Erzincan Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
2021, 14(1), 322-330
ISSN: 1307-9085, e-ISSN: 2149-4584
Araştırma Makalesi

Erzincan University
Journal of Science and Technology
2021, 14(1), 322-330
DOI: 10.18185/erzifbed.891322
Research Article

Multi-Scale Aircraft Detection from Satellite Images

Hussein M.A. Mohammed, Merve Polat, Abdullatif Ali Tahlil, İbrahim Yücel Özbek*

Department of Electrical and Electronic Engineering, Faculty of Engineering
Ataturk University, Erzurum, Turkey

Geliş / Received: 04/03/2021, Kabul / Accepted: 29/03/2021

Abstract

Satellite image analysis is a research area in which many research studies are carried out for civil and military applications in the field of image processing. Satellite imagery has many applications including recognition, detection and classification of regions, buildings, roads, aircraft and other man-made objects. Among these, especially aircraft detection is strategically important for military applications, and forms the basis of this study. In the first phase of the study, a new dataset of aircrafts is created from Google Earth images to compensate the shortage of data set in this area. In the second stage, the detection of air vehicles was carried out using algorithms based on Convolutional Neural Network (CNN). Region-based Fully Convolutional Network (R-FCN), Single Shot Multi Box Detector (SSD) and Faster R-CNN methods are used for this process. The obtained accuracy rate for R-FCN, SSD and Faster R-CNN are 98.01%, 69.71% and 96.56%, respectively.
Keywords: Aircraft Detection, Satellite Image Analysis, Object Detection, Region-based Fully Convolutional Network (R-FCN), Single Shot Multi Box Detector (SSD) and Faster Region-based Convolutional Neural Network (Faster R-CNN).

Uydu Görüntülerinden Çok Ölçekli Uçak Tespiti

Öz

Uydu görüntü analizi, görüntü işleme alanında sivil ve askeri uygulamalar için birçok çalışmanın yapıldığı geniş bir araştırma alanıdır. Uydu görüntüleri, bölgelerin, binaların, yolların, uçakların ve diğer insan yapımı nesnelerin tanınması, algılanması ve sınıflandırılması olmak üzere birçok uygulamada kullanılmaktadır. Uçak tespiti özellikle askeri uygulamalar için stratejik öneme sahiptir ve bu çalışmanın temelini oluşturmaktadır. Bu çalışmada uçak tespiti uygulamalarına yönelik veri seti eksikliğini telafi etmek için Google Earth görüntülerinden yeni bir uçak veri seti oluşturulmuştur. Hava araçlarının tespiti için Konvolüsyonel Sinir Ağlarına (CNNs) dayalı algoritmalar kullanılmıştır. Uçak tespitine yönelik yapılan deneylerde Bölge Tabanlı Tam Evrişimli Ağ (R-FCN), Tek Atışlı Çoklu Kutu Detektörü (SSD) ve Daha Hızlı R-CNN (Faster R-CNN) yöntemlerinin performansı karşılaştırılmıştır ve sırasıyla % 98.01,% 69.71 ve % 96.56 doğruluk oranları elde edilmiştir.
Anahtar Kelimeler: Uçak Algılama, Uydu Görüntü Analizi, Nesne Algılama, Bölge Tabanlı Tam Evrişimli Ağ (R-FCN), Tek Atışlı Çoklu Kutu Detektörü (SSD) ve Daha Hızlı Bölge-tabanlı Evrişimli Sinir Ağı (Daha Hızlı R-CNN)

