Advances in Artificial Intelligence Research (AAIR) **Vol. 1 (No. 2), pp. 73-79, 2021**
ISSN 2757-7422
www.dergipark.com/aair/ **Published online: Aug 31, 2021**

A Review on Deep Learning Models for Satellite Imagery

Hasan Ersan Yağcı [1,*], Abdullah Atçılı [2], Şükrü Sezer [3]

[1] Artificial Intelligence Department, Mugla Sitki Kocman University; 0000-0001-7556-8811
[2] Artificial Intelligence Department, Mugla Sitki Kocman University; 0000-0001-6872-6754
[3] College of Computing, Georgia Institute of Technology; 0000-0003-3045-2596

Abstract

Object detection and image classification from remote sensing data are used in many different fields. It has been the subject of many studies in recent years. Research in this field has increased with the development of deep learning techniques and remote sensing data, which can be satellite images or unmanned aerial vehicles (UAV), providing high resolution spatial and spectral data. In this review, we survey modern deep learning techniques are trained on remote sensing data. Term remote sensing data is widely used for satellite imagery, however the term also refers to UAV collected data. It is chosen as a topic of the this review that 'how green the metropolitans?'. There are two approaches for this question. First one is the detection of green (vegetation) in all metropolitan and the other one is classification of green types. Convolutional neural networks (CNN), generative adversarial networks (GAN), and autoencoder (AE) were compared on tensorflow's UC Merced dataset.

Keywords: *Deep Learning; Satellite Imagery; Remote Sensing.*

1. Introduction

It is important where we get the data needed. And most of the time, data collection method can and will determine the approach to the problem or data and which method will be used. Remote sensing of images are used in many area like land use land cover (LULC), urbanization, classification, vegetation, change detection [1]. The remote sensing data is used mainly for these problems. This study will cover calculation of green ratio and green classification of the metropolitans with modern deep learning techniques.

Remote sensing is the satellite views mainly. But use of UAV sensor data has recently increased with the development of as a remote sensing drone technology. UAV imagery cannot give the same performance compared to the satellite imagery. The area of coverage and the number the images cannot be compared to those provided by satellites.

There are different studies performed on UAV's remote sensing data. In this study [2], classify the vegetation of the urban according to healthy classes by the drone (UAV) images. Study is performed on aerial images which is taken from distance between 20 to 30 m, on sunny day. In [3], Unmanned Aerial Vehicle (UAV) is used to detect changes of urban areas. For the change detection different times UAV images are needed. In other study [4], UAV images are used to get cultivated land information. In this study [5], UAV images are used to classify vegetation in vegetated areas under clear sky conditions. This study claims that UAV offers considerable advantages with high-resolution capability but adds that UAV has limits on number of sensors that can be mounted.

UAV has some advantages, UAVs can be loaded with specialized sensors according to the purpose of the study. For example, for the vegetation the normalized difference vegetation index (NDVI) can be needed so the spectral resolution high sensors can be mounted. The UAVs can move/fly low height and gives high spatial resolution. However, UAV has some disadvantages too. The study area should be small. The UAV cost is high and needs more time to cover large areas. In addition, number of sensors carried at the same time is strictly limited to weight capacity of UAV. Weather conditions can also limit data collection operations with UAV and it is also challenging to collect data pertaining large areas such as a whole city at a time.

As mentioned above, the study area should be small. UAV imagery can be used for the agriculture, classify the trees in a forest or garden, for the urbanization of the county, for the solar panel's anomaly detection etc. however it is somewhat limited for large-scale data collection operations including over metropolitan areas.

Mostly in the articles, the term “remote sensing data” is used interchangeably with satellite imagery. It can be thought that you can give more details rather than satellite imagery. However the state of art satellite systems are the main source for the remote sensing work area.

There are three major remote sensing academic societies; [6] All of these societies use satellite imagery as the data source as mainly. The remote sensing data will covered in detail at the data section. The remote sensing term is used for the satellite imagery until end of article.

