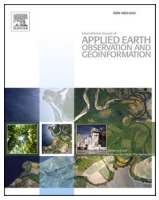
International Journal of Applied Earth Observations and Geoinformation 102 (2021) 102456

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



A review on deep learning in UAV remote sensing

Lucas Prado Osco [a,*], José Marcato Junior [b], Ana Paula Marques Ramos [c,d], Lúcio André de Castro Jorge [e], Sarah Narges Fatholahi [f], Jonathan de Andrade Silva [g], Edson Takashi Matsubara [g], Hemerson Pistori [g,h], Wesley Nunes Gonçalves [b,g], Jonathan Li [f]

[a] Faculty of Engineering and Architecture and Urbanism, University of Western São Paulo, Rod. Raposo Tavares, km 572 - Limoeiro, Pres. Prudente 19067-175, SP, Brazil
[b] Faculty of Engineering, Architecture, and Urbanism and Geography, Federal University of Mato Grosso do Sul, Av. Costa e Silva, Campo Grande 79070-900, MS, Brazil
[c] Environment and Regional Development Program, University of Western São Paulo, Rod. Raposo Tavares, km 572 - Limoeiro, Pres. Prudente 19067-175, SP, Brazil
[d] Agronomy Program, University of Western São Paulo, Rod. Raposo Tavares, km 572 - Limoeiro, Pres. Prudente 19067-175, SP, Brazil
[e] National Research Center of Development of Agricultural Instrumentation, Brazilian Agricultural Research Agency, R. XV de Novembro, 1452, São Carlos 13560-970, SP, Brazil
[f] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[g] Faculty of Computing, Federal University of Mato Grosso do Sul, Av. Costa e Silva, Campo Grande 79070-900, MS, Brazil
[h] Inovisão, Dom Bosco Catholic University, Av. Tamandaré, 6000, Campo Grande 79117-900, MS, Brazil

ARTICLE INFO

Keywords:
Convolutional neural networks
Remote sensing imagery
Unmanned aerial vehicles

ABSTRACT

Deep Neural Networks (DNNs) learn representation from data with an impressive capability, and brought important breakthroughs for processing images, time-series, natural language, audio, video, and many others. In the remote sensing field, surveys and literature revisions specifically involving DNNs algorithms' applications have been conducted in an attempt to summarize the amount of information produced in its subfields. Recently, Unmanned Aerial Vehicle (UAV)-based applications have dominated aerial sensing research. However, a literature revision that combines both "deep learning" and "UAV remote sensing" thematics has not yet been conducted. The motivation for our work was to present a comprehensive review of the fundamentals of Deep Learning (DL) applied in UAV-based imagery. We focused mainly on describing the classification and regression techniques used in recent applications with UAV-acquired data. For that, a total of 232 papers published in international scientific journal databases was examined. We gathered the published materials and evaluated their characteristics regarding the application, sensor, and technique used. We discuss how DL presents promising results and has the potential for processing tasks associated with UAV-based image data. Lastly, we project future perspectives, commentating on prominent DL paths to be explored in the UAV remote sensing field. This revision consisting of an approach to introduce, commentate, and summarize the state-of-the-art in UAV-based image applications with DNNs algorithms in diverse subfields of remote sensing, grouping it in the environmental, urban, and agricultural contexts.

Abbreviations: AdaGrad, Adaptive Gradient Algorithm; AI, Artificial Intelligence; ANN, Artificial Neural Network; CEM, Context Enhanced Module; CNN, Convolutional Neural Network; DCGAN, Deep Convolutional Generative Adversarial network; DDCN, Deep Dual-domain Convolutional neural Network; DL, Deep Learning; DNN, Deep Neural Network; DEM, Digital Elevation Model; DSM, Digital Surface Model; FPS, Frames per Second; GAN, Generative Adversarial Network; GPU, Graphics Processing Unit; KL, Kullback-Leibler; LSTM, Long Short-Term Memory; IoU, Intersection over Union; ML, Machine Learning; MAE, Mean Absolute Error; MAPE, Mean Absolute Percentage Error; MRE, Mean Relative Error; MSE, Mean Squared Error; MSLE, Mean Squared Logarithmic Error; MSM, Multi-Stage Module; MVS, Multiview Stereo; NAS, Network Architecture Search; PCA, Principal Component Analysis; PPM, Pyramid Pooling Module; r, Correlation Coefficient; RMSE, Root Mean Squared Error; RNN, Recurrent Neural Network; ROC, Receiver Operating Characteristics; RPA, Remotely Piloted Aircraft; SAM, Spatial Attention Module; SGD, Stochastic Gradient Descent; SfM, Structure from Motion; UAV, Unmanned Aerial Vehicle; WOS, Web of Science.