*Corresponding Author: iozbek@atauni.edu.tr

1. Introduction

Extensive research on object detection with deep convolutional neural networks have dominated the scientific world for the past few years. Detecting the presence of object instances and locating them accurately in images are proved to be quite challenging. However, recent advances in the methods driving object detection –like region based convolutional neural networks and region proposal methods – achieve almost accurate and real-time detection rates. These deep convolutional neural networks require huge datasets for training, and as a result, new large datasets are being introduced every now and then. In this work, ATA-Plane, a new dataset comprising of 2705 aerial images of planes, is created and used for both training and detecting airplane instances in images. The detection of aircraft plays a vital role in making more sound and successful decisions in both military and civil applications. In the literature, the detection of aircraft by low-resolution infrared sensors (Anon 2010), the hierarchical classification algorithm to accurately recognize the aircraft from satellite images(Hsieh et al. 2005) and the real-world conditions based on the use of a hierarchical database of object models are among the studies conducted for aircraft detection system (Das et al. 1994). In this study, different object detection algorithms like Faster Region-based Convolutional Neural Network (Faster R-CNN), Region-based Fully Convolutional Networks (R-FCN) and Single Shot Detector (SSD) will be used to evaluate our dataset, ATA-Plane.

This study is organized as follows: in Section II, previous work in this field and object detection algorithms will be discussed. In section III, dataset preparing and labeling will be presented. In the last section, shortcomings of our work, suggestions and future studies will be presented.

2. Material and Methods

2.1. Object Detection

The size, position and background of objects is one of the greatest hindrances of detecting objects accurately. Viola-Jones has proposed a machine learning approach that has the ability to locate object instances very fast(Viola and Jones 2001). Using HOG attributes, Navneet Dalal and Bill Triggs proposed a method for detecting pedestrian in images (Dalal and Triggs 2005). A model based on multi-scale deformation component detection, known as Deformable Part Model (DPM), was proposed by Felzenszwalb (Felzenszwalb et al. 2010). For the past few years, deep convolutional neural networks such as Region-based Convolutional Network (R-CNN), proposed by Girschick et al. in 2014 (Girshick et al. 2014), with relatively good performance have inspired researchers to fine tune them and come up with algorithms which are faster and more accurate than them without building everything from scratch. Fast R-CNN(Girshick 2015), an improved version of R-CNN, and R-CNN both use selective search algorithm which makes them incompetent for real time object detection applications. Faster R-CNN, which is a modified version of Fast R-CNN, is not only faster than R-CNN and Fast R-CNN but also more accurate due to fact that it uses Region Proposal Network (RPN) rather than selective search algorithm(Ren et al. 2017).

2.1.1. Single Shot Multi Box Detector (SSD)

SSD (Liu et al. 2016) is an important algorithm that is especially used to detect objects in real time problems and widely

