Object Detection Methods for Improving UAV Autonomy and Remote Sensing Applications

Panagiotis Aposporis
School of Electrical and Computer Engineering, National Technical University of Athens.
9, Iroon Polytechniou St, PC: 157 80, Athens, Greece

***Abstract*— The last decades the Unmanned Aerial Systems (UASs) are being used in a variety of applications, such as civil protection, security, agriculture, armed forces, that need real time object detection of observed information by their sensors. Moreover, the development of fully autonomous UAS is heavily dependent on their capability to detect and track steady or moving objects in a robust, powerful and reliable manner. In this review, we present a comprehensive literature survey and discussion on object detection methodologies for improving UAV autonomy and remote sensing applications. Emphasis is placed on Convolutional Neural Networks (CNN) implementing different object detectors and exploiting cloud processing. Based on these works, we provide a brief discussion and summary of related proposals for UAV-based object detection using different methodologies and approaches, share views for future research directions and draw conclusive remarks.**

Keywords—Deep learning, drones, machine learning, neural networks, object detection, remote sensing, unmanned aerial vehicle.

I. INTRODUCTION

During the recent years, the widespread proliferation of Unmanned Aerial Systems (UAS or drones) among both state and non-state actors, has remarkably influenced remote sensing and object detection capabilities, in many areas, such as search and rescue, surveillance, inspection. With increasingly affordable, agile, flexible and available modern technologies almost all government and private domains, are exploiting drones' usage to on-line detection, identification and tracking of target objects. One particularly important application domain where UASs could be proved a very powerful asset is in the area of natural disasters, such as floods, volcano explosions, earthquakes or man-made disasters caused by terrorists or anarchists. In such cases, emergency response authorities often require a real-time situation awareness capability to monitor the development and the status of reactions in the affected area. Unmanned aerial vehicles equipped with the appropriate technologies offer an ideal solution for developing the necessary situation awareness in order to support decision making. Due to the unknown dangers of the specific environment there are no other options in getting the necessary information.

Traditional command and control of drones, for Line of Sight (LOS) or Beyond LOS (BLOS) flights is mainly based on radio communications between the unmanned vehicle and the ground station. Recently the focus has been on both high autonomy, incorporating Artificial Intelligence capabilities with or without cloud-based technologies and low-level autonomy, using traditional control and advanced avionics systems and their integration in distributed software architectural frameworks which support robust autonomous operation in complex operational environments such as those in disasters [1]. In any case the cornerstone of the UAS Autonomy, which is a huge challenge nowadays, is the development of a robust, powerful, reliable and simultaneously real-time capability to detect and track steady or moving objects. Even though this challenge has already been studied and implemented in many commercial products (i.e. DJI Smart Track function) the UAV domain sets additional restrictions such as the size and weight of the hardware which limits the computational capacity, the low-quality images due to the distance of the sensor from the target and the high pose variations, especially for humans. Relatively larger objects such as buildings, vehicles, large animals can be detected in optical images whilst locating people could be a relatively more difficult problem due to the small target size, low contrast of the target with the background or presence of clutter [2]

Object detection using aerial images is a very challenging and complex task but this task is even more difficult when using real time images from UAVs. The key challenges in deploying an "Object Detection Component" (ODC) on a UAV platform are as follows:

- Power. The ODC should consume minimal power in order to minimize its effect on battery consumption and flight time of the drone
- Computational load. The ODC should have low processing demands and less memory requirements as typical commercial UAVs have resource limitations.
- Size & weight. All parts (acquisition unit, processing unit etc.) of the ODC and their support equipment should be light-weight, small and flexible to be attached on a drone's frame without affecting its aerodynamic characteristics

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00 © 2020 IEEE

- Latency. Communication between components and processing the input data with low latency and faster performance for utilization in real time.

There are mainly two different approaches to overcome the above drone limitations, foregrounding specific advantages and disadvantages. One solution is to keep on board this high computational load of modern computer vision techniques such as deep learning, by exploiting parallel architectures with prominence of GPUs as accelerators. This approach offers fast processing and independence of external communications but is heavily dependent on the size and the produced power of the unmanned vehicle. The second one is transferring the heavy processing burden to the ground station or to a cloud-based solution, which, however, poses additional challenges such as availability and stability of networks, latency in communication links as well as security risks. Notwithstanding, this approach ensures dynamic computational resources capable to handle very demanding models and algorithms. [7]

The overarching aim of this paper is to provide an integrated review of the recent progress in the area of object detection with special focus on unmanned aerial vehicles and their challenges. Different from previously published reviews which concentrated on object detection methods implemented on "large vehicles" (fixed ground devices, cars, airplanes etc.), the significance of this paper is that it addresses the research gap on the challenges of small-scale UAVs with limited computational resources that can incorporate object detection. Presenting and comparing an indicative number of publications this review will be helpful for the researchers to have better understanding of object detection challenges posed by UAVs. This paper begins with a description of UAVs characteristics and main challenges related to remote sensing, while next presents a basic background about remote sensing focusing on UAV sensors. It also describes the different object detection methodologies and then provides fundamentals of Convolutional Neural Networks and the main object detectors categories. It also presents a qualitative comparison of the techniques implementing CNN-based object detection both on-board and off-board, and finally concludes with the challenges and directions for future research.

