USAGE OF DEEP LEARNING FOR IMAGE SEGMENTATION FROM HIGH RESOLUTION IMAGES

O. Ozturk[a]*, D. Z. Seker[a], B. Bayram[b], Z. Duran[a]

[a] ITU, Civil Engineering Faculty, 34469 Maslak-Istanbul, TURKEY - (oozturk16, seker, duranza)@itu.edu.tr
[b] YTU, Civil Engineering Faculty, 34220 Esenler-Istanbul, TURKEY (bayram@yildiz.edu.tr)

KEYWORDS: Deep Learning, Feature Extraction, High Resolution Images, Convolutional Neural Network

ABSTRACT:

Image segmentation studies conducted by analyzing of aerial and satellite images play a key role in many studies, such as military applications, environment, agriculture, urban management and updates of Geographic Information System (GIS). In the remote sensing and photogrammetry, image segmentation studies are defined as determining whether the relevant signal contains one or more objects and finding the location of each object in the image. In this context, while extracting different objects such as cars, airplanes, ships and buildings which are independent from background and objects such as land use and vegetation classes which are also difficult to discriminate from the background can be extracted. Obtaining features by means of digitizing techniques in the digital base maps are time consuming process. Moreover, in image segmentation studies, generally, various difficulties are often encountered such as projection center error, image blockage, disorder of background, lighting, shading that cause to fundamental modifications in the appearance of features. Use of low resolution satellite or aerial images were insufficient to detect artificial and natural objects in the past years. With the development of technology, obtaining high spatial resolution satellite and aerial images contain detailed texture information become easier. Thus, the regional characteristics, artificial and natural objects can be perceived and interpreted. By now, many different operators have been used to automatically image segmentation from high resolution images. However, these methods require complex operation that have serious problems, such as incorrect or insufficient detection. In recent years, the deep learning approach has been started to be used in the discipline of photogrammetry and remote sensing widely for segmentation and detection. In deep learning, there is a structure based on the learning of multiple feature levels or representations of data and high-level features are derived from lower-level features to create a hierarchical representation. Although the image segmentation does not seem to be a very recent subject, it is necessary to successfully image segmentation from high resolution images, which is one of the important subject of today's Geomatics Engineering with the help of deep learning. In this study, the related methods, data sources and preliminary results of deep learning application on high resolution images to extract buildings and roads are widely discussed.

1. INTRODUCTION

In remote sensing and photogrammetry, image segmentation studies was widely used since 1980s as one of the main interested area which cover several difficulties to be overcome due to the features and methods. The aim of segmentation is to change the representation of an image into something that is more meaningful and easier to analyze. Compared with other image processing approach, image segmentation that is important in understanding the content of images and finding target objects is one of the most difficult task. In deep learning, computer vision and image processing, image segmentation is typically used to locate objects and boundaries in images. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity and texture (Linda and Stockman, 2001; Barghout et al., 2003).

Image segmentation in optical remote sensing images which include aerial and satellite images generally suffers from several increasing challenges such as shadow and occlusion. For this reason, large variations in the visual appearance of objects can be detected and extracted. Studies in this topic for overcame challenges are doing on since 1980s. Early studies, since low spatial images was used, manmade and natural objects couldn't be extracted from this images. With the advances of remote sensing and photogrammetry, high resolution images that have high spatial and texture resolution can be available such as worldview and unnamed aerial images. Thus, aside from region properties, a greater range of objects become recognizable and even can be separately identified (Cheng and Han, 2016).

During the last decades, considerable efforts have been made to develop various methods for the segmentation of different types of objects in satellite and aerial images such as roads, car, building, roof and airplanes. Deep learning and segmentation have been used several in academic area from remote sensing to environmental science (Figure 1).

