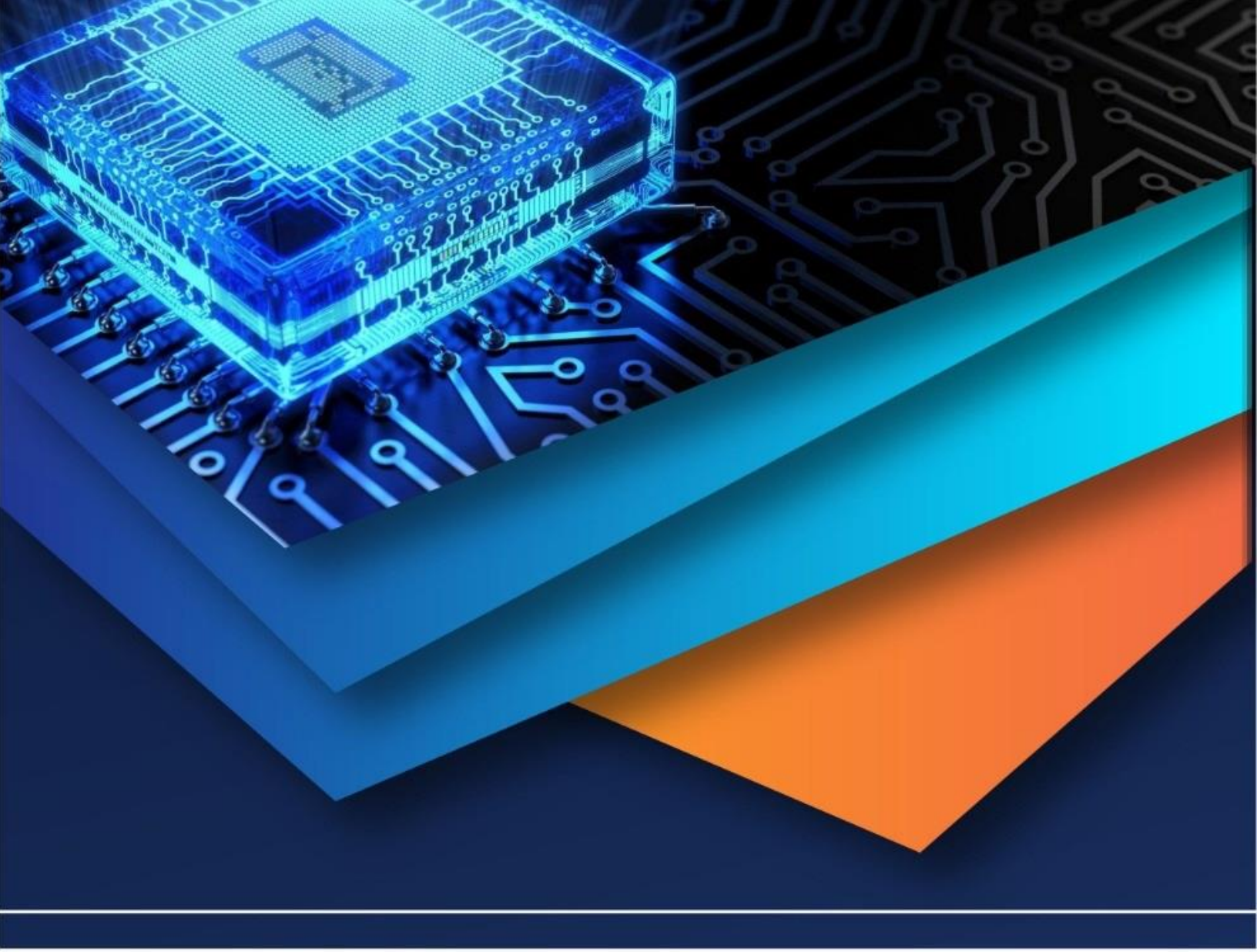
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The paper offers important insights for researchers and professionals in choosing appropriate methodologies by carefully examining conventional techniques alongside deep learning approaches. Furthermore, it tackles issues like scarce data accessibility and model interpretability when investigating potential future research directions. Adding real-life examples through practical case studies enhances the conversation, providing concrete instances of deep learning usage in analyzing satellite images. In the end, this review paper is a crucial tool that provides information and direction for using satellite imagery for various purposes, influencing the future development of technology in this field.