II. UAVs CHARACTERISTICS AND CHALLENGES

UAVs can be categorized in a variety of ways based on vehicle attributes, such as size and weight, aerodynamic features (fixed wing or rotorcraft) or flight characteristics (altitude, speed, etc.). In general, larger aircraft use larger and more powerful engines that provide higher altitude, longer endurance and more payload capacity than smaller vehicles. Different organizations (NATO, European Union, NASA) each have defined UAV classifications based on weight and altitude or speed. Classification of UAV platforms for civil or scientific applications has generally followed existing military descriptions of the platforms with characteristics such as size, flight endurance, and capabilities.

There are two types of UAVs that are most widely investigated and developed considering their aerodynamic features. Fixed-wing UAVs have been popular and commonly used for a variety of applications, particularly for long distance/long endurance tasks [11], [12] and for increased weight-lifting capacity (cargo, agriculture, military equipment etc.). Rotary-wing UAVs have many unique capabilities such as vertical take-off and landing, hovering and high level of maneuverability [22]. The last decade there has been an increase in interest on rotary UAV specifically multirotor because of the ease of construction and control [23] [24] and the rapid development of microelectronics. Compared to a conventional helicopter, a multicopter (more than two rotors) is more suited for flight that requires high maneuverability and simple control. The most common types of configuration for a multicopter is quadcopter, hexacopter, and octocopter. These platforms use multiple sensors and advanced electronic control system to stabilize the aircraft [25] [26] facilitating the control of the remote pilot.

Being a platform for remote sensing, multicopters offer better stability to the photographic equipment as opposed to conventional single-rotor helicopters due to their omission of a vertical tail rotor and complex mechanical components that adjust the pitch of the fast-spinning primary blade. Additionally, the placement of rotors on the periphery of multicopters allows more room for both housing the gimbal and the camera in the center of the vehicle. Their simpler structure and their increased flight stability make multicopters easier to operate and maintain as well as less costly to acquire and modify. The inherent advantages associated with multicopters combined with measuring capabilities render them ideal for a diversity of remote sensing applications. Table 1 summarizes the benefits and drawbacks of various types of UAVs, focusing on remote sensing applications.

III. REMOTE SENSING AND UAV SENSORS

Remote sensing is the science of obtaining information about objects or areas on the Earth from a distance, using images acquired from airborne or spaceborne vehicles by measuring reflected or emitted electromagnetic radiation [39]. Based on their operation, remote sensors can be either passive or active. Passive sensors detect and record natural energy that is reflected or emitted from the Earth's surface. The most common source of radiation detected by passive sensors is reflected sunlight. On the other hand, active sensors use internal energy to collect data about Earth. This is clearly illustrated by a laser-beam remote sensing system projects a laser onto the surface of Earth and measures the time that it takes for the laser to reflect back to its sensor. This review is focusing on applications in which the general target of the observation is an "object" on the Earth's surface, the measured energy is electromagnetic radiation, the sensors are positioned on UAVs platforms, and the recorded data are available as two-dimensional digital images. It is worth mentioning that other techniques of remote sensing such as laser, acoustical or sonar technologies as well as Earth's surface recognition (i.e. for precision agriculture) are excluded from this paper.

IEEE/ACM ASONAM 2020, December 7-10, 2020 978-1-7281-1056-1/20/$31.00 © 2020 IEEE

Any remote sensing application consists of two distinct processes: data acquisition (detection and recording of electromagnetic radiation), and data analysis (extraction of information from the recorded data). Electronic sensors convert electromagnetic radiation into electronic signals that can be stored as digital images locally or transmitted to a remote position. While the acquisition depends on the chosen sensor (camera), the last factor will vary in terms of the hardware and deep-learning algorithms used to process the interpretation of the images. Even though, near infrared, thermal, and depth-sensitive cameras can also be used for image recognition, the most common are RGB cameras, equipped with a sensor that collects the same bands of light as the human eye (i.e., red, green, and blue). During the second step of the remote sensing process, the remotely sensed data, must be analyzed in order to provide useful information about the observed features. The final product of the remote sensing process is usually a map showing the spatial distribution of the objects of interest.

Table 1. Advantages & Disadvantages of fixed wings and multicopter UAVs for remote sensing applications

Fixed wing vs Multicopters - Advantages and disadvantages		
UAV Type	Advantages	Disadvantages
Fixed wing	Long range	Take-off and landing space
	Long endurance	Low manoeuvrability
	High altitude	Minimum flight speed
	Efficiency	Flights at high altitudes affected by clouds
	Weight-lifting capacity	Higher cost
Multicopter	Vertical take-off	Low payload
	Hovering	Short range
	High Manoeuvrability	Short flight time
	Simple start-up & take-off	Wind susceptibility
	Low weight	
	Lower cost	
	Simple control	

Even though large military or commercial unmanned vehicles were capable to carry many sensors from the beginning, nowadays, small UAVs can also be equipped with extensive range of sensors and cameras. The miniaturization and the cost-effectiveness of electronics on one side, and the high-resolution cameras on board on the other, make UAVs flexible and adaptive for several high-performance applications.