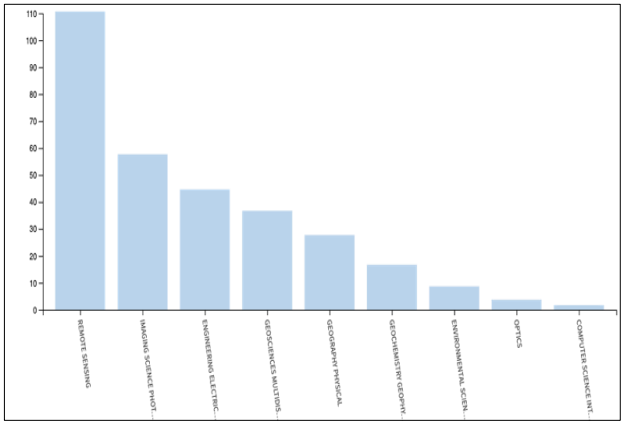

Figure 1. Deep learning based segmentation applications and record count in web of science (2010-2019) (URL-1)

By now, object detection and segmentation methods can be divided as matching-based methods, knowledge-based methods, machine learning-based and deep learning based. Many methods and algorithm have been used in this task but these methods can be complex processing and mislearn. With development of technology, learning based approach have begun to increase their effectiveness in such tasks. In these approaches, rules are obtained by processing of data and answers. Moreover, layered

representations learning and hierarchical representations of data can be carried out in deep learning (Cheng and Han., 2016).

In this study, deep learning based image segmentations were summarized. Additionally, Convolutional Neural Networks (CNN) based segmentation for roads and buildings from Unnamed Aerial Vehicle (UAV) based images in Istanbul Technical University Campus Area was examined and initial results were presented. Also, the challenges of current studies and proposed promising research directions in future was discussed.

2. DEEP LEARNING

In recent years, artificial intelligence has been common in both the media and the academic community. Machine learning, deep learning and artificial intelligence have been the subject of papers such as intelligent cities, meteorological estimates and change detection analysis. It is necessary to define artificial intelligence and machine learning before understanding that what deep learning is? Thus, the relationship and distinction between artificial intelligence, machine learning and deep learning must be clearly identified (Chollet, 2008). In Figure 2, the relationship among them are clearly indicated.

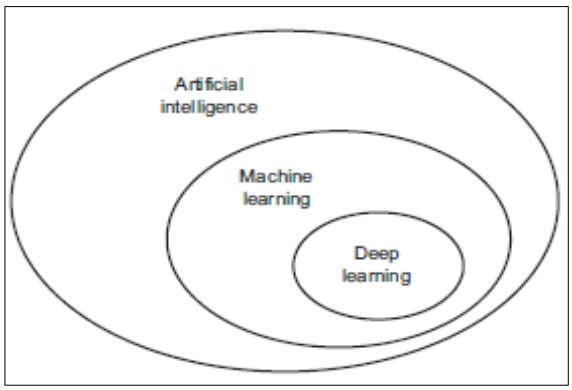

Figure 2. Artificial intelligence, machine learning, and deep learning (Chollet, 2008).

The Artificial intelligence which was born in 1950, emerged with the question of whether the problems solved by people can be solved by computers. More precisely, test is proposed to determine if a computer could think. The main goal is to automate the activities performed by people. In this context, artificial intelligence encompasses machine learning and deep learning. Early studies in artificial intelligence, due to manipulate the knowledge, programmers believed that a clear and broad set of rules should be created. This approach is known as symbolic artificial intelligence. Popularity increased with the rise of expert systems. For instance, chess that was one of the first applications of artificial intelligence, was able to solve logical problems quickly and accurately. However, it was insufficient to solve fuzzy logic problems such as image classification, speech recognition and language translation. Thus, machine learning was born newly to replace symbolic artificial intelligence(Chollet, 2008).

The basis of machine learning is formed by shaping artificial intelligence with the approach that computers can learn and have originality. Machine learning basically consists of questions such as whether the computer can learn to perform a specific task, when processing the data, whether the rules can be learned automatically. In classical programming rules and data are processed and results are produced. However, in machine learning, rules are obtained by processing of data and answers (Figure 3).

