International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT2065124

Satellite Imagery Classification with Deep Learning : A Survey

Niharika Goswami[1], Keyurkumar Kathiriya[2], Santosh Yadav[3], Janki Bhatt[4], Dr. Sheshang Degadwala[5]
[1,2,3,4] U.G. Scholar, Sigma Institute of Engineering, Vadodara, Gujarat, India
[5] Associate Professor, Sigma Institute of Engineering, Vadodara, Gujarat, India

Article Info
Volume 6, Issue 6
Page Number: 36-46
Publication Issue :
November-December-2020

Article History
Accepted : 01 Nov 2020
Published : 05 Nov 2020

ABSTRACT

Object detection from satellite images has been a challenging problem for many years. With the development of effective deep learning algorithms and advancement in hardware systems, higher accuracies have been achieved in the detection of various objects from very high-resolution satellite images. In the past decades satellite imagery has been used successfully for weather forecasting, geographical and geological applications. Low resolution satellite images are sufficient for these sorts of applications. But the technological developments in the field of satellite imaging provide high resolution sensors which expands its field of application. Thus, the High-Resolution Satellite Imagery (HRSI) proved to be a suitable alternative to aerial photogrammetric data to provide a new data source for object detection. Since the traffic rates in developing countries are enormously increasing, vehicle detection from satellite data will be a better choice for automating such systems. In this research, a different technique for vehicle detection from the images obtained from high resolution sensors is reviewed. This review presents the recent progress in the field of object detection from satellite imagery using deep learning.

Keywords: FCNN, SSD, RCNN, YOLY, CNN, YOLO, CRF and FCN

I. INTRODUCTION

Computer viewing techniques have made great strides in the last few years since the introduction of convolutional neural networks [5] in the ImageNet competition [3]. The availability of large, high-quality data labels such as ImageNet [3], PASCAL VOC [2] and MS COCO [6] has helped to promote further advances in the availability of fast-moving objects near real-time; the three best ones are Faster R-CNN [2], SSD [6], and YOLO [9]. Fast IR-CNN captures 1000 x 600 pixels, while the SSD uses input images of 300 x 300 or 512 x 512 pixels, and YOLO works with 416 x 416 pixels input or 544 x 544. While the performance of all these structures is impressive, no one can come close to importing ~ 16; 000x16; 000 Input size of satellite imagery. In all three categories, YOLO has shown high acquisition speed and high scores on the PASCAL VOC database. The authors of the also have also shown that this framework is largely transferred to new domains by demonstrating the superior performance of other frameworks (i.e., SSD and Faster R-CNN) in the Picasso Dataset [3] and the People-Art Dataset.

Because of the speed, accuracy, and flexibility of YOLO, we appropriately use this program as the inspiration for our satellite image acquisition framework.

The use of in-depth learning methods in traditional discovery pipes is no small feat for a variety of reasons. The distinctive features of satellite imagery require the provision of an algorithm to address challenges related to the location range of the target object, the full orbit of rotation, and the large search area. The use of in-depth learning methods in traditional discovery pipes is no small feat for a variety of reasons. different features of satellite imagery require algorithmic rendering to deal with challenges related to the scope of pre-targeted objects, total consistency, and a large search space. In addition to the start-up details, algorithms should be suitable for:

Small scope in satellite imagery objects of interest are then much smaller and more concentrated, than the larger and more prominent headlines common in ImageNet data. In a satellite environment, the resolution is usually defined as the ground sample range (GSD), which defines the pixel body size of a single image. Commercially available images range from 30 cm GSD for Digital-Globe sharp objects, up to 3 to 4 meters for Planet GSD images. That means that for small objects such as cars each item will be only 15 pixels per maximum resolution.

Complete rotating objects viewed from the top can be positioned (e.g. vessels can have a head between 0 and 360 degrees, and trees in ImageNet data stand reliably). In most cases the training model has a relative shortage of training data (although attempts like SpaceNet1 try to improve this problem)

Ultra-high-resolution Input images are very large (usually hundreds of megapixels), so simply reducing the input size required for most algorithms (a few hundred pixels) is not an option.