The technology used depends on the size of the UAV and the type and detailed data to be collected. The range of advanced imaging and sensor technologies that can be hooked up on a small commercial UAV usually includes GPS, INS, standard cameras, hyperspectral and multispectral cameras, thermal sensors, as well as several other specialized sensors such as LiDAR and Radar sensors [18]. The GPS/INS data are mainly used for navigation and control but also allow measurement of the position and the orientation of the vehicle at all times. Especially when the remote sensing data is acquired, GPS/INS information are used to support the analysis of those data.

A. UAV Imaging sensors

Even though UAVs provide a flexible flight platform, the success of a monitoring mission depends on the sensors they are equipped with. With the improvement of UAV performance, different imaging sensors have also been developed rapidly. Almost all modern commercial UAV are equipped with at least one onboard imaging sensor, which could be used in many applications providing low weights and high resolution. Representative sensors widely used in both scientific research and business applications are digital cameras, spectral imaging sensors and thermal infrared cameras [19].

Interestingly enough UAV cameras provide fast images and real time videos of the target area while in some cases they could also be used as a vision navigation system. As compared to other types of sensors, there exist a wide range of RGB cameras on the market, with a great variety of features and costs. The low-cost RGB digital camera is widely used in remote sensing techniques, providing a high spatial resolution of radiation values in the red (~600 nm), green (~550 nm), and blue (~450 nm) spectral bands. Most commercial UAV platforms provide RGB sensors with differing spatial resolution determining the image quality. Common parameters for selecting RGB cameras for UAVs, include camera lens (better lens come with less geometric distortions), spectral range, resolution and quality of the charge coupled device (CCD)/complementary metal oxide semiconductor (CMOS) chips, as well as payload.

Spectral sensors are divided into multispectral and hyperspectral sensors. The classification criteria are the number of spectrum bands and the width of each spectrum band. A multispectral sensor generally detects five to twelve spectral bands in each pixel while a hyperspectral sensor can acquire imagery data with hundreds or thousands of spectrum bands in each pixel through narrow widths (5–10 nm) in the visible–infrared region. Multispectral cameras are one of the most commonly used sensors in addition to RGB cameras in the

IEEE/ACM ASONAM 2020, December 7-10, 2020 978-1-7281-1056-1/20/$31.00 © 2020 IEEE

UAV sensors family, because of their benefits of obtaining spectral information in the red-edge and near-infrared (NIR) band. Multispectral imagery generally refers to 3 to 10 bands that range from the visible to NIR. Hyperspectral sensors contain bands with narrow wavelengths while multispectral sensors contain bands with broad wavelengths [19]. Hyperspectral sensors consist of much narrower bands and generate more than 200 spectral bands that range from the visible to short wave infrared [18].

Thermal cameras typically carry a sensor that detects the infrared radiation emitted by a body, displaying its temperature in a digital radiometric image. Two types of thermal cameras are currently available: scanning devices that allow for capturing a point or a line and those with a two-dimensional infrared focal plane array [43]. All bodies emit electromagnetic energy in the infrared (IR)wavelength range depending on temperature according to the principle of black body radiation. A thermal sensor detects this invisible energy (with wavelengths from 3–14 μm), which is then converted into visible images showing the temperature of the target. A thermal sensor is prone to errors owing to fluctuating environmental conditions in the air and other objects emitting or reflecting thermal infrared radiation. Thus, periodical calibration for thermal sensors is crucial for collecting accurate data [42].

Table 2. Sensors for commercial unmanned aerial vehicle (UAV) platforms.

	Brand/model	Spectral range	Spatial Resolution	Weight
RGB Camera	Sony A9	~400–700 nm	24.2 MP	588 g
	Canon EOS 5D mark IV	~400–700 nm	30.4 MP	~800 g
	Nikon D850	~400–700 nm	45.7 MP	915 g
Multispectral sensors	Sentera Quad Sensor	RED 655 nm RED EDGE 725 nm Near infrared 800 nm	1248 × 950	170 g
	MicaSense ALTUM	BLUE 475 nm GREEN 560 nm RED 668 nm RED EDGE 717 nm Near infrared 840 nm	2064 × 1544	357 g
	Parrot Sequoia +	GREEN 550 nm RED 660 nm RED EDGE 735 nm Near infrared 790 nm	1280 × 960	72 g
Thermal infrared sensors	DJI Zenmuse XT	7.5–13.5 μm	640 × 512 336 × 256	270g
	Yuneec CGOET	8–14 μm	1920 × 1080	275g
	FLIR Duo Pro R	7.5–13.5	336 × 256	325g

IV. Object Detection Methodologies

G. Cheng et al. [17] presented a thorough review of the literature concerning generic object detection and the related methodologies. Different approaches have been developed for object detection from aerial images which could be generally divided into template matching-based methods, knowledge-based methods, OBIA-based methods and machine learning-based methods. Among the first approaches developed for object detection are the template matching-based methods. At the first stage, the template is generated and at the second stage the similarity is measured. During the template generation, a template for each to-be-detected object class should be firstly generated by hand-crafting or learning from the training set, while at the similarity measure, the stored template T is used to match the image at each possible position to find the best matches.