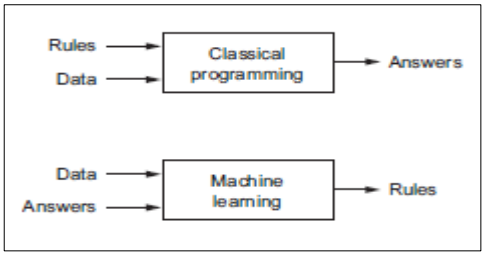

Figure 3. Classical programming and machine learning

A machine-learning system is trained rather than explicitly programmed. It is mostly important to train the networks for determining the model and rules in the machine learning approach. For instance, to automate traffic flow density, a machine-learning system would learn statistical rules for associating specific pictures with many different data such as weather, changing lighting and traffic condition (Chollet, 2008).

Although machine learning is tightly related to with mathematical statistics, machine learning distinguishes from mathematical statistics due to the logical processing structure. While the statistics are inadequate in studies such as processing, voice recognition and objet recognition, these complex sets of data can be analyzed and interpreted statistically using one of the methods used in machine learning. In another word, machine learning and deep learning are engineering oriented with little mathematical theory. It's a hands-on discipline in which ideas are proven empirically more often than theoretically (Chollet, 2008).

Meaningfully transforming data from input to output is the main problem in machine learning and deep learning. Machine learning algorithms aren't usually creative in finding these transformations; they're merely searching through a predefined set of operations, called a hypothesis space. Machine learning models are all about finding appropriate representations for their input data transformations of the data that make it more amenable to the task at hand, such as a classification task. Machine learning tend to focus on learning only one or two layers of representations of the data. Thus, they are sometimes called shallow learning. Although deep learning is a specific subfield of machine learning, modern deep learning consists of hundreds of layers of representations. The deep in deep leaning is how many layers contribute to a model of the data is called the depth of the model. More precisely, goal of deep leaning is layered representations learning and hierarchical representations(Chollet, 2008).

In particular, deep learning has been achieved successfully results compering others approaches such as near human level image classification, digital assistants and ability to answer natural-language questions etc. It is obvious that deep learning, which is becoming more common day by day, will open a new era in many fields from image classification to natural language processing.

Generally, deep learning methods can be divided into four categories according to the basic method. They are derived from: Convolutional Neural Networks (CNNs), Restricted Boltzmann Machines (RBMs), Autoencoder and Sparse Coding. The categorization of deep learning methods along with some representative works is shown in Figure 4.

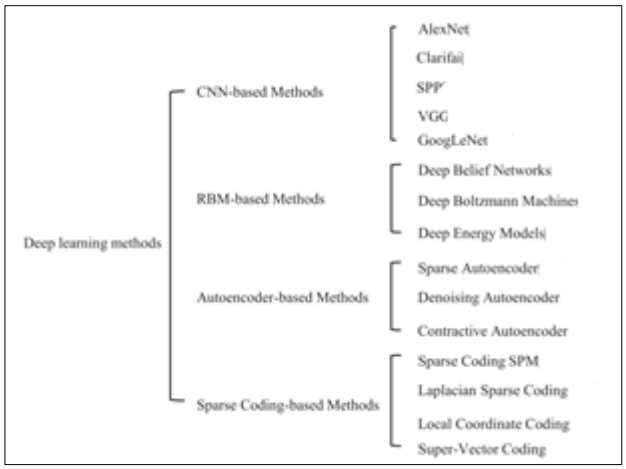

Figure 4. A categorization of the deep learning methods and their representative works (Guo et al., 2016).