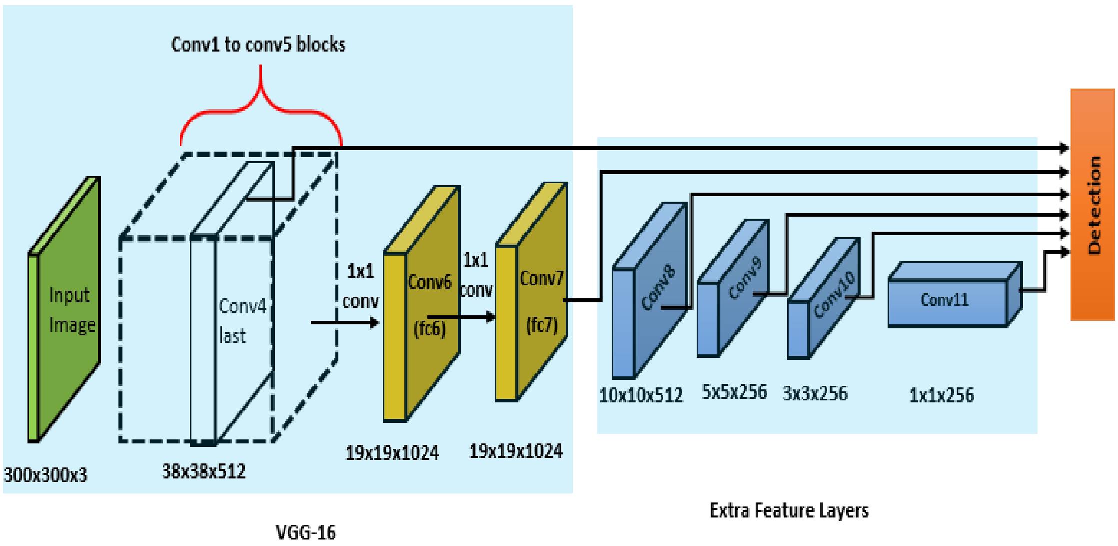

Figure 1. Single Shot Multi-Box Detector (SSD) Architecture

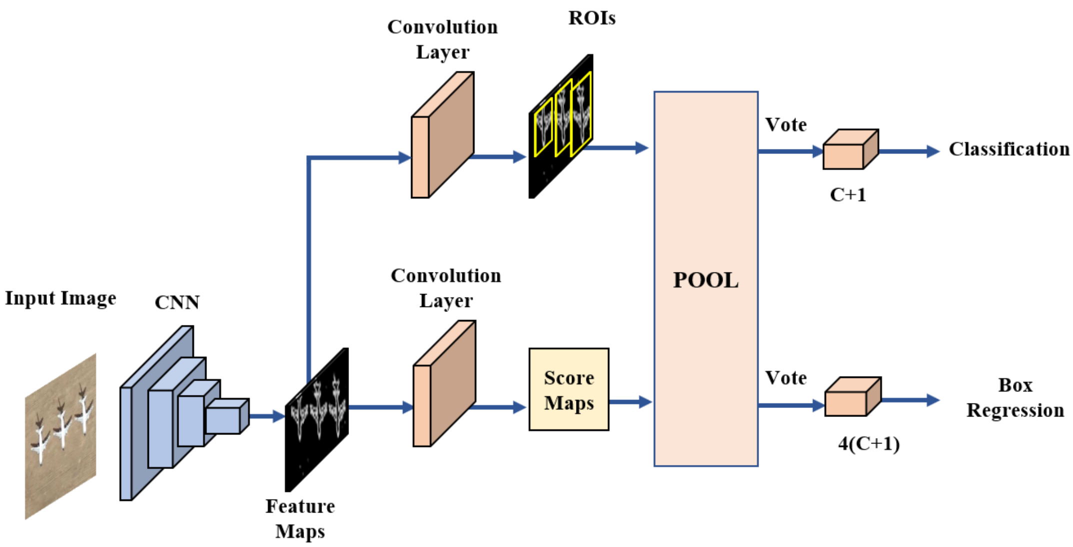

Figure 2. Region-based Fully Convolutional Networks (R-FCN) Flowchart

preferred. Faster R-CNN defines a regional proposal network by creating a bounding box and uses these regional proposals to classify objects. While two shots are required for Regional proposals and feature extraction in R-FCN and Faster R-CNN, in SSD these two are combined to one deep neural network step.

The feature that makes the SSD detector different from other single-shot detectors is the use of multiple layers that provide more accurate detection of objects of different scales. The foundation of the SSD architecture that can be used with any deep network-based model, such as ResNet (He et al. 2016) and Inception,(Pandit et al. 2020) is based on VGG16(Simonyan and Zisserman 2014). Some additional layers are also embodied for processing large objects. While the first convolution layer of the network architecture is responsible for detecting the smallest objects, the last convolution layer is responsible for detection the largest objects. Single Shot Multi-Box Detector (SSD) Architecture is shown in Fig.1.

.