Another very popular approach for object detection is the knowledge-based methods through which object detection problem is translated into hypotheses testing problem by establishing various knowledge and rules. The main challenge of knowledge-based object detection methods is how to effectively transform the implicit knowledge understanding on target objects into the explicit detection rules. A new methodology for object detection is the Object-Based Image Analysis (OBIA) methods which classifies or maps high resolution imagery into meaningful objects. During the first stage of these methods the image is segmented into homogeneous regions (segments also called objects) representing a relatively homogeneous group of pixels by selecting desired scale, shape, and compactness. At the next stage a classification process is applied to these objects. OBIA methods prevail over conventional pixel-based image classification methods because they are capable to incorporate spatial context or object shape in the classification.

IEEE/ACM ASONAM 2020, December 7-10, 2020 978-1-7281-1056-1/20/$31.00 © 2020 IEEE

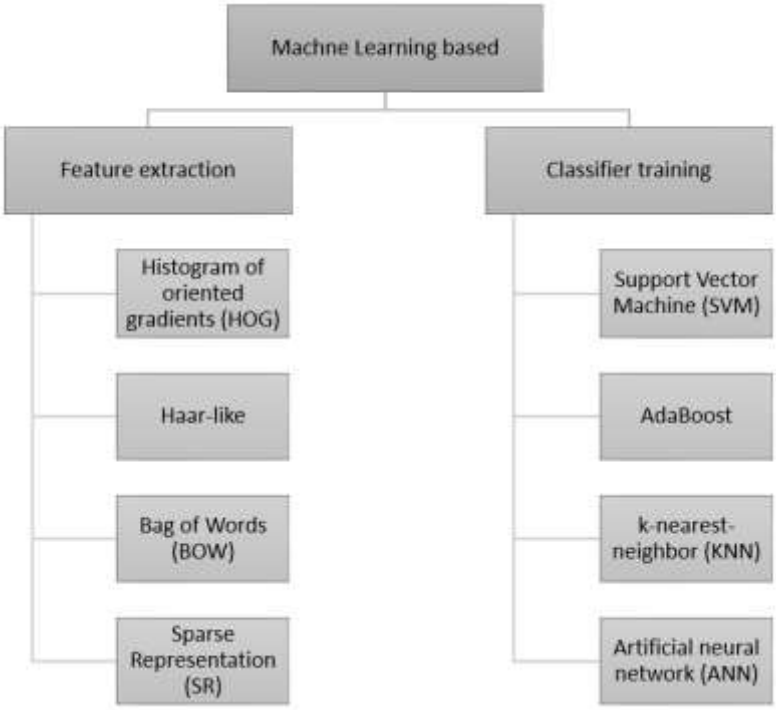

Fig. 1. Machine-learning Feature Extraction methods and Classifier Trainings.

The latest trend for object detection is machine learning-based methods where object detection can be performed by learning a classifier that captures the variation in object appearances and views from a set of training data in a supervised or semi-supervised framework. The input of the classifier is a set of object proposals with their corresponding feature representations and the output is their corresponding predicted labels, namely an object or not. Most popular feature extraction mechanisms, used at machine learning-based methods, are Histogram of oriented gradients (HOG) feature and Haar-like features. The next important step after feature extraction, is classifier training using a number of possible approaches with the objective of minimizing the misclassification error on the training dataset. There are different learning approaches such as Support Vector Machine (SVM was proposed by (Vapnik and Vapnik, [27]), AdaBoost (AdaBoost algorithm (Freund, [28] and Freund and Schapire, [29]) and Artificial Neural Networks (ANN) which are very capable to learn complicated patterns whose complexity impedes analysis especially via usage of other conventional approaches. A brief summary of feature extraction methods and classifier training methods is presented in Figure 1.

V. Convolutional Neural Network Detectors

Convolutional Neural Network (CNN) is a specific neural network architecture, [30], [31] which has been demonstrated as a powerful class of models in the computer vision field, beating state-of-the-art results on many tasks such as object detection, image segmentation and object recognition [13] – [14] – [15]. A Convolutional Neural Network (CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and able to differentiate one from the other. The pre-processing required in a CNN is much lower compared to other classification algorithms. In primitive methods filters are hand-engineered, with enough training, CNNs have the ability to learn these filters/characteristics.

CNNs are composed of multiple layers, with higher layers built on top of lower ones capturing more abstract representations of the input data. The structure of a CNN typically comprises a feature extractor stage followed by a classifier. The objective of the Convolution Operation is to extract high-level features such as the edges, from the input image. Conventionally, the first Convolutional Layer is responsible for capturing Low-Level features such as the edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well, giving a network, which has the wholesome understanding of images in the dataset, similar to human interpretation [8].

During the last decade, a variety of object detectors have been proposed by researchers, aiming at improving the accuracy of the detection while at the same time decreasing the computational complexity of their methods in order to achieve real-time performance for mobile and embedded platforms [32] The CNN-based object detectors, such as Faster R-CNN [33], R-FCN [34], YOLO [35] and SSD [36] can be divided into two categories with respect to their high-level structure: region-based detectors, and single-shot detectors (SSD).