*Corresponding author
E-mail address: hersanyagci@gmail.com

Received: July 15, 2021; Received in revised form: July 29, 2021; Accepted: August 04, 2021.

As mentioned above, the satellites keep providing us with the remote sensing data of the Earth. Remote sensing of images are used in many area like classification and change detection. There is a big data, this data is the source of many studies. **Figure 1** [6]

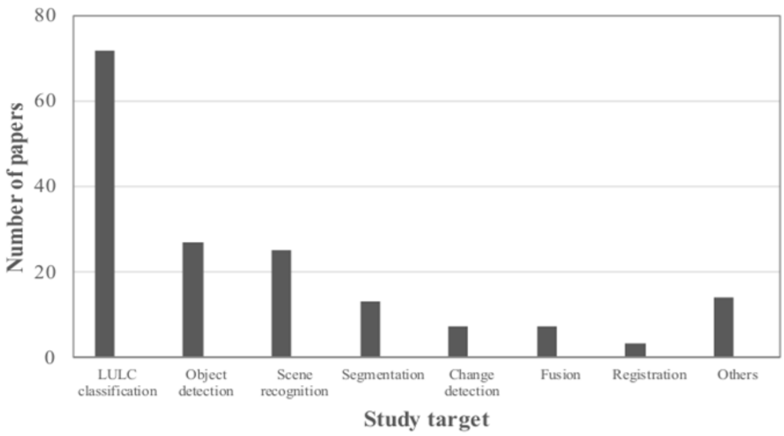

Figure 1. *Remote sensing study targets.*

2. Data

The remote sensing data used for mainly LULC, object detection, vegetation, scene recognition, classification, segmentation, change detection etc. All of these study target mainly use the spectral and spatial resolutions of remote sensing data. The first artificial satellite Sputnik-1 launched at 1957, after this year many satellites launched for different aims by different countries. At 1972 Landsat program has begun and The Earth Resources Technology Satellite is launched.

In 1986 The Spot can take 3D, At 1999 IKONOS first VHR (very high resolution) satellite. It can give 82 cm spatial resolution. At 2008 GeoEye-1 gives under the 50 cm spatial resolution, 2009 worldview-2 9 bant spectral VHR, at 2015 worldview-3 gives under 30 cm spatial VHR and 9 spectral band. With advancing technology, the data becomes more detailed and more precise. With this situation State of the art AI is obligatory. Because high spatial resolution and wide spectral resolution can be possible. There are 3 different orbit levels, as we investigate them, Low Earth Orbit (LEO), Geostationary Earth Orbit (GEO) and Medium Earth Orbit (MEO).

LEO provides high spatial resolution with low temporal resolution while GEO provides for low spatial resolution, but high temporal resolution. We need low orbit altitude to more spatial resolution, it is LEO. There is a two important point for the remote sensing spatial and spectral resolutions.

Resolution is the important issue for the remote sensing studies. The resolution of an image gives the possible detailed information provided by the imagery. In remote sensing, there are three types of resolution: spatial, spectral and temporal.

Spatial Resolution refers to the size of the smallest piece that can be detected by a satellite system or presented in a satellite image. It is given as a value symbolizing the length of one side of a square. For example, a spatial resolution of 30m means that one pixel represents an area 30 by 30 meters on the ground. [7] The display of low, medium and high spatial resolution objects **Figure 2** [8].

Spectral Resolution has to do with how a satellite sensor measures certain wavelengths of the electromagnetic spectrum. The finer the spectral resolution, the narrower the wavelength range for a particular channel or band [7].

Temporal resolution relates to the time among images. At our task, temporal resolution is not important, temporal differences are irrelevant as we do not expect to change the green ratio of a metropolitan over the course of minutes or days. We need constant image of the metropolitan we need high spatial and spectral resolution. We need LEO satellite imagery. The quality of the spatial and spectral resolution will define our approach (object based or pixel based) style.