REFERENCES

[1] Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., & Nemani, R. (2015, September 11). DeepSat - A Learning framework for Satellite Imagery. arXiv.org. https://arxiv.org/abs/1509.03602v1
[2] Chandra, N., & Ghosh, J. K. (2017, March 1). A Cognitive Method for Building Detection from High-Resolution Satellite Images | Chandra | Current Science. A Cognitive Method for Building Detection From High-Resolution Satellite Images | Chandra | Current Science. https://i-scholar.in/index.php/CURS/article/view/140018/128285
[3] A Comparison of Deep Learning Vehicle Group Detection in Satellite Imagery. (n.d.). A Comparison of Deep Learning Vehicle Group Detection in Satellite Imagery | IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9006415
[4] Automatic feature extraction from satellite images using LVQ neural network. (n.d.). View Article. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=EErhtNQAAAAJ&citation_for_view=EErhtNQAAAAJ:Tyk-4Ss8FVUC
[5] Image Enhancement and Feature Extraction Based on Low-Resolution Satellite Data. (n.d.). Image Enhancement and Feature Extraction Based on Low-Resolution Satellite Data | IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/7084576
[6] Pan, Z., Xu, J., Guo, Y., Hu, Y., & Wang, G. (2020, May 15). Deep Learning Segmentation and Classification for Urban Village Using a Worldview Satellite Image Based on U-Net. MDPI. https://doi.org/10.3390/rs12101574
[7] Etten, A. V. (2018, May 24). You Only Look Twice: Rapid Multi-Scale Object Detection In Satellite Imagery. arXiv.org. https://arxiv.org/abs/1805.09512v1
[8] Gudžius, P., Kurasova, O., Darulis, V., & Filatovas, E. (2021, June 22). Deep learning-based object recognition in multispectral satellite imagery for real-time applications - Machine Vision and Applications. SpringerLink. https://doi.org/10.1007/s00138-021-01209-2
[9] Abraham, L., & Sasikumar, M. (2012, January). Automatic Building Extraction from Satellite Images using Artificial Neural Networks. Procedia Engineering, 50, 893–903. https://doi.org/10.1016/j.proeng.2012.10.097
[10] Nath, Swapan Kumar Deb, R. K. (n.d.). Water-Body Area Extraction From High Resolution Satellite Images-An Introduction, Review, and Comparison. Water-Body Area Extraction From High Resolution Satellite Images-An Introduction, Review, and Comparison. https://www.cscjournals.org/library/manuscriptinfo.php?mc=IJIP-106
[11] IEEE Xplore Full-Text PDF: (n.d.). IEEE Xplore Full-Text PDF: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7084576
[12] Lin, Y., Xu, D., Wang, N., Shi, Z., & Chen, Q. (2020, September 14). Road Extraction from Very-High-Resolution Remote Sensing Images via a Nested SE-Deeplab Model. MDPI. https://doi.org/10.3390/rs12182985
[13] Urban feature shadow extraction based on high-resolution satellite remote sensing images. (2023, July 12). Urban Feature Shadow Extraction Based on High-resolution Satellite Remote Sensing Images - ScienceDirect. https://doi.org/10.1016/j.aej.2023.06.046
[14] Advanced road extraction using CNN-based U-Net model and satellite imagery. (2023, August 12). Advanced Road Extraction Using CNN-based U-Net Model and Satellite Imagery - ScienceDirect. https://doi.org/10.1016/j.prime.2023.100244
[15] ELsharkawy, A. S., & Abdalla, M. (2019, April 1). Feature extraction from high resolution satellite images using K-means and colour threshold approach. Feature Extraction From High Resolution Satellite Images Using K-means and Colour Threshold Approach. https://doi.org/10.1088/1757-899X/610/1/012045

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)