IEEE/ACM ASONAM 2020, December 7-10, 2020 978-1-7281-1056-1/20/$31.00 © 2020 IEEE

Region-based detectors separate the prediction of the bounding box position from the object class prediction. There are two major components that make region-based CNN architectures powerful at the object detection task. [37] The first component replaces the low-level hand engineered features like HOG [38] or SIFT [39], with CNN feature maps which have larger representation capacity. But this larger representation capacity requires more computational effort to process the CNN features. The second component is a region proposal algorithm, which is used to propose regions of interest (ROI), that will contain the object of interest. Hence the features computation time is reduced by focusing the network attention on a smaller set of ROIs. Region-based CNN (R-CNN) [40] was the approach that pioneered using region proposal on top of CNN features as an object detector. Improvements of R-CNN are Fast R-CNN, Faster R-CNN, Mask R-CNN and Mesh R-CNN.

Single-Shot Detectors aim to avoid the performance bottlenecks of the 2-step region-based systems [7]. Single-Shot Detector algorithms like YOLO (You Only Look Once) [35] and SSD (Single-Shot Detector) [36] use a fully convolutional approach in which the network is able to find all objects within an image in one pass (hence 'single-shot' or 'look once') through the convolutional network. The region proposal algorithms usually have slightly better accuracy but slower to run, while single-shot algorithms are more efficient. The YOLO [35] framework casts object detection to a regression problem and in contrast to the Region Proposal Network and the classifier design of Faster R-CNN, employs a single CNN for the whole task. YOLO divides the input image into a grid of cells and for each cell outputs predictions for the coordinates of a number of bounding boxes, the confidence level for each box and a probability for each class. Compared to Faster R-CNN, YOLO is designed for real-time execution and by design provides a trade-off that favors high performance over detection accuracy.

In [44] a new family of scalable and efficient object detectors, named EfficientDet was introduced, based upon a previous work on scaling neural networks (EfficientNet), incorporating a novel bi-directional feature network (BiFPN) and new scaling rules. EfficientDet achieves state-of-the-art accuracy while being up to 9x smaller and using significantly less computation compared to prior state-of-the-art detectors. The idea behind EfficientDet arose from the effort to find solutions to improve computational efficiency by conducting a systematic study of prior state-of-the-art detection models. In general, object detectors have three main components: a backbone that extracts features from the given image; a feature network that takes multiple levels of features from the backbone as input and outputs a list of fused features that represent salient characteristics of the image; and the final class/box network that uses the fused features to predict the class and location of each object. By examining the design choices for these components, several key optimizations were identified, in order to improve performance and efficiency [44]. Such optimizations include:

- the implementation of the EfficientNet backbone, which offers a much better efficiency
- a new bi-directional feature network, BiFPN, which incorporates the multi-level feature fusion idea from FPN/PANet/NAS-FPN that enables information to flow in both the top-down and bottom-up directions, while using regular and efficient connections
- a new compound scaling method for object detectors, which jointly scales up the resolution/depth/width. Each network component, (i.e., backbone, feature, and box/class prediction network), will have a single compound scaling factor that controls all scaling dimensions using heuristic-based rules.

VI. Related Work of Machine Learning-Based Methods for UAV's Imagery

P. Doherty et al. [1] proposed a technique using two video sources (thermal and color) and allows for high rate human detection at larger distances than in the case of using the video sources separately with standard techniques. A thermal image is analyzed first to find human body sized silhouettes. Corresponding regions in a color image are subjected to a human body classifier which is configured to allow weak classifications. The classifier which is in fact a cascade of boosted classifiers working with Haar-like features requires training with a few hundred positive and negative examples. During learning the structure of a classifier is learned using boosting.

V. Reilly [4] proposed a system based on the geometric constraints of the orientation of shadow cast by a person in the scene with respect to the metadata (global position, time) and the relationship between average person height and the size of its corresponding shadow. The authors utilize the projection of shadow orientation to obtain a set of potential shadow candidates and then obtain a refined set of human candidates, which are pairs of shadow and normal blobs that are of correct geometric configuration, and relative size. Once the refined set of candidates has been obtained, they extract wavelet features from each human candidate, and classify it as either human or clutter using a Support Vector Machine (SVM). The rationale behind their geometric constraints was to improve the performance of any detection method by avoiding full frame search. Hence other models, features, and classification schemes suitable for aerial imagery can be used.

O. Oreifej et al. [3] presented a system using Histogram of Oriented Gradients (HOG) for people detection and identification from airborne optical images. For detection, they trained a support vector machine (SVM) classifier based on the HOG descriptor using a dataset of pedestrian images in aerial view. The HOG descriptor captures the most important cues of the human body, such as head and shoulders in good detail. Although reported results are promising, the scenarios are presented in simple and uncluttered environments which limit the general application of this technique.