* Corresponding author.
E-mail addresses: lucasosco@unoeste.br (L.P. Osco), jose.marcato@ufms.br (J. Marcato Junior), anaramos@unoeste.br (A.P. Marques Ramos), lucio.jorge@embrapa.br (L.A. de Castro Jorge), nfatholahi@uwaterloo.ca (S.N. Fatholahi), jonathan.andrade@ufms.br (J. de Andrade Silva), edsontm@facom.ufms.br (E.T. Matsubara), pistori@ucdb.br (H. Pistori), wesley.goncalves@ufms.br (W.N. Gonçalves), junli@uwaterloo.ca (J. Li).

https://doi.org/10.1016/j.jag.2021.102456
Received 22 January 2021; Received in revised form 30 June 2021; Accepted 17 July 2021

1. Introduction

For investigations using remote sensing image data, multiple processing tasks depend on computer vision algorithms. In the past decade, applications conducted with statistical and Machine Learning (ML) algorithms were mainly used in classification/regression tasks. The increase of remote sensing systems allowed a wide collection of data from any target on the Earth's surface. Aerial imaging has become a common approach to acquiring data with the advent of Unnamed Aerial Vehicles (UAV). These are also known as Remotely Piloted Aircrafts (RPA), or, as a commonly adopted term, drones (multi-rotor, fixed wings, hybrid, etc). These devices have grown in market availability for their relatively low cost and high operational capability to capture images quickly and in an easy manner. The high-spatial-resolution of UAV-based imagery and its capacity for multiple visits allowed the creation of large and detailed amounts of datasets to be dealt with.

The surface mapping with UAV platforms presents some advantages compared to orbital and other aerial sensing methods of acquisition. Less atmospheric interference, the possibility to fly within lower altitudes, and mainly, the low operational cost have made this acquisition system popular in both commercial and scientific explorations. However, the visual inspection of multiple objects can still be a time-consuming, biased, and inaccurate operation. Currently, the real challenge in remote sensing approaches is to obtain automatic, rapid, and accurate information from this type of data. In recent years, the advent of Deep Learning (DL) techniques has offered robust and intelligent methods to improve the mapping of the Earth's surface.

DL is an Artificial Neural Network (ANN) method with multiple hidden layers and deeper combinations, which is responsible for optimizing and returning better learning patterns than a common ANN. There is an impressive amount of revision material in the scientific journals explaining DL-based techniques, its historical evolution, general usage, as well as detailing networks and functions. Highly detailed publications, such as Lecun (Lecun et al., 2015) and Goodfellow (Goodfellow et al., 2016) are both considered important material in this are. As computer processing and labeled examples (i.e. samples) became more available in recent years, the performance of Deep Neural Networks (DNNs) increased in the image-processing applications. DNN has been successfully applied in data-driven methods. However, much needs to be covered to truly understand its potential, as well as its limitations. In this regard, several surveys on the application of DL in remote sensing were developed in both general and specific contexts to better explain its importance.

The context in which remote sensing literature surveys are presented is variated. Zhang et al. (2016) organized a revision material which explains how DL methods were being applied, at the time, to image classification tasks. Later, Cheng and Han (2016) investigated object detection in optical images, but focused more on the traditional ANN and ML. A complete and systematic review was presented by Ball et al. (2017) in a survey describing DL theories, tools, and its challenges in dealing with remote sensing data. Cheng et al. (2017) produced a revision on image classification with examples produced at their experiments. Also, focusing on classification, Zhu et al. (2017) summarized most of the current information to understand the DL methods used for this task. Additionally, a survey performed by Li et al. (2018) helped to understand some DL applications regarding the overall performance of DNNs in publicly available datasets for image classification task. Yao et al. (2018) stated in their survey that DL will become the dominant method of image classification in remote sensing community.