II. RELATED WORKS

In [1] Nyan Linn Tun, Alexander Gavrilov and Naing Min Tun, used FCNN to separate Satellite images. The proposed model performs high precision separation in very small periods without high computer performance. They have a breakdown of 0.70% accuracy. Many image classification technologies have been developed and the effects of classification can also achieve significant improvements in both categories of classification. They created a confusion matrix in reference database (with 21 classes) and gained 70.48% accuracy. The satellite image classification has been developed to compare various key regions of the region to use using in-depth learning technology.

In [2] Adam Van Etten, he used a complete neural network pipeline (SIMRDWN) to make cars and airports faster on satellite images. This pipeline integrates leading acquisition techniques such as SSD, Faster RCNN, R-FCN, and YOLT into a single framework that quickly analyses test images of opposing size. For large confirmation images, we use a split with two different scales: 200m, and 5000m. The results of R-FCN and Faster-RCNN do not follow the conclusions that these models live in a "good place" in terms of speed and accuracy. The YOLT design works much better than other fast object detection components, indicating that it seems better to be able to disrupt things from behind with smaller training sets.

In [3] Yohei Koga, Hiroyuki Miyazaki and Ryosuke Shibasaki, proposed an uncontrolled domain modification (DA) approach to the deterioration of performance caused by image aspect differences between data domains. They used the correlation of Correlation (CORAL) DA and the opposition DA on a regional-based vehicle detector and improved the accuracy of receiving more than 10% of the targeted domain. Their proposed method achieved better performance than the accuracy obtained by labelling data with the target domain label. The DA method is effective even in a situation where a well-labelled database may not achieve the desired maximum accuracy due to the complexity of the target area, such as the complexity of the image element.

In [4] Rodrigo F. Berriel, Andre Teixeira Lopes, Alberto F. de Souza, and Thiago Oliveira-Santos, proposed a system of categorization of large zebra satellite satellites. The system automatically detects images of intersected and unpaved roads around the world using the Google Static Maps API, Google Maps Directions API and OpenStreetMap. The experiment was performed in this novel dataset with 245,768 photographs from three different continents, 9 countries and more than 20 cities. The test results confirmed the strength of the proposed system and showed 97.11% accuracy in the global survey.

In [5] Tanguy Ophoff, Steven Puttemans, Vasileios Kalogirou, Jean-Philippe Robin and Toon Goedeme, investigated the possibility of the detection of small spontaneous objects, such as cars and ships, in satellite images with a resolution of between 0.3 and 0.5 m. The main challenges of this work are small objects, as well as the spread of object sizes, with objects ranging from 5 to a few hundred pixels in length. They trained and tested four different detection networks: YOLOV2, YOLOV3, D-YOLO and YOLT, adjusting multiple hyperparameters to achieve maximum accuracy. They performed various tests to better understand the performance and differences between the models. The most efficient model, D-YOLO, has achieved accuracy between 60% of vehicles and 66% of vessels and can process 1 Gpx image in 14 s. They concluded that D-YOLO appears to be a complete detector, achieving high accuracy (APvehicle: 60%, APvessel: 66%) and fast operating time (± 4 ms per 416 × 416 patch).

In [6] Jiangye Yuan, he has a flexible network design with a final phase that integrates performance from multiple previous stages of pixel intelligence prediction, and introduces a signed range of construction parameters as output, with improved representation power. They introduced a new architecture approach that combines the ConvNet framework with rich GIS data. The qualified system is tested on a large real-world database and provides accurate results with high efficiency. The method proposed in this paper separates semantic objects (structures) into complex squares, which is a special case of land division. Test results show that with enough labelled data their ConvNet model correctly distinguishes the elements behind them with invisible data.

In [7] Mark Pritt and Gary Chern, presented an in-depth reading program that distinguishes objects and resources in satellite imagery with high definition. The program contains CNN's integration with post-processing neural networks that include predictions from CNN and satellite metadata. In the IARPA fMoW database of one million images in 63 categories, including the false detection phase, the system obtains an accuracy of 0.83 and a F1 rating of 0.797. It divides 15 classes with 95% or better accuracy and beats the John Hopkins APL model by 4.3% in the fMoW TopCoder challenge. By monitoring a satellite imagery store, it can help law enforcement to detect illegal mining operations or illegal fishing vessels, assist disaster response teams by mapping mud slides or storm damage, and empowering investors to monitor crop growth or oil resource development more effectively.