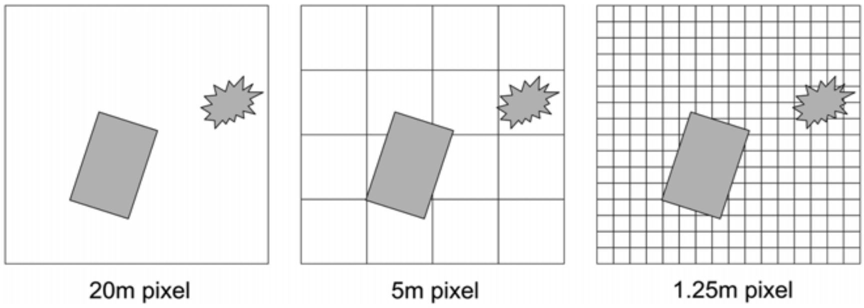

Figure 2. *Comparison of spatial resolutions*

In the comparison of methods have used the UC Merced dataset. UC Merced dataset has 100 images for each one of 21 classes. These are; agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, tennis court; **Figure 3.** Each image measures 256x256 pixels. The images were manually taken out from large images from the imagery collection for various urban areas around the country. The pixel resolution of this public domain imagery is 1 foot [9].

Figure 3. *UC Merced dataset*

3. Methods

There are two approaches for the remote sensing data; these are pixel based and object based. There are two aims; classify each green object and calculate the all green ratio. For the calculation of green ratio; best way is the unsupervised methods (clustering pixels). We can use unsupervised pixel based approach and detect green objects by the spectral resolution/bands. We can utilize some indexes evolved from spectral bands like NDVI. For the classification, we can use supervised methods, we need training data with labeled. We have to define spectral information for the each classes in the training data. We will define classes, forest, urban, lake Etc. and define spectral info. Each pixel will be interpreted numerically according to its spectral characteristics. We use some classification techniques to the define pixel's characteristics. Like; maximum likelihood classifier, minimum distance classifier. We can use object based classification method too for classification task. In this method, instead of pixel characteristics, we use segments.

Pixel-based classification methods analyze spectral properties of each pixel, whereas object based classification methods consider spectral, spatial and contextual information of segments [10]. An object-based approach deals with image segments, or "patches/objects of reality", rather than individual pixels [11].

In this study [11], pixel-based approach is used for to see change of the green area in the city, object-based approach is used to define the change as an expansion, shrinking, new or lost etc. by the size. Before the deep learning techniques, traditional pixel based methods were used mostly. Pixel based methods use spectral information. Object based methods use both spectral and spatial information. Object based approach use spectral characteristics and the class's shape, texture and their (neighbor pixel) spatial relationships. Pixel based classification is based solely on the spectral information in each pixel, object-based classification is based on information from a set of similar pixels called objects or image objects.

In object based approach; firstly, segmentation is done then the classification. Segmentation is the process that pixel grouping according to spectral and spatial characteristics for the classes. Scale is the very important term for object based approach. Three methods were examined in this review. Under the following headings; the results of the articles that made classification by applying the CNN, GAN and AE using UC Merced dataset were examined.

3.1. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is the one of the most commonly used deep learning model and its originally designed to process data in the form of multiple arrays [12] CNN is well-suited for processing multiband remote-sensing image data. CNN consists four different types of hierarchical structures: convolutional layers, pooling layers, fully connected layers and softmax layer. The architecture of CNNs; **Figure 4** [13] At each layer, the input image is convolved with a set of K kernels W = { W 1, W 2, , W K } and added biases = { b 1, , b K }, each generating a new feature map X k.