The Convolutional Neural Networks (CNN) that it is the most commonly used computer vision applications is one of effective deep learning approaches. Multiple layers are trained in a robust manner in CNN. Generally, a CNN consists of three main neural layers, which are convolutional layers, pooling layers, and fully connected layers. Different kinds of layers show different roles. Deep learning has been widely adopted in various tasks of computer vision such as image classification, object detection, object recognition and image segmentation. Image segmentation is the most important task for deep learning as well as it is yielding promising for future studies. Successful results obtained from CNN models which are capable of tackling the pixel-level predictions with the pre-trained networks on large-scale datasets (Guo et al., 2016).

CNN-based image segmentation methods can be divided into 3 categories according to the basic method. They are derived from: detection based, fully convolutional networks (FCN) and weakly supervised segmentation. As for semantic segmentation, recent and advanced CNN based methods can be summarized as follows:

Detection-based segmentation is that segments images based on the candidate windows outputted from object detection (Chen et al., 2014). RCNN and SDS first generated region proposals for object detection, and then utilized traditional approaches to segment the region and to assign the pixels with the class label from detection (Girshick et al., 2014; Guo et al., 2016).

Fully Convolutional Neural Networks (FCNN) based segmentation, replacing the fully connected layers with more convolutional layers, has been a popular strategy and baseline for semantic segmentation (Chen et al., 2014). Long et al. (2015) defined architecture that combined semantic information from a deep, coarse layer with appearance information from a shallow, fine layer to produce accurate and detailed segmentations. Chen et al., 2014 proposed a similar FCN model, but also integrated the strength of conditional random fields (CRFs) into FCN for detailed boundary recovery.

Weakly supervised segmentation, researchers studied the more challenging segmentation with weakly annotated training data such as bounding boxes or image-level labels (Papandreou et al., 2015). Likewise, the BoxSup method made use of bounding box annotations to estimate segmentation masks, which are used to update network iteratively (Dai et al., 2015). These works both showed excellent performance when combining a small number of pixel-level annotated images with a large number of bounding box annotated images (Guo et al., 2016). There are various architectures of CNN and it is promised that the architecture will be faster and more accurate in foreseeable future. Some of them are LeNet, VGGNet, GoogleNet, ResNet and ZFNet.

3. METHODOLOGY USED

In this study, we aimed to test the performance of CNN based segmentation of roads and buildings details. Convolutional Network (CCN) learns a mapping from pixels to pixels, without extracting the region proposals. The main idea is to make the classical CNN take as input arbitrary sized images. The pipeline of the general CNN architecture is shown in Figure 5.

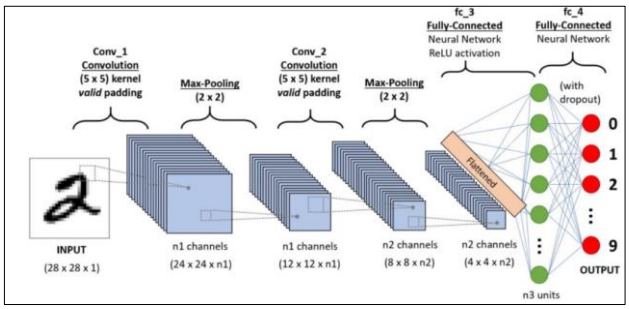

Figure 5. A CNN architecture sequence to classify digits (URL-2)

A CNN consists of three main neural layers, which are convolutional layers, pooling layers, and fully connected layers. Different kinds of layers play different roles. A general CNN architecture for image classification is shown layer by layer. There are two stages for training the network: a forward stage and a backward stage. First, the main goal of the forward stage is to represent the input image with the current parameters (weights and bias) in each layer. Then the prediction output is used to compute the loss cost with the ground truth labels. Second, based on the loss cost, the backward stage computes the gradients of each parameter with chain rules. All the parameters are updated based on the gradients, and are prepared for the next forward computation. After sufficient iterations of the forward and backward stages, the network learning can be stopped. At the end, using activation functions segments are produced and classifiers label the feature which are extracted as building or road.

4. CASE STUDY

In this study, segmentation process by means of deep learning has been realized using high resolution UAV images for extraction of buildings and roads. In the study, CNN architecture based on the Tensorflow library was implemented.