In [8] Milena Napiorkowska, David Petit and Paula Martí, used the FCN-VGG network to detect three different objects or features in satellite imagery: roads, palm trees and cars. In remote sensing, satellite images are also used for feature extraction and often classic machine learning techniques are used for the classification of the pixels in the image. Results are very promising and on-going work shows that they can be improved further, reaching in some cases 98%-99% accuracy. Results also show that, their approach is good at finding objects that might have different colours and slightly varying shapes, which cannot be achieved as easily using more common techniques in remote sensing such as Random Forest or Support Vector Machine.

In [9] Vladimir Iglovikov, Sergey Mushinskiy and Vladimir Osin, approach is based on an adaptation of fully convolutional neural network for multispectral data processing. Their approach includes several steps, such as the adaptation of fully convolutional network to multispectral satellite images with joint training objective and analysis of boundary effects, reflectance indices. Its accuracy is comparable to the first two places, but unlike those solutions, it doesn't rely on complex assembling techniques and thus can be easily scaled for deployment in production as a part of automatic feature labelling systems for satellite imagery analysis.

In [10] Ekaterina Kalinicheva, Jer emie Sublime and Maria Trocan, have presented an end-to-end approach for the detection and clustering of non-trivial changes between two bi-temporal high-resolution satellite images. The presented baseline firstly deploys a neural network autoencoder for feature extraction and comparison to detect some meaningful changes. Their approach has shown promising results on the experimental dataset and outperformed concurrent approaches.

In [11] Qiling Jiang, LiujuanCao, Ming Cheng, Cheng Wang and Jonathan Li, propose a vehicle detection method in satellite images using Deep Convolutional Neural Network (DNN). DNN is a model of deep learning and it has a high learning capacity when dealing with images. Deep Neuro Network (DNN) based classifier is adopted for classifying whether the target super pixel is vehicle or not. Their experiment shows that DNN has excellent performance in comparison to alternative approaches. They will further investigate the possibility to transfer the trained detector across different resolutions, which is a common-sense challenge in vehicle detection in satellite images.

In [12] Tomohiro Ishii, Edgar Simo-Serra, Satoshi Iizuka, Yoshihiko Mochizuki, Akihiro Sugimoto, Hiroshi Ishikawa, and Ryosuke Nakamura, present an approach for the detection of buildings in multispectral satellite images. Their approach consists of training a Convolutional Neural Network (CNN) from scratch to classify multispectral image patches taken by satellites as whether or not they belong to a class of buildings. They adapt the classification network to detection by converting the fully-connected layers of the network to convolutional layers, which allows the network to process images of any resolution. They present a new dataset for the detection of solar power plants in multispectral images to evaluate their approach, although it can be applied to detect any type of building. They provide an in-depth evaluation of the seven different spectral bands provided by the satellite images and show it is critical to combine them to obtain good results.

In [13] Ahmad Mansour, Ahmed Hassan, Wissam Hussein and Ehab Said, they have use two states of art algorithms for object detection (Faster RCNN and SSD). They construct vehicle dataset collected by Google Earth and other satellite samples such as JF-2 and WORLD-VIEW satellites. Mean average precision (MAP) used for performance evaluation. Based on the results obtained, Faster R-CNN Inception-V2 gives better accuracy than SSD Inception-V2. but the SSD Inception-V2 performs in a shorter time for image detection. The study will extend for general vehicle detection (bicycle, motorcycle, bus, truck).

In [14] Y Harold Robinson, S Vimal, Manju Khari, Fernando Carlos Lopez Hernandez and Ruben Gonzalez Crespo, describes a method for the effective semantic segmentation of satellite images, and compares different object classifiers in terms of accuracy and execution time. the image spectrum is used to reduce the computational cost during the segmentation and classification steps. Firstly, artifacts are corrected from the satellite images for facilitating the feature extraction process. They have evaluated and reported the performance of the proposed techniques. The experiments indicate that our proposed tree-based CNN has a higher classification performance and lower execution time than the other techniques. The augmented information will utilize the deeper neural networks and enhance the efficiency by including the neural network concepts.