Although DL does provide promising results, many observations and examinations are still required. Interestingly enough, multiple remote sensing applications using hyperspectral imagery (HSI) data were in the process, which gained attention. In Petersson et al. (2017), probably one of the first surveys on hyperspectral data was performed. In (Signoroni et al., 2019), is presented a multidisciplinary review about how DL models have been widely used in the field of HSI dataset processing. These authors highlighted that, among the distinct areas of applications, remote sensing approaches are one of the most emerging. Regarding the use of DL models to process highly detailed remotely sensed HSI data, Signoroni et al. (2019) summarized usage into classification tasks, object detection, semantic segmentation, and data enhancement, such as denoising, spatial super-resolution, and fusion. Ado et al. (2020) present a recent review on hyperspectral imaging acquired by UAV-based sensors for agriculture and forestry applications, and show that there are manifold DL approaches to deal with HSI dataset complexity.

A more recent survey is presented by Jia et al. (2021) regarding DL for hyperspectral image classification considering few labeled samples. They commentate how there is a notable gap between deep learning models and HSI datasets because DL models usually need sufficient labeled samples, but it is generally difficult to acquire many samples in HSI dataset due to the difficulty and time-consuming nature of manual labeling. However, the issues of small-sample sets may be well defined by the fusion of deep learning methods and related techniques, such as transfer learning and a lightweight model. Deep learning is also a new approach for the domain of infrared thermal imagery processing to attend different domains, especially in satellite-provided data. Some of these applications are the usage of convolutional layers to detect potholes on roads with terrestrial imagery (Aparna et al., 2019), detection of land surface temperatures from combined multispectral and microwave observations from orbital platforms (Wang et al., 2020b), or determining sea surface temperature patterns to identify ocean temperatures extremes (Xavier Prochaska et al., 2021) from orbital imagery.

Yet in the literature revision theme, a comparative review by Audebert et al. (2019) was conducted by examining various families of networks' architectures while providing a toolbox to perform such methods to be publicly available. In this regard, another paper written by Paoletti et al. (2019) organized the source code of DNNs to be easily reproduced. Similar to Cheng et al. (2017), Li et al. (2019a) conducted a literature revision while presenting an experimental analysis with DNNs' methods. As of recently, literature revision focused on more specific approaches within this theme. Some of which included DL methods for enhancement of remote sensing observations, as super-resolution, denoising, restoration, pan-sharpening, and image fusion techniques, as demonstrated by Tsagkatakis et al. (2019) and Signoroni et al. (2019). Also, a meta-analysis by Ma et al. (2019) was performed concerning the usage of DL algorithms in seven subfields of remote sensing: image fusion and image registration, scene classification, object detection, land use and land cover classification, semantic segmentation, and object-based image analysis (OBIA).

Although, from these recent reviews, various remote sensing applications using DL can be verified, it should be noted that the authors did not focus on specific surveying in the context of DL algorithms applied to UAV-image sets, which is something that, at the time of writing, has gained the attention of remote sensing investigations. We verified in the literature that, in general, similar DL methods are used for imagery acquired at different levels, resolutions and domains, such as the ones from orbital, aerial, terrestrial and proximal sensing platforms. However, as of recently, some of the proposed deep neural networks are maintaining high resolution images into deeper layers (Kannojia and Jaiswal, 2018). This type of deep networks may benefit from UAV-based data, taking advantage of its resolutions. Indeed, there are orbital images with high spatial resolutions, but these are not as commonly available to the general public as UAV-based images. Because of that, these kinds of architectures associated with UAV-based data may be a surging trend in remote sensing applications.

Another interesting take on DL-based methods was related to image segmentation in a survey by Hossain and Chen (2019), which its theme was expanded by Yuan et al. (2021) and included state-of-the-art algorithms. A summarized analysis by Zheng et al. (2020) focused on remote sensing images with object detection approaches, indicating some of the challenges related to the detection with few labeled samples, multi-scale issues, network structure problems, and cross-domain detection

difficulties. In more of a "niche" type of research, environmental applications and land surface change detection were investigated in literature revision papers by Yuan et al. (2020) and Khelifi and Mignotte (2020), respectively.