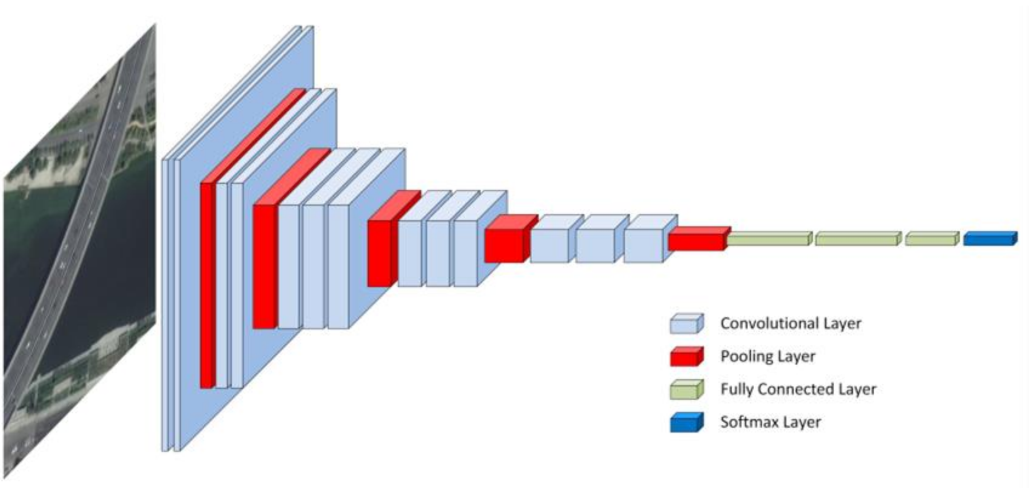

Figure 4. *Architecture of CNN*

There are a lot of articles to evaluate the different kinds of CNN networks. With GoogleNet, Xia et al. got 94.31 accuracy score. And with VGG-16 model, they got 95.21 accuracy score [14]. Zhang et al. used VGG-16-CapsNet and got 98.81 accuracy score **Table 1** [15].

Table 1. *Results of CNN models.*

CNN Model	Accuracy Score
GoogleNet	94.31
VGG-16 CapsNet	95.21

3.2. Generative Adversarial Networks (GANs)

Generative adversarial networks (GANs) [16] have become a very popular category of unsupervised deep learning models. GAN model has two sub-models. These are generative network and discriminative network; **Figure 5** [13].

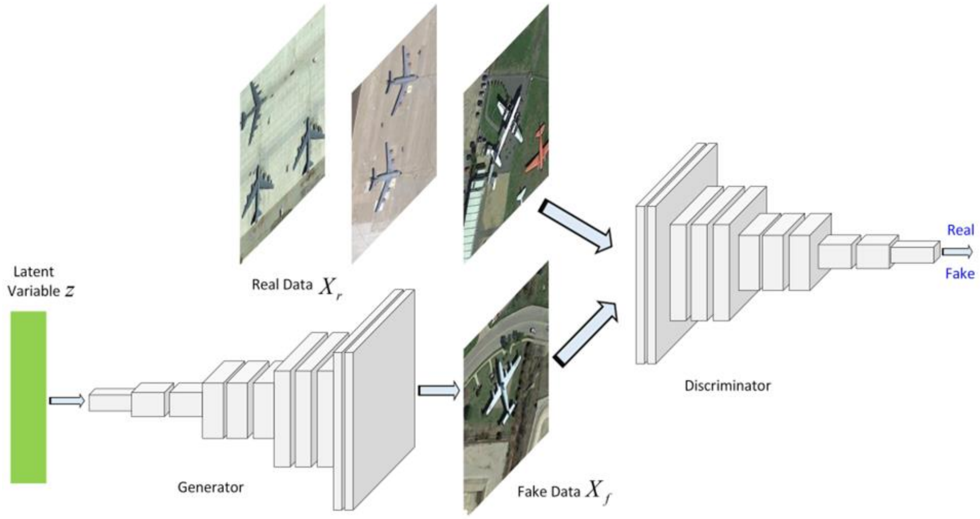

Figure 5. *Architecture of GAN*

Generally, the generative network is trained to generate samples that keep in line with real data in a manifold, whereas the discriminative network is trained to determine whether a sample is from real data or the generator. These two networks compete with each other so that the distribution captured from the generator is as similar as possible to the distribution of the real data. [17]

Reseachers have studied about GAN. There are several techniques applied to low-level computer vision tasks. According to the different kinds of model trainings, Lin et al. used MARTA GANs model and the got 94.86 accuracy score [18]. On the orther hand, Yu et al. used Attention GANs model and got 97.69 accuracy score in **Table 2** [19].