In [15] Nevrez Imamoglu, Pascual Martínez-Gómez, Ryuhei Hamaguchi, Ken Sakurada, Ryosuke Nakamura, explored the efficiency of recurrent and feedback connections in shallow CNNs on multispectral satellite images for solar power plant classification. They proposed and implemented recurrent-CNN (R-Net) and Feedback-CNN (F-Net) based on a state-of-the-art feed-forward model for multi-spectral image classification. Their experiments demonstrated that using top-down signals (especially recurrent and feedback features together) on CNNs can provide good representation of multi-spectral images which can in turn improve classification accuracy drastically. As a future work, more investigation on recurrent networks can be done on different approaches such as convolutional (Long-Short Term Memory networks) or GRU (Gated Recurrent Units).

III. METHODOLOGY

A. Datasets [1]:

i. Images were extracted from the National Agriculture Imagery Program (NAIP) dataset. The NAIP dataset consists of a total of 330,000 scenes spanning the whole of the Continental United States (CONUS). We used the uncompressed digital Ortho quarter quad tiles (DOQQs) which are GeoTIFF images and the area corresponds to the United States Geological Survey (USGS) topographic quadrangles. The average image tiles are ~6000 pixels in width and ~7000 pixels in height, measuring around 200 megabytes each.

ii. SAT-4 consists of a total of 500,000 image patches covering four broad land cover classes. These include - barren land, trees, grassland and a class that consists of all land cover classes other than the above three. 400,000 patches (comprising of four-fifths of the total dataset) were chosen for training and the remaining 100,000 (one-fifths) were chosen as the testing dataset. We ensured that the training and test datasets belong to disjoint set of image tiles. Each image patch is size normalized to 28x28 pixels. Once generated, both the training and testing datasets were randomized using a pseudo-random number generator.

iii. SAT-6 consists of a total of 405,000 image patches each of size 28x28 and covering 6 landcover classes - barren land, trees, grassland, roads, buildings and water bodies. 324,000 images (comprising of four-fifths of the total dataset) were chosen as the training dataset and 81,000 (one fifths) were chosen as the testing dataset. Similar to SAT-4, the training and test sets were selected from disjoint NAIP tiles. Once generated, the images in the dataset were randomized in the same way as that for SAT-4. The specifications for the various landcover classes of SAT-4 and SAT-6 were adopted from those used in the National Land Cover Data (NLCD) algorithm.

B. Deep learning approaches

1) SAE[2]: SAE is a relatively simple deep learning model that has been successfully applied for remote sensing image scene classification. An SAE consists of multiple layers of autoencoders in which the outputs of each layer are wired to the inputs of the successive layer. Specifically, one should first train the first layer on raw input data to obtain parameters and transfer the raw data into an intermediate vector consisting of activations of the hidden units. Then, this process is repeated for subsequent layers by using the output of each layer as input for the subsequent layer. This method trains the parameters of each layer individually while freezing parameters for the remainder of the model. To obtain better results, after layer-wise training is completed, fine-tuning is performed to tune the parameters of all layers at the same time with a smaller learning rate. Compared to a single autoencoder as mentioned in previous subsection, the feature representation power of SAE can be significantly strengthened. This can be easily explained: with the composition of multiple autoencoder that each transforms the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level, we can learn very powerful representations.
2) CNNs[10]: CNNs are designed to process data that come in the form of multiple arrays, for example a multi-spectral image composed of multiple 2D arrays containing pixel intensities in the multiple band channels. Starting with the impressive success of AlexNet, many representative CNN models including Overfeat, VGGNet, GoogLeNet, SPPNet, and ResNet have been proposed in the literature. There exist four key ideas behind CNNs that take advantage of the properties of natural signals, namely, local connections, shared weights, pooling, and the use of many layers.
3) DNN [3]: The Deep Convolutional Neural Networks is one of the Deep Learning models. It has a strong ability of learning. It is trained with the back-propagation algorithm. Convolutional Neural Networks can be designed to recognize visual patterns directly from pixels, it extracts image features while the pixels forward the networks. The structure of DNN using in this paper is showed in figure 1. It contains three convolutional layers and three max-pooling layers. It also uses ReLU and Local Response Normalization (LRN) mentioned in [4]. All the parameters in DNN are determined by training. They are randomly initialized at first time. And then the caffe will update the networks over and over again by using a large number of labelled data. Each training iteration consist of two parts, the forward part and the back-propagation part. In the forward part, the labelled input image patch will go forward through DNN networks layer by layer. After forward part, Caffe will do the back-propagation part. It uses stochastic gradient descent (SGD) with momentum to deal with the back propagation. The parameters will update layer by layer from. For training the networks, we set 0.0005 as the delay weight rate. The terminational iteration is 120000 times.
4) TSC[19]: A two-layer sparse coding (TSC) model is designed to discover the “true” neighbours of the images and bypass the intensive learning phase of the satellite image classification. ... The images are classified according to a newly defined “image-to-category” similarity based on the coding coefficients.
5) SSD[6]: We present a method for detecting objects in images using a single deep neural network. Our approach, named SSD, discretises the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default