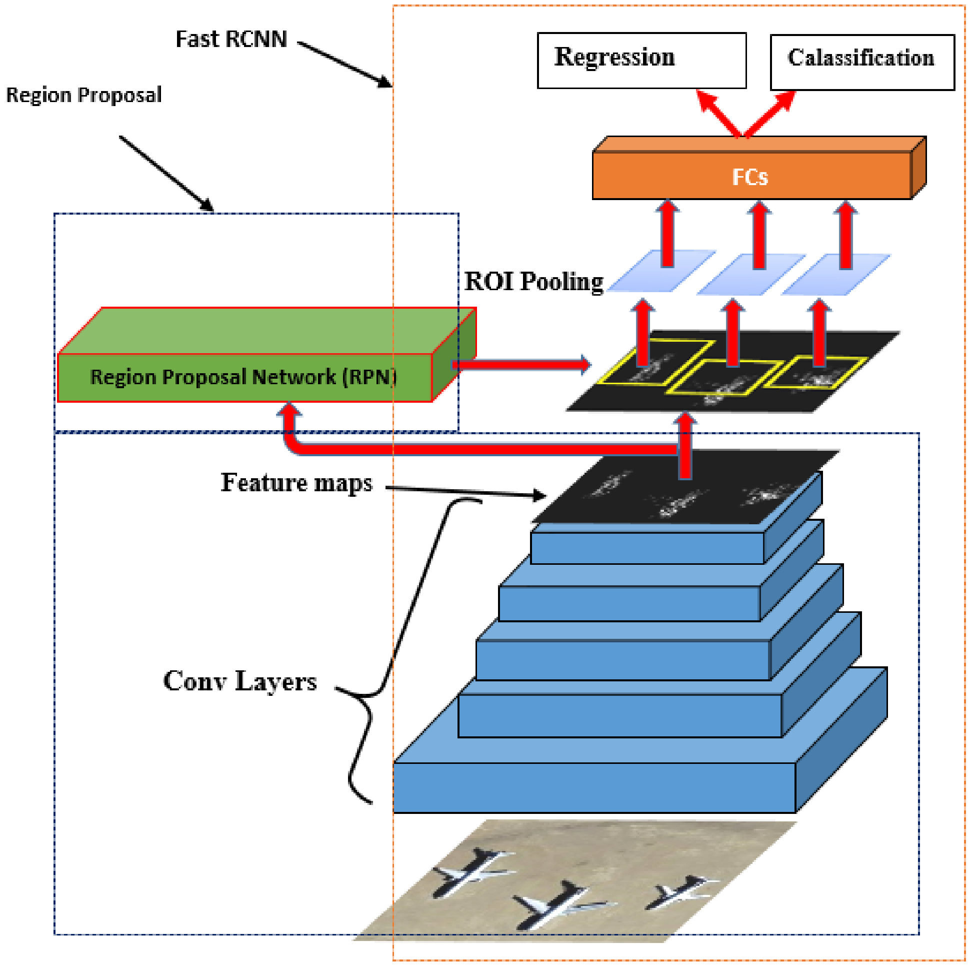

Figure 3. Faster R-CNN Architecture

2.1.2. Region-based Fully Convolutional Network (R-FCN)

R-FCN is an object detection technique based on convolutional neural network.
The main difference between R-FCN and R-CNN series is the removal of FC layers that are located after the RoI pooling layer. Also, in R-FCN, position-sensitive score maps are proposed to enable FCN efficiently represent translation variance. These changes made R-FCN 2.5-20 times faster than the Faster R-CNN (Networks and Dai 2016) R-FCN Architecture shown in Fig. 2.

2.1.3. Faster R-CNN

Faster R-CNN is a further development of fast R-CNN, and the biggest difference between the two is the determination of Regions of Interest, RoI. While Fast R-CNN uses selective search algorithm to detect regions, Faster R-CNN uses Region Proposal Network (RPN) for this task. Faster R-CNN consists of four steps. Extraction of feature map from the input image in the first step. In the second step, object proposals with their objectivness score are produced by the application of RPN on the extracted feature maps. Then a RoI pooling layer is applied on these proposals to bring them to the same size.