A. Gąszczak et al. [2] presented an approach for the automatic detection of vehicles based on using multiple trained cascaded Haar classifiers with secondary confirmation in thermal imagery. Additionally, they presented a related approach for people detection in thermal imagery based on a similar combined set of cascaded Haar classifiers [21] with additional multivariate Gaussian shape matching for secondary

IEEE/ACM ASONAM 2020, December 7-10, 2020 978-1-7281-1056-1/20/$31.00 © 2020 IEEE

confirmation. The presented results showed the successful detection of vehicle and people under varying conditions in both isolated rural and cluttered urban environments with minimal false positive detection.

J. Gleason et al. [16] introduced an approach consisting of a cascade detection algorithm with the first stage serving as a fast detection solution, rejecting most of the background and selecting patterns corresponding to man-made objects. The patterns selected by the first stage are further refined in the second stage using four image classification methods (KNN, SVM, decision trees and random trees) and two feature extraction techniques (histogram of gradients and Gabor coefficients). The proposed system achieved best overall results using Gabor derived histograms and random trees classifiers.

Table 3. Comparison of machine learning-based methods for UAVs object detection.

Reference	Date	Processing Unit	Feature Extraction	Classifier Training	Datasets	CNN Detectors	Indicative Experimental Results (mean average precision – mAP)	Targets
P. Doherty [1]	2007	On-board	Cascaded Haar features	AdaBoost Decision trees				Human Body Detection
V. Reilly [4]	2010	On-board	Wavelet features	Support Vector Machine (SVM)				Humans
O. Oreifej [3]	2010	On-board	Histogram of Gradients (HOG)	Support Vector Machine (SVM)				Human identity
A. Gaszcza [2]	2011	Ground station	Cascaded Haar features	AdaBoost				People and Vehicles
J. Gleason [16]	2011	Off-board	Histogram of Gradients (HOG) Gabor coefficients	Support Vector Machine (SVM) K -nearest-neighbor (KNN) Decision trees (DTrees) Random trees (RTrees)			HOG/SVM-mAP=93.3% HOG/KNN-mAP=83.3% HOG/RTrees-mAP=85.7% Gabor/SVM-mAP=100% Gabor/KNN-mAP=100% Gabor/RTrees-mAP=98.9%	
J. Lee [10]	2017	On-board & Cloud		Convolutional Neural Networks - Region-based detectors	Pascal VOC 2007 Pascal VOC 2012	Fast YOLO YOLO SSD300 SSD500 Faster R-CNN	mAP=78.3% mAP=79.4% mAP=81.6% mAP=82.6% mAP=83.9%	Objects in an Indoor Environment
C. Kyrkou [8]	2018	On-board		Convolutional Neural Networks - Single-Stage Detectors (CNN-SSD)	Custom (about 5k images)	Modified Tiny-YOLO	mAP=95%	Vehicles
Subrahmanyam [7]	2019	On-board		Convolutional Neural Networks	VisDrone	Deep Feature Pyramid Network (DFPN) architecture	ResNet mAP=30.6% MobileNet mAP=29.2%	Object Detection
P. Nousi [5]	2019	On-board		Convolutional Neural Networks - Single-Stage Detectors (CNN-SSD)	Custom (about 12k images)	SSD YOLO Tiny YOLO		Object Detection and Tracking

IEEE/ACM ASONAM 2020, December 7-10, 2020 978-1-7281-1056-1/20/$31.00 © 2020 IEEE

D. Safadino [6]	2020	On-board		Convolutional Neural Networks - Single-Stage Detectors (CNN-SSD)	Microsoft Common Objects in Context (COCO)	SSD SSDLite		Human Detection

J. Lee et al. [10] proposed a hybrid solution namely moving the computation to an off-board computing cloud, while keeping low-level object detection and short-term navigation on-board. Using the cloud system, they were able to apply Faster R-CNNs [20], to detect not one or two but hundreds of object types in near real-time. To minimize the unpredictable lag from communication latencies they retained some visual processing locally, including a triage step that quickly identified region(s) of an image that are likely to correspond with objects of interest, as well as low-level feature matching needed for real-time navigation and stability. Their findings suggest that the cloud-based approach could allow speed-ups of nearly an order of magnitude, approaching real-time performance even when detecting hundreds of object categories, ignoring these additional communication lags.

C. Kyrkou et al. [8] presented a holistic approach for designing a single-shot object detector based on deep convolutional neural networks (CNNs) that enabling UAVs to perform vehicle detection. The CNN architecture, and the optimizations necessary to efficiently map such a CNN on a lightweight embedded processing platform suitable for deployment on UAVs. They focused on designing an efficient and lightweight network to accelerate the execution of the model with minimal compromise on the achieved accuracy. They adapted the Tiny-YOLO, model [9] to detect only one class (top-view vehicles) and then they explored the impact on performance by changing the structure of a CNN network such as the number of filters, the number of layers, the image size, the number of convolution and the pooling layers.