Table 2. *Results of GAN models.*

GAN Model	**Accuracy Score**
MARTA GANs	94.68
Attention GANs	97.69

3.3. Autoencoder (AE)

Autoencoder is one of the most popular models in deep learning which an unsupervised learning model is. It consists of symmetrical neural network and consists of input layer, hidden layer, and output layer. Autoencoder contains encoder and decoder. Encoding part is the reduction in the number of nodes at our hidden layers until bottleneck. And decoding part is the increasing in the number of nodes at our hidden layers after bottleneck. The architectures of (a) autoencoder and (b) stacked autoencoder; **Figure 6** [13].

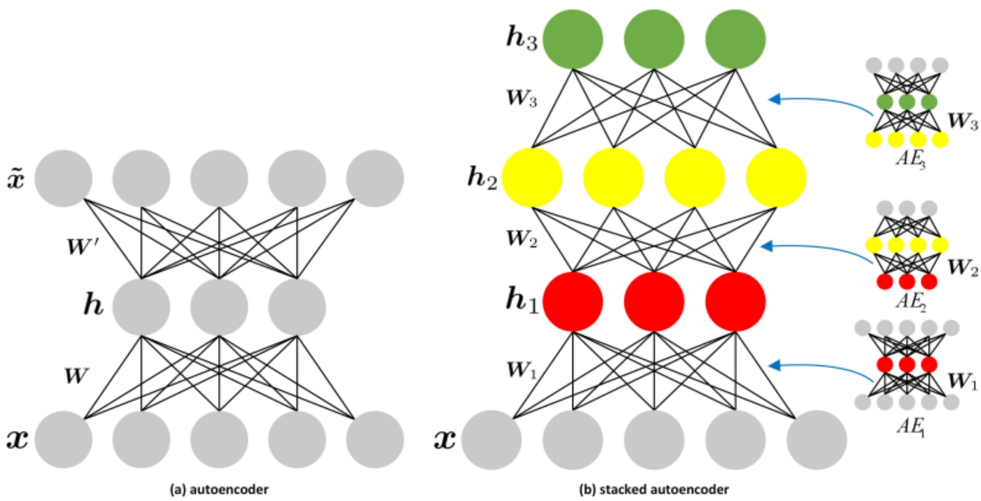

Figure 6. *Architectures of AE and SAE*

Autoencoder is also widely used in remote-sensing. There are some researches and their scores about autoencoders. Zhang et al. run Saliency-guided unsupervised feature learning (SGUFL) and they got 82.72 accuracy score. [20] Xiong et al. run Stacked convolutional denoising auto-encoders (SCDAE) model and they got 93.7 accuracy score. **Table 3** [21]

Table 3. *Results of AE models*

AE Model	Accuracy Score
SGUFL	82.72
SCDAE (Stacked)	93.7

4. Results

This review aims that to gather methods and approach styles to the 'how green the metropolitans?' task. The importance of the data set to be used and the studies using drone imagery as remote sensing instead of satellite imagery were mentioned. In line with the information provided by the satellites, 2 basic approaches (pixel and object based) were investigated. The articles in which 3 methods as CNN, GAN and AE were used on the UC Merced dataset were reviewed. As a result of the studies examined, the accuracy score success is shown below; **Table 4.** The best success has been achieved with GAN. However, the most used method in remote sensing data is CNN. [6]

Table 4. *Results of models*

Method	Accuracy Score
Attention GANs	97.69
VGG-16 CapsNet (CNN)	95.21
MARTA GANs	94.68
GoogleNet (CNN)	94.31
SCDAE (Stacked AE)	93.7
SGUFL (AE)	82.72

This review can be carried forward by using high-resolution satellite images of a particular metropolitan area as data and choosing the appropriate approach mentioned above, measuring both the green ratio and green classification of metropolitans.