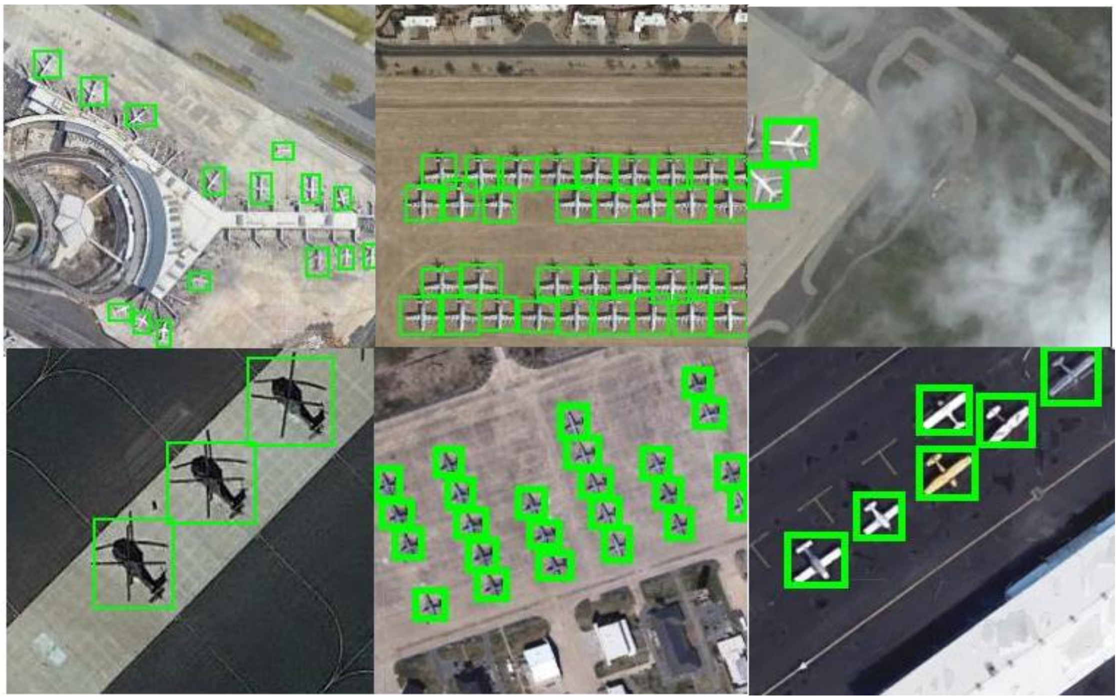

Figure 4. Examples of labeled images

of the ConvNets. Finally, in the fourth step, classification and generation of the bounding boxes for objects are carried out in the last layers (softmax layer and a linear regression layer) of the network. Faster R-CNN Architecture is depicted in Fig. 3

2.2. Datasets

The use of large datasets such as ImageNet (ImageNet 2016) and MSCOCO (Lin et al. 2014) played a vital role in the success of deep learning research studies especially in object detection and classification areas. However, it is not likely to have datasets with enough number of images tuned for specific applications such as aerial problems. Recently, many studies have been conducted in satellite imagery, and various data sets have been created. Some of these include but are not limited to NWPU VHR-10, NUDT Aircraft, NUDT Aerial-Vehicle and NUDT SAR-Ship (Deng et al. 2018), DOTA, AID. DOTA is a Large-scale Dataset for Object Detection in Aerial Images (Xia et al. 2018). This dataset consists of 15 classes of aircraft, helicopter, port, ship, swimming pool, storage tank etc. AID(Xia et al. 2017) also consists of aerial images.

As far as aircraft detection problem is concerned, there are very few publicly available aircraft datasets. The shortcomings of these datasets are the large-scale high-resolution of aircrafts images and the small number of aircraft instances in the dataset. For this reason, we have built a new dataset, namely ATA-Plane. It consists of mostly fixed-wing civil and military aircrafts along with rotary-wing samples. The following two sections briefly introduce this data set.

2.2.1. Image Collection

The ATA-Plane data set containing 2705 satellite images is an extended version of our previous dataset (Polat et al. 2019) containing 1030 images . This data set was created using Google Earth. Each dataset image contains at least one target of different sizes collected from different parts of the world with 8584 instances. In order to increase the number of images in the dataset, we also captured images of same locations in different times using Google Earth historical imagery feature.