Subrahmanyam et al. [7] proposed an object detection model which is computationally less expensive, memory efficient and fast without compromising the detection performance, running on-board a drone. They proposed a Deep Feature Pyramid Network (DFPN) architecture and a modified loss function to avoid class imbalance and achieved real time object detection performance on real drone environment [41]. Their experiments were conducted using a low-cost quadcopter drone as a hardware platform, in the scenario of detecting target objects in an environment containing objects like pedestrians, buses, bicycles etc. They considered two networks namely ResNet and MobileNet as backbone convolutional bases for their detection model, concluding that ResNet provided better results in terms of detection accuracy, while combination of MobileNet resulted in real time speeds without compromising the detection accuracy.

P. Nousi et al. [5] proposed a combination of one-stage deep neural detectors and correlation-based trackers as it seemed to provide the best balance between accuracy and real-time performance, under the energy and computational constraints imposed by the UAV setting. Although Region-based detectors, such as Faster R-CNN, are more accurate, they tend to be slower than single-stage detectors, so their study is focused on Single-Stage Detectors, namely SSD and YOLO with MobileNet v1 and Inception v2 backbones detectors. A specific modular software system incorporating a range of detectors and trackers was implemented in a Robot Operating System (ROS) environment and evaluated on a number of relevant datasets. Their results indicated that a sophisticated, neural network-based detection and tracking system can be deployed at real-time even on embedded devices.

D. Safadinho et al. [6] proposed a solution that performed human detection from an aerial perspective through techniques of Computer Vision with the objective of estimating a safe landing location near a person. Their research is concentrated on low-cost equipment (i.e., camera, processing device) on-board a commercial UAV, to understand if the solution can be cost effective. The CNN models tested were the SSD-MobileNet-V2 and the SSDLite-MobileNet-V2, designed for devices with low computing capabilities. Both models presented patterns for the average time, precision, recall, and confidence. Their solution was tested iteratively in five different contexts deducing that the time elapsed for processing the images with the SSD takes more than twice the time than with the SSDLite. Finally, they proposed a synthesis of the CNN models, since:

- at higher altitudes the SSDLite is the proper algorithm to be used, offering lower processing time and better detection range and
- SSD model at lower altitudes due to better precision and confidence values.

To sum up, the analysis of the related works, is shown in Table 3. They were compared by the location of the processing unit, the Feature Extraction architecture, the Classifier Training method, the CNN Detectors and Datasets being used as well as the target objects. A column with indicative experimental results is also presented.

VII. Conclusions and Future Directions.

Object detection for improving UAV autonomy and remote sensing applications has always been an essential but challenging issue in the field of UAV image analysis, due to the restrictions and the 3-axis movement of the remote vehicle. In this paper, a review of indicative studies and experimentations was presented focusing on convolutional neural networks and the new capabilities provided by cloud processing. The (almost)

IEEE/ACM ASONAM 2020, December 7-10, 2020 978-1-7281-1056-1/20/$31.00 © 2020 IEEE

unlimited cloud-based computation resources could be the ideal solution for the high processing demands of computationally expensive state-of-the-art object detection algorithms such as CNNs. A promising future direction would be to exploit the capabilities of IoT, which could guarantee the cloud-based computational power and improve the potentially high and unpredictable communication lag.

REFERENCES

[1] Doherty, Patrick &Rudol, Piotr. (2007). A UAV Search and Rescue Scenario with Human Body Detection and Geolocalization. Adv. Artif. Intell.. 4830/2007. 1-13. 10.1007/978-3-540-76928-6_1.

[2] A. Gaszczak, T. P. Breckon, and J. Han, "Real-time people and vehicle detection from UAV imagery," Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques, vol. 7878, International Society for Optics and Photonics, 2011.

[3] O. Oreifej, R. Mehran, and M. Shah, "Human identity recognition in aerial images," Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 709-716 (2010)

[4] V. Reilly and M. Solmaz, "Geometric Constraints for Human Detection in Aerial Imagery," Proc. European Conference on Computer Vision, 252-265 (2010)

[5] Nousi, Paraskevi &Mademlis, Ioannis &Karakostas, Iason&Tefas, Anastasios & Pitas, Ioannis. (2019). Embedded UAV Real-Time Visual Object Detection and Tracking. 10.1109/RCAR47638.2019.9043931.

[6] Safadinho, David & Ramos, & Ribeiro, Roberto & Filipe, & Barroso, Joao & Pereira, António. (2020). UAV Landing Using Computer Vision Techniques for Human Detection. Sensors. 20. 613. 10.3390/s20030613.

[7] Vaddi, Subrahmanyam, "Efficient object detection model for real-time UAV applications" (2019). Graduate Theses and Dissertations. 17592.

[8] Kyrkou, Christos &Plastiras, George &Theocharides, Theo &Venieris, Stylianos &Bouganis, Christos. (2018). DroNet: Efficient convolutional neural network detector for real-time UAV applications. 967-972. 10.23919/DATE.2018.8342149.

[9] A. De Bruin and M. J. Booysen, "Drone-based traffic flow estimation and tracking using computer vision," 2015.

[10] Lee, Jangwon& Wang, Jingya& Crandall, David &Sabanovic, S. & Fox, Geoffrey. (2017). Real-Time, Cloud-Based Object Detection for Unmanned Aerial Vehicles. 36-43. 10.1109/IRC.2017.77.