box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. Our SSD model is simple relative to methods that require object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stage and encapsulates all computation in a single network. This makes SSD easy to train and straightforward to integrate into systems that require a detection component. Experimental results on the PASCAL VOC, MS COCO, and ILSVRC datasets confirm that SSD has comparable accuracy to methods that utilize an additional object proposal step and is much faster, while providing a unified framework for both training and inference. Compared to other single stage methods, SSD has much better accuracy, even with a smaller input image size.

6) Faster R-CNN[14]: Detecting Objects Without the Wait Advances in the field of computer vision have been spearheaded by the adoption of Convolutional Neural Networks (CNNs). There are a number of related architectures available, among them the Region-CNN, used for object detection.
7) Region-CNN (R-CNN)[15], originally proposed in 2014 by Ross Girshik et. al., is a deep learning object detection algorithm that aims to find and classify multiple objects within an image. The algorithm doesn't know in advance how many objects there will be in the image. This makes it difficult to use a Convolutional Neural Network (CNN), because the input is of variable length. here is a dilemma with regard to identifying objects in the image—you can arbitrarily choose a few regions and classify them, but then risk missing the important objects. Or check every possible region in the image, which would take too long to run.
8) YOLT[7]: YOLT is an extension of the YOLO v2 framework that can evaluate satellite images of arbitrary size, and runs at ~50 frames per second. Current applications include vehicle detection (cars, airplanes, boats), building detection, and airport detection. The YOLT code alters a number of the files in src/*.c to allow further functionality. We also built a python wrapper around the C functions to improve flexibility. We utilize the default data format of YOLO, which places images and labels in different directories.
9) YOLO[3]: You Only Look Once (YOLO) is a network that uses Deep Learning (DL) algorithms for object detection. YOLO performs object detection by classifying certain objects within the image and determining where they are located on it. For example, if you input an image of a herd of sheep into a YOLO network, it will generate an output of a vector of bounding boxes for each individual sheep and classify it.
10) A Boltzmann machine [11]: (also called stochastic Hopfield network with hidden units or Sherrington–Kirkpatrick model with external field or stochastic Ising-Lenz-Little model) is a type of stochastic recurrent neural network. It is a Markov random field. It was translated from statistical physics for use in cognitive science. The Boltzmann machine is based on stochastic spin-glass model with an external field, i.e., a Sherrington–Kirkpatrick model that is a stochastic using Model and applied to machine learning.
11) U-Net[21]: The architecture looks like a 'U' which justifies its name. This architecture consists of three sections: The contraction, The bottleneck, and the expansion section. The contraction section is made of many contraction blocks. Each block takes an input applies two 3X3 convolution layers followed by a 2X2 max pooling. The number of kernels or feature maps after each block doubles so that architecture can learn the complex structures effectively. The

bottommost layer mediates between the contraction layer and the expansion layer. It uses two 3X3 CNN layers followed by 2X2 up convolution layer. But the heart of this architecture lies in the expansion section. Similar to contraction layer, it also consists of several expansion blocks. Each block passes the input to two 3X3 CNN layers followed by a 2X2 up sampling layer. Also, after each block number of feature maps used by convolutional layer get half to maintain symmetry. However, every time the input is also get appended by feature maps of the corresponding contraction layer. This action would ensure that the features that are learned while contracting the image will be used to reconstruct it. The number of expansion blocks is as same as the number of contraction block. After that, the resultant mapping passes through another 3X3 CNN layer with the number of feature maps equal to the number of segments desired.