References

[1] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, Brian Alan Johnson, Deep learning in remote sensing applications: A meta-analysis and review, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 152, 2019, Pages 166-177.

[2] Moreno-Armendáriz, M.A.; Calvo, H.; Duchanoy, C.A.; López-Juárez, A.P.; Vargas-Monroy, I.A.; Suarez-Castañon, M.S. Deep Green Diagnostics: Urban Green Space Analysis Using Deep Learning and Drone Images.Sensors 2019,19, 5287.

[3] Qin, R. An Object-Based Hierarchical Method for Change Detection Using Unmanned Aerial Vehicle Images.Remote Sens.2014,6, 7911-7932.

[4] Lu, H., Fu, X., Liu, C. et al.Cultivated land information extraction in UAV imagery based on deep convolutional neural network and transfer learning.J. Mt. Sci.14,731–741 (2017).

[5] Tetsuro Ishida, Junichi Kurihara, Fra Angelico Viray, Shielo Baes Namuco, Enrico C. Paringit, Gay Jane Perez, Yukihiro Takahashi, Joel Joseph Marciano, A novel approach for vegetation classification using UAV-based hyperspectral imaging, Computers and Electronics in Agriculture, Volume 144, 2018, Pages 80-85.

[6] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, Brian Alan Johnson, Deep learning in remote sensing applications: A meta-analysis and review, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 152, 2019, Pages 166-177, ISSN 0924-2716.

[7] Satellite Applications for Geoscience Education.

[8] T. Blaschke,Object based image analysis for remote sensing,ISPRS Journal of Photogrammetry and Remote Sensing,Volume 65, Issue 1,2010,Pages 2-16,ISSN 0924-2716.

[9] Yi Yang and Shawn Newsam, "Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification," ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), 2010.

[10] Elif Sertel & Ugur Alganci (2015): Comparison of pixel and object-based classification for burned area mapping

using SPOT-6 images, Geomatics, Natural Hazards and Risk.

[11] Jing Wang, Weiqi Zhou, Yuguo Qian, Weifeng Li, Lijian Han, Quantifying and characterizing the dynamics of urban greenspace at the patch level: A new approach using object-based image analysis, Remote Sensing of Environment, Volume 204, 2018, Pages 94-108.

[12] LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

[13] G. Cheng, X. Xie, J. Han, L. Guo and G. -S. Xia, "Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 3735-3756, 2020.

[14] Xia, G.S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A benchmark data set for performance evaluation of aerial scene classification. IEEE Trans. Geosci. Remote Sens. 2017, 55, 3965–3981.

[15] Wei Zhang 1,2, Ping Tang 1 and Lijun Zhao 1, Remote Sensing Image Scene Classification CNN-CapsNet, Remote Sens. 2019, 11, 494; doi:10.3390/rs11050494.

[16] Goodfellow, I., Abadie, J., Mirza, M., Xu, B., Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, arXiv: 1406.2661v1.

[17] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu and J. Jiang, “Edge-Enhanced GAN for Remote Sensing Image Superresolution,” in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 8, pp. 5799–5812, Aug. 2019, doi: 10.1109/TGRS.2019.2902431.

[18] D. Lin, K. Fu, Y. Wang, G. Xu, and X. Sun, “Marta gans: Unsupervised representation learning for remote sensing image classification,” IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 11, pp. 2092– 2096, 2017.

[19] Y. Yu, X. Li, and F. Liu, “Attention gans: Unsupervised deep feature learning for aerial scene classification,” IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 1, pp. 519–531, 2019.

[20] F. Zhang, B. Du, and L. Zhang, “Saliency-guided unsupervised feature learning for scene classification,” IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 4, pp. 2175–2184, 2014.

[21] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, “Stacked convolutional denoising auto-encoders for feature representation,” IEEE transactions on cybernetics, vol. 47, no. 4, pp. 1017–1027, 2016.