The aforementioned studies were evaluated with a text processing method that returned a word cloud in which the word size denotes the frequency of the word within these papers (Fig. 1). An interesting observation regarding this world-cloud is that the term "UAV" is under or not represented at all. This revision gap is a problem since UAV image data is daily produced in large amounts, and no scientific investigation appears to offer a comprehensive literature revision to assist new research on this matter. In the UAV context, there are some revision papers published in important scientific journals from the remote sensing community. As of recently, a revision-survey (Bithas et al., 2019) focused on the implications of ML methods being applied to UAV image processing, but no investigation was conducted on DL algorithms for this particular issue. This is an important theme, especially since UAV platforms are more easily available to the public and DL-based methods are being tested to provide accurate mapping in highly detailed imagery.

As mentioned, UAVs offer flexibility in data collection, as flights are programmed under users' demand; they are low-cost when compared to other platforms that offer similar spatial-resolution images; produce high-level of detail in its data collection; presents dynamic data characteristics since it is possible to embed RGB, multispectral, hyperspectral, thermal and, LiDAR sensors on it; and are capable of gathering data from difficult to access places. Aside from that, sensors embedded in UAVs are known to generate data at different altitudes and point-of-views. These characteristics, alongside others, are known to produce a higher dynamic range of images than common sensing systems. This ensures that the same object is viewed from different angles, where not only their spatial and spectral information is affected, as well as form, texture, pattern, geometry, illumination, etc. This becomes a challenge for multidomain detection. As such, studies indicate that DL is the most prominent solution for dealing with these disadvantages. These studies, which most are presented in this revision paper, were conducted within a series of data criteria and evaluated DL architectures in classifying, detecting, and segmenting various objects from UAV scenes.

To the best of our knowledge, there is a literature gap related to review articles combining both "deep learning" and "UAV remote sensing" thematics. This survey is important to summarize the direction of DL applications in the remote sensing community, particularly related to UAV-imagery. The purpose of this study is to provide a brief review of DL methods and their applications to solve classification, object detection, and semantic segmentation problems in the remote sensing field. Herein, we discuss the fundamentals of DL architectures, including recent proposals. There is no intention of summarizing existing literature, but to present an examination of DL models while offering the necessary information to understand the state-of-the-art in which it encounters. Our revision is conducted highlighting traits about the UAV-based image data, their applications, sensor types, and techniques used in recent approaches in the remote sensing field. Additionally, we relate how DL models present promising results and project future perspectives of prominent paths to be explored. In short, this paper brings the following contributions:

1. A presentation of fundamental ideas behind the DL models, including classification, object detection, and semantic segmentation approaches; as well as the application of these concepts to attend UAV-image based mapping tasks;
2. The examination of published material in scientific sources regarding sensors types and applications, categorized in environmental, urban, and agricultural mapping contexts;
3. The organization of publicly available datasets from previous researches, conducted with UAV-acquired data, also labeled for both object detection and segmentation tasks;
4. A description of the challenges and future perspectives of DL-based methods to be applied with UAV-based image data.

2. Deep neural networks overview

DNNs are based on neural networks which are composed of neurons (or units) with certain activations and parameters that transform input data (e.g., UAV remote sensing image) to outputs (e.g., land use and land cover maps) while progressively learning higher-level features (Ma et al., 2019; Schmidhuber, 2015). This progressive feature learning occurs, among others, on layers between the input and the output, which are referred to as hidden layers (Ma et al., 2019). DNNs are considered as