IV. COMPARATIVE STUDY

TABLE III. COMPARATIVE STUDY

Method	Advantage	Limitation
FCNN[1]	-deep convolutional networks are flexible and work well on image data. -there is no need of feature extraction.	-Classification of Images with different Positions - Coordinate Frame
SSD[6]	-feature maps improve the detection of objects at different scale. -Design better default boundary boxes wi ll help accuracy.	-performs worse than Faster R-CNN for small-scale objects. -Training the data is unwieldy and too long
Faster RCNN[14]	-both region proposal generation and objection detection tasks are all done by the same conv networks. -With such design, object detection is much faster.	-It cannot be implemented real time as it takes around 47 seconds for each test image. -The selective search algorithm is a fixed algorithm.
R-FCN[16]	-R-FCN is even faster than Faster R-CNN with competitive mAP.	Complex to implement
YOLT[7]	-Up sampling via a sliding window to look for small, densely packed objects -Augment training data with re-scaling and rotations.	-It struggles to generalize objects in new or unusual aspect ratio or configurations.
CNN[2]	-Robust to Noise Can handle multi-step forecasts Can handle multi-variate inputs	-Cannot learn temporal dependence
YOLO[3]	-It's incredibly fast and can process 45 frames per second. -YOLO also understands generalized object representation.	-Struggles to detect small objects. - Comparatively low recall and more localization error compared

		to faster R_CNN
TSC[19]	-Enhanced Efficiency -Optimised Performance	-No Congestion Avoidance -Best Effort Delivery
SIC[8]	-It is easy to install and manage the ground station sites.	-Redundant components are used in the network design. This incur more cost in the installation phase.
CRF[7]	-Compared with HMM: Since CRF does not have as strict independence assumptions as HMM does, it can accommodate any context information. Its feature design is flexible. -Compared with MEMM: Since CRF computes the conditional probability of global optimal output nodes, it overcomes the drawbacks of label bias in MEMM	-CRF is highly computationall y complex at the training stage of the algorithm. It makes it very difficult to re-train the model when newer data becomes available.
FCN[5]	-The integration of multi-sources remotely sensed data. -The integration of information over multiple scales	FCN is a network that does not contain any “Dense” layers (as in traditional CNNs).
UNET[21]	-The U-Net combines the location information from the down sampling path with the contextual information in the up-sampling path to finally obtain a general information combining localization and context, which is necessary to predict a good segmentation map.	-The decoders are disconnected—deeper U-Nets do not offer a supervision signal to the decoders of the shallower U-Nets in the ensemble.

V. CONCLUSION

Satellite Imaginary object classification is requiring for traffic monitoring, Land classification and Military applications. Past method is based on classify only one or two three Land type. It would not provide all object information. In this research, we summarize different types of deep learning methods for classification of objects. For that different learning approaches best strategy is U-net. The main idea is to supplement a usual contracting network by successive layers, where pooling operators are replaced by up sampling operators. Hence, these layers increase the resolution of the output. It’s combines the location information from the down sampling path with the contextual information in the up-sampling path to finally obtain a general information combining localization and context, which is necessary to predict a good segmentation map for future object prediction in satellite images.

VI.REFERENCES

[1] N. L. Tun, A. Gavrilov, and N. M. Tun, “Multi-classification of satellite imagery using fully convolutional neural network,” Proc. - 2020 Int. Conf. Ind. Eng. Appl. Manuf. ICIEAM 2020, pp. 7–11, 2020, doi: 10.1109/ICIEAM48468.2020.9111928.

[2] A. Van Etten, “Satellite imagery multiscale rapid detection with windowed networks,” Proc. - 2019 IEEE Winter Conf. Appl. Comput. Vision, WACV 2019, pp. 735–743, 2019,

[3] A. Van Etten, “You Only Look Twice: Rapid Multi-Scale Object Detection in Satellite Imagery,” 2018.

[4] Y. Koga, H. Miyazaki, and R. Shibasaki, “Correction: A Method for Vehicle Detection in High-Resolution Satellite Images That Uses a Region-Based Object Detector and Unsupervised Domain Adaptation. [Remote Sensing 2020, 12, 575] doi: 10.3390/rs12030575,” Remote Sens., vol. 12, no. 7,

[5] R. F. Berriel, A. T. Lopes, A. F. De Souza, and T. Oliveira-Santos, “Deep Learning-Based Large-Scale Automatic Satellite Crosswalk Classification,” IEEE Geosci. Remote Sens. Lett., vol. 14, no. 9, pp. 1513–1517, 2017, doi: 10.1109/LGRS.2017.2719863.