[11] Corrigan,C.E.,Roberts,G.C.,Ramana,M.V.,Kim,D.,Ramanathan,V.,2008. Capturing vertical profiles of aerosols and black carbon over the Indian Ocean using autonomous unmanned aerial vehicles. Atmos. Chem. Phys. 8,737–747.

[12] Mak,J.E.,Su,L.,Guenther,A.,Karl,T.,2013.A novel whole airs ample profiler (WASP) for the quantification of volatile organic compounds in the boundary layer. Atmos. Meas. Tech. 6,2703–2712.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.

[14] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2625–2634.

[15] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in Advances in neural information processing systems, 2012, pp. 2843–2851.

[16] J. Gleason, A. V. Nefian, X. Bouyssounousse, T. Fong, and G. Bebis, "Vehicle detection from aerial imagery," Proc. of IEEE International Conference on Robotics and Automation, pp. 2065-2070, 2011.

[17] Cheng, Gong & Han, Junwei. (2016). A Survey on Object Detection in Optical Remote Sensing Images. ISPRS Journal of Photogrammetry and Remote Sensing. 117. 11-28. 10.1016/j.isprsjprs.2016.03.014.

[18] Arfaoui, Aymen. (2017). Unmanned Aerial Vehicle: Review of Onboard Sensors, Application Fields, Open Problems and Research Issues. International Journal of Image Processing. 11. 12-24.

[19] L.-J. Ferrato, and K. W. Forsythe, "Comparing hyperspectral and multispectral imagery for land classification of the Lower Don River, Toronto," Journal of Geography and Geology 5(1), 92 (2013).

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Advances in Neural Information Processing Systems (NIPS), 2015.

[21] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 511-518 (2001)

[22] McGonigle,A.J.S., Aiuppa,A., Giudice,G., Tamburello,G., Hodson,A.J., Gurrieri,S.,2008. Unmanned aerial vehicle measurements of volcanic carbon dioxide fluxes. Geophys.Res.Let.35(6)L06303.

[23] Austin, Reg, Unmanned Aircraft Systems. Great Britain, UK. Wiley. 2010

[24] Carleton UAV Research Group, Carleton University, UAV FAQs, 5 M, 2014

[25] Mahen M.A., Anirudh S. Naik, Chethana H.D, Shashank A.C. Design and Development of Amphibious Quadcopter. International Journal of Mechanical and Production Engineering, Vol. 2 No. 2. 2014.

[26] Magnussen, Øyvind, and KjellEivindSkjønhaug. Modeling, design and experimental study for a quadcopter system construction. Department of Engineering University of Adger. 2011.

[27] Vapnik, V.N., Vapnik, V., 1998. Statistical learning theory. Wiley New York.

[28] Freund, Y., 1995. Boosting a weak learning algorithm by majority. Inf. Comput. 121, 256-285.

[29] Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. In: Proc. Int. Conf. Mach. Learn., pp. 148-156.

[30] Jin, X., Davis, C.H., 2007. Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks. Image Vis. Comput. 25, 1422-1431.

[31] Wang, J., Song, J., Chen, M., Yang, Z., 2015. Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine. Int. J. Remote Sens. 36, 3144-3169.

[32] G. D. T. N. SubarnaTripathi, Byeongkeun Kang, "Low-complexity object detection with deep convolutional neural network for embedded systems," Proc.SPIE, vol. 10396, pp. 10 396 – 10 396 – 15, 2017.

[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.

[34] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object Detection via Region based Fully Convolutional Networks," in NIPS, 2016, pp. 379–387.

[35] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.

[36] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," ECCV, pp. 21–37, 2016.

[37] Muhammad, Amgad&Moustafa, Mohamed. (2018). Improving Region Based CNN Object Detector Using Bayesian Optimization. 10.1109/IPAS.2018.8708859.

[38] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in In CVPR, 2005, pp. 886–893.

[39] D. G. Lowe, "Distinctive image features from scale invariant key points," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[40] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," CoRR, vol. abs/1311.2524, 2013.

[41] T.-Y. Lin, P. D. (2017). Feature pyramid networks for object detection. CVPR.Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces (Translation Journals style)," *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9thAnnu. Conf. Magnetics* Japan, 1982, p. 301].

[42] Jang, Gyujin& Kim, Jaeyoung& Yu, Ju-Kyung & Kim, Hak-Jin& Kim, Yoonha& Kim, Dong-Wook& Kim, Kyung-Hwan & Lee, Chang & Chung, Yong Suk. (2020). Review: Cost-Effective Unmanned Aerial Vehicle (UAV) Platform for Field Plant Breeding Application. Remote Sensing. 12. 998. 10.3390/rs12060998.

[43] Messina, Gaetano &Modica, Giuseppe. (2020). Applications of UAV Thermal Imagery in Precision Agriculture: State of the Art and Future Research Outlook. Remote Sensing. 12. 1491. 10.3390/rs12091491.

[44] Tan, Mingxing& Pang, Ruoming& Le, Quoc. (2019). EfficientDet: Scalable and Efficient Object Detection.

IEEE/ACM ASONAM 2020, December 7-10, 2020 978-1-7281-1056-1/20/$31.00 © 2020 IEEE