2.2.2. Image Labeling

A bounding box surrounding the target objects in each dataset sample is annotated by hand. Matlab Image Labeler APP and the LabelImg (GitHub 2017) programs were used for the annotation. Fig. 4 shows some labeled samples of ATA-Plane dataset. The parameters used in Matlab Image Labeler APP to define the bounding boxes is a vector in [x y width height] format, which correspond to the upper left corner horizontal and vertical coordinates, width and height values of the bounding box, respectively. On the other hand, $[x_{min}\ y_{min}\ x_{max}\ y_{max}]$ format was used in the *LabelImg* program.

3. Resarch Findings

ATA-Plane dataset contains 2705 images of which 2401 were used for training while the remaining 304 were used for testing. Precision-Recall curves were obtained to evaluate the object detection performance of all models. The Jaccard index is also used to measure the similarity between the ground truth and estimated regions. Jaccard index represent Intersection over Union between these two bounding boxes (IoU).

The Precision-Recall values are calculated as follows:

$$Precison = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

where TP (True Positive) is the number of correctly detected aircraft; FP (False Positive) is the number of cases where aircraft was identified mistakenly and FN (False Negative) is the number of missed targets.

A comparison of Precision- Recall curves for Faster R-CNN, R-FCN and SSD models with a threshold of 0.5 are shown in Fig. 5. Also, the obtained Precision-Recall curves at different threshold values are shown in Fig.6 for all object detection models used in this work. Fig.7 illustrates some test results of different images from the test data. In this figure, green boxes represent the ground truth of target objects while yellow boxes Show The correct detections of the different models used in this study.

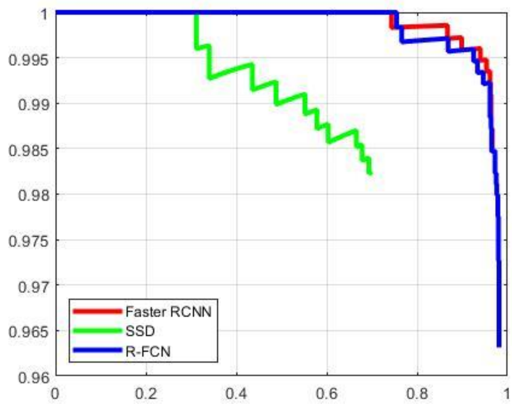

Figure 5 Precision - Recall curve at 0.5 threshold for (Faster R-CNN, R-FCN and SSD.

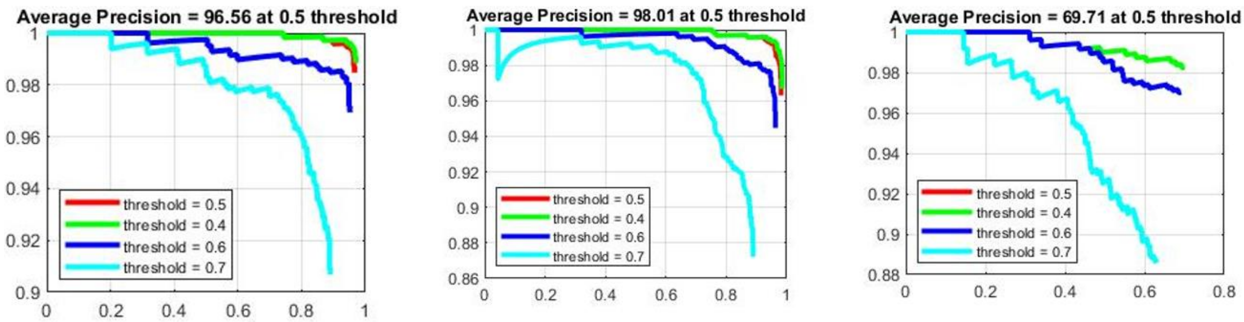

Figure 6. Precision - Recall curve (Left =Faster R-CNN, Center=R-FCN, Right=SSD)

Figure 7. Testing the model on images from test Data

4. Results

In this study, a data set with aircraft satellite images taken from Google Earth at different times and locations was created. This data set consists of a total of 2705 images, 2401of which are allocated for training whereas the remaining 304 are used for testing. SSD, a very popular real time object detection algorithm, faster R-CNN and R-FCN were trained on our dataset. Performance comparison of R-FCN, SSD and Faster R-CNN algorithms was carried out and their accuracy rates were 98.01%, 69.71% and 96.56% respectively. In future studies, the ATA-plane dataset will be increased. Also, the current state of the art object detection algorithms will be fine-tuned to improve their performance and new faster and more accurate ones will be designed.