[6] T. Ophoff, S. Puttemans, V. Kalogirou, J. P. Robin, and T. Goedemé, “Vehicle and vessel detection on satellite imagery: A comparative study on single-shot detectors,” Remote Sens., vol. 12, no. 7, pp. 1–21, 2020, doi: 10.3390/rs12071217.

[7] J. Yuan, “Automatic Building Extraction in Aerial Scenes Using Convolutional Networks,” 2016, [Online]. Available: http://arxiv.org/abs/1602.06564.

[8] M. Pritt and G. Chern, “Satellite image classification with deep learning,” Proc. - Appl. Imag. Pattern Recognit. Work., vol. 2017-October, pp. 1–7, 2018, doi: 10.1109/AIPR.2017.8457969.

[9] “Three Applications Of Deep Learning Algorithms For Object Detection Milena Napiorkowska (1), David Petit (1), Paula Martí (2) (1) Deimos Space UK Ltd ., (2) Deimos Engenharia,” no. 1, pp. 4839–4842, 2018.

[10] V. Iglovikov, S. Mushinskiy, and V. Osin, “Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition,” 2017, [Online]. Available: http://arxiv.org/abs/1706.06169.

[11] E. Kalinicheva, J. Sublime, and M. Trocan, “Object-Based Change Detection in Satellite Images Combined with Neural Network Autoencoder Feature Extraction,” 2019 9th Int. Conf. Image Process. Theory, Tools Appl. IPTA 2019, pp. 1–6, 2019, doi: 10.1109/IPTA.2019.8936085.

[12] C. Wang, Q. Jiang, M. Cheng, J. Li, and L. Cao, “Deep Neural Networks-Based Vehicle Detection In Satellite Images Fujian Key Laboratory of Sensing and Computing for Smart City School of Information Science and Engineering, Xiamen University Xiamen, China,” IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., 2016.

[13] T. Ishii et al., “Detection by classification of buildings in multispectral satellite imagery,” Proc. - Int. Conf. Pattern Recognit., vol. 0, pp. 3344–3349, 2016, doi: 10.1109/ICPR.2016.7900150.

[14] A. Mansour, A. Hassan, W. M. Hussein, and E. Said, “Automated vehicle detection in satellite images using deep learning,” IOP Conf. Ser.

Mater. Sci. Eng., vol. 610, no. 1, 2019, doi: 10.1088/1757-899X/610/1/012027.

[15] G. Cheng, J. Han, and X. Lu, "Remote Sensing Image Scene Classification: Benchmark and State of the Art," Proc. IEEE, vol. 105, no. 10, pp. 1865–1883, 2017, doi: 10.1109/JPROC.2017.2675998.

[16] Y. H. Robinson, S. Vimal, M. Khari, F. C. L. Hernández, and R. G. Crespo, "Tree-based convolutional neural networks for object classification in segmented satellite images," Int. J. High Perform. Comput. Appl., 2020, doi: 10.1177/1094342020945026.

[17] N. Imamoglu, P. Martínez-Gómez, R. Hamaguchi, K. Sakurada, and R. Nakamura, "Exploring recurrent and feedback CNNs for multi-spectral satellite image classification," Procedia Comput. Sci., vol. 140, pp. 162–169, 2018, doi: 10.1016/j.procs.2018.10.325.

[18] A. Hosny and A. Parziale, "A Study on Deep Learning," vol. 9, no. 4, pp. 21482–21483, 2019.

[19] D. Dai and W. Yang, "Satellite image classification via two-layer sparse coding with biased image representation," IEEE Geosci. Remote Sens. Lett., vol. 8, no. 1, pp. 173–176, 2011, doi: 10.1109/LGRS.2010.2055033.

[20] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Han_etal_IEEE_TGRS_2015_Object_detection_in_optical_remote_sensing_images_based_weakly.pdf," pp. 1–26.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[22] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 12, no. 7, pp. 2217–2226, 2019, doi: 10.1109/JSTARS.2019.2918242.

Cite this article as :

Niharika Goswami, Keyurkumar Kathiriya, Santosh Yadav, Janki Bhatt, Dr. Sheshang Degadwala, "Satellite Imagery Classification with Deep Learning : A Survey ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 6, pp. 36-46, November-December 2020. Available at doi : https://doi.org/10.32628/CSEIT2065124
Journal URL : http://ijsrcseit.com/CSEIT2065124