The images are divided into training images, testing images and validation images in a predefined directory structure. The second step is training model by training dataset that consist of satellite images and its binary images. Tensorflow initialize a deep neural network of four hidden layers of sizes 100, 150, 100, 50 neurons respectively, then loads the training and testing, and use it to train the neural network for a predefined number of iterations, the testing data is used to evaluate the accuracy of the model at the end of the training.

Classification of an input image is realized, it takes the names of the input and output image names as command line arguments, it

loads the model and the input image from the 'image-input' directory, generate the feature vector for each pixel, classify it using the loaded classifier, and generate the output image with zero/one value for each pixel based on the classifier prediction for that pixel, then it saves the generated image in 'image-output' directory using the output image name entered in the command line arguments. In the study segmentation accuracy has been calculated as 83%.

Together with the original image used in the study and obtained results in different color combination were displayed on the Figure 6.

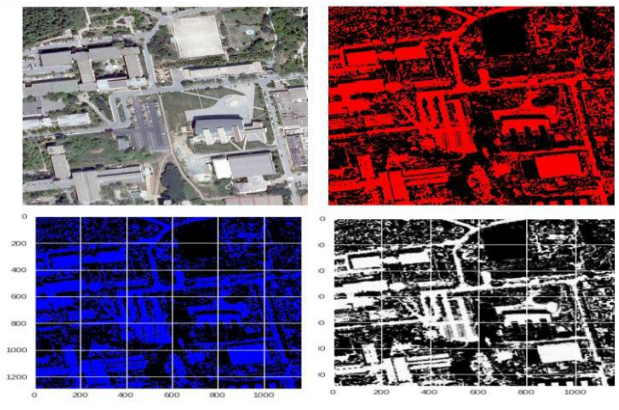

Figure 6. Extracted roads and buildings from the original image used in the study

5. CONCLUSIONS

Although current studies have achieved promising results, it is clear that deep learning is an important tool for further progress. However, the GPU based problems of the deep learning must be overcome. Thus, training and classification time can be reduced.

In the current studies, especially in developed countries mostly regular areas such as urban areas are selected as the test area. Thus, using these data set as the training data may bring several problems for other counties such as Turkey. The study areas in Turkey may differ from the other countries and this may result with the problem of extracting the linear features such as roads in rural areas, which might be very difficult. For this reason, the data sets and rules should be considered and created for every county separately according to their land use/cover characteristics.

ACKNOWLEDGEMENTS

The authors would like to thank to Volodymyr Mnih (University of Toronto) for his permission to use his labelled data set. They also would like to thank to Mahmoud Mohsen for helpful suggestions contributed directly through GitHub.

REFERENCES

Barghout, L., and Lee, L. 2004. U.S. Patent Application No. 10/618,543.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062.

Cheng, G., and Han, J. (2016). A survey on object detection in optical remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 117, 11-28.

Chollet, F. 2018. Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek. MITP-Verlags GmbH & Co. KG.

Dai, J., He, K., and Sun, J. (2015). Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1635-1643).

Girshick, R., Donahue, J., Darrell, T., and Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., and Lew, M. S. (2016). Deep learning for visual understanding: A review. Neurocomputing, 187, 27-48.

Hariharan, B., Arbeláez, P., Girshick, R., and Malik, J. (2015). Hypercolumns for object segmentation and fine-grained localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 447-456).

Mnih, V. (2013). Machine learning for aerial image labeling. University of Toronto (Canada).

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

Papandreou, G., Chen, L. C., Murphy, K. P., and Yuille, A. L. (2015). Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In Proceedings of the IEEE international conference on computer vision (pp. 1742-1750).

Shapiro, L., and Stockman, G. (2001). Computer Vision, pp Prentice-Hall. New Jersey, USA.

URL–1: http://apps.webofknowledge.com

URL-2: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53