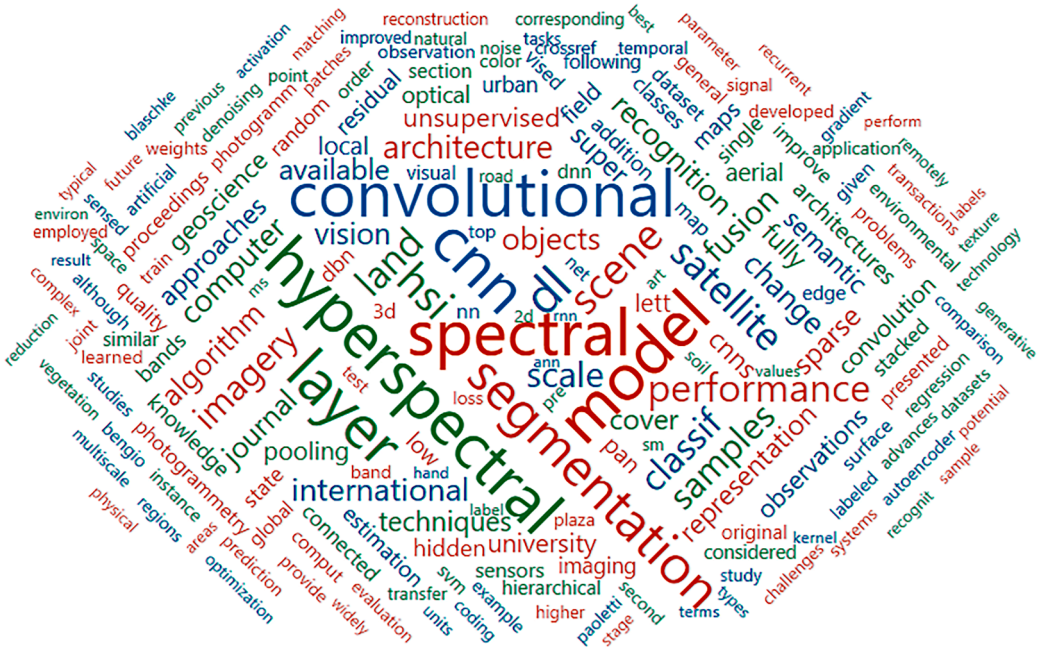


Fig. 1. Word-cloud of different literature-revision papers related to the "remote sensing" and "deep learning" themes.

a DL method in their most traditional form (i.e. with 2 or more hidden layers). Their concept, based on an Artificial Intelligence (AI) modeled after the biological neurons' connections, exists since the 1950s. But only later, with advances in computer hardware and the availability of a high number of labeled examples, its interest has resurged in major scientific fields. In the remote sensing community, the interest in DL algorithms has been gaining attention since mid 2010s decade, specifically because these algorithms achieved significant success at digital image processing tasks (Ma et al., 2019; Khan et al., 2020).

A DNN works similarly to an ANN, when as a supervised algorithm, uses a given number of input features to be trained, and that these feature observations are combined through multiple operations, where a final layer is used to return the desired prediction. Still, this explanation does not do much to highlight the differences between traditional ANNs and DNNs. LeCun et. al. (Lecun et al., 2015), the paper amongst the most cited articles in DL literature, defines DNN as follows: "Deep-learning methods are representation-learning methods with multiple levels of representation". Representation-learning is a key concept in DL. It allows the DL algorithm to be fed with raw data, usually unstructured data such as images, texts, and videos, to automatically discover representations.

The most common DNNs (Fig. 2) are generally composed of dense layers, wherein activation functions are implemented in. Activation functions compute the weighted sum of input and biases, which is used to decide if a neuron can be activated or not (Nwankpa et al., 2018). These functions constitute decision functions that help in learning intrinsic patterns (Khan et al., 2020); i.e., they are one of the main aspects of how each neuron learns from its interaction with the other neurons. Known as a piecewise linear function type, ReLu defines the 0 valor for all negative values of X. This function is, at the time of writing, the most popular in current DNNs models. Regardless, another potential activation function recently explored is Mish, a self regularized non-monotonic activation function (Khan et al., 2020). Aside from the activation function, another important information on how a DNN works is related to its layers, such as dropout, batch-normalization, convolution, deconvolution, max-pooling, encode-decode, memory cells, and others. This layer is regularly used to solve issues with covariance-shift within feature-maps (Khan et al., 2020). The organization in which the layers are composed, as well as its parameters, is one of the main aspects of the architecture.