References

Anon. 2010. “DETECTING AIRCRAFT WITH A LOW RESOLUTION INFRARED SENSOR J ´ LTCI , UMR 5141 37-39 Rue Dareau DOTA MPSO Chemin de La Huni ` Ere 91761 Palaiseau.” 2475–77.

Dalal, Navneet, and Bill Triggs. 2005. “Histograms of Oriented Gradients for Human Detection.” in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*.

Das, Subhodev, Bir Bhanu, Xing Wu, and R. Neil Braithwaite. 1994. "System for Aircraft Recognition in Perspective Aerial Images." *IEEE Workshop on Applications of Computer Vision - Proceedings* 168–75. doi: 10.1109/acv.1994.341305.

Deng, Zhipeng, Hao Sun, Shilin Zhou, Juanping Zhao, Lin Lei, and Huanxin Zou. 2018. "Multi-Scale Object Detection in Remote Sensing Imagery with Convolutional Neural Networks." *ISPRS Journal of Photogrammetry and Remote Sensing*. doi: 10.1016/j.isprsjprs.2018.04.003.

Felzenszwalb, Pedro, Ross Girshick, David Mcallester, and Deva Ramanan. 2010. "Object Detection with Discriminatively Trained Part-Based Models." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32:1627–45. doi: 10.1109/TPAMI.2009.167.

Girshick, Ross. 2015. "Fast R-CNN." in *Proceedings of the IEEE International Conference on Computer Vision*.

Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

GitHub. 2017. "LabelImg." Retrieved (https://github.com/tzutalin/labelImg).

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. *Deep Residual Learning for Image Recognition*.

Hsieh, Jun-Wei, J. M. Chen, Chi-Hung Chuang, and K. C. Fan. 2005. "Aircraft Type Recognition in Satellite Images." *Vision, Image and Signal Processing, IEE Proceedings* - 152:307–15. doi: 10.1049/ip-vis:20049020.

ImageNet. 2016. "ImageNet." Retrieved (http://www.image-net.org/).

Lin, Tsung Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. "Microsoft COCO: Common Objects in Context." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8693 LNCS(PART 5):740–55. doi: 10.1007/978-3-319-10602-1_48.

Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng Yang Fu, and Alexander C. Berg. 2016. "SSD: Single Shot Multibox Detector." in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

Networks, Region-based Fully Convolutional, and Jifeng Dai. 2016. "R-FCN : Object Detection Via." *ArXiv Preprint*.

Pandit, Tejas, Akshay Kapoor, Rishi Shah, and Rushi Bhuva. 2020. *UNDERSTANDING INCEPTION NETWORK ARCHITECTURE FOR IMAGE CLASSIFICATION*.

Polat, M., H. M. A. Mohammed, E. A. Oral, and I. Y. Ozbek. 2019. "Aircraft Detection from Satellite Images Using ATA-Plane Data Set." Pp. 1–4 in *2019 27th Signal Processing and Communications Applications Conference (SIU)*.

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2017. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence*. doi: 10.1109/TPAMI.2016.2577031.

Simonyan, Karen, and Andrew Zisserman. 2014. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *ArXiv 1409.1556*.

Viola, Paul, and Michael Jones. 2001. *Rapid Object Detection Using a Boosted Cascade of Simple Features*. Vol. 1.

Xia, Gui Song, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. 2018. "DOTA: A Large-Scale Dataset for Object Detection in Aerial Images." in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Xia, Gui Song, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu. 2017. "AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification." *IEEE Transactions on Geoscience and Remote Sensing*. doi: 10.1109/TGRS.2017.2685945.