Multiple types of architectures were proposed in recent years to improve and optimize DNNs by implementing different kinds of layers, optimizers, loss functions, depth-level, etc. However, it is known that one of the major reasons behind DNNs' popularity today is also related to the high amount of available data to learn from it. A rule of thumb conceived among data scientists indicates that at least 5,000 labeled examples per category was recommended (Goodfellow et al., 2016). But, as of today, DNNs' proposals focused on improving these network's capacities to predict features with fewer examples than that. Some applications which are specifically oriented may benefit from it, as it reduces the amount of labor required at sample collection by human inspection. Even so, it should be noted that, although this pursuit is being conducted, multiple takes are performed by the vision computer communities and novel research includes methods for data-augmentation, self-supervising, and unsupervised learning strategies, as others. A detailed discussion of this manner is presented in (Khan et al., 2020).

2.1. Convolutional and recurrent neural networks

A DNN can be formed by different architectures, and the complexity of the model is related to how each layer and additional computational method is implemented. Different DL architectures are proposed regularly, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Deep Belief Networks (DBN) (Ball et al., 2017), and, more recently yet, Generative Adversarial Networks (GAN) (Goodfellow et al., 2016). However, the most common DNNs in the supervised networks categories are usually classified as CNNs (Fig. 3) and RNNs (Khan et al., 2020).

As a different kind of DL network structure, RNNs refer to another supervised learning model. The main idea behind implementing RNNs regards their capability of improving their learning by repetitive observations of a given phenom or object, often associated with a time-series collection. A type of RNN being currently implemented in multiple tasks is the Long Short-Term Memory (LSTM)(Hochreiter and Schmidhuber, 1997). In the remote sensing field, RNN models have been applied to deal with time series tasks analysis, aiming to produce, for example, land cover mapping (Ienco et al., 2017; Ho Tong Minh et al.,

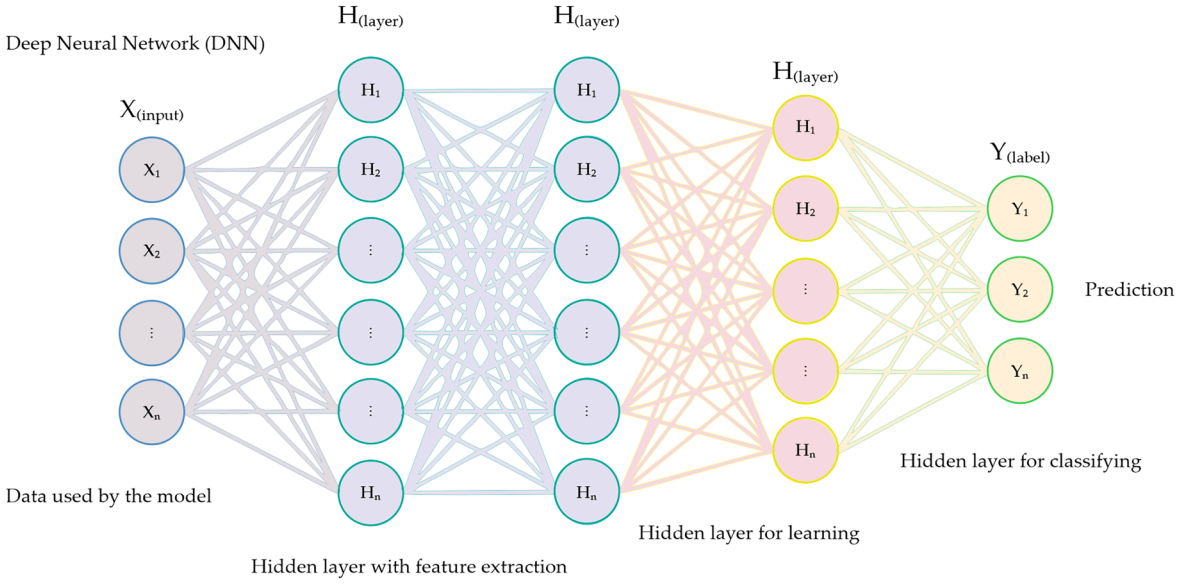


Fig. 2. A DNN architecture. This is a simple example of how a DNN may be built. Here the initial layer (X_{input}) is composed of the collected data samples. Later this data information can be extracted by hidden layers in a back-propagation manner, which is used by subsequent hidden layers to learn these features' characteristics. In the end, another layer is used with an activation function related to the given problem (classification or regression, as an example), by returning a prediction outcome (Y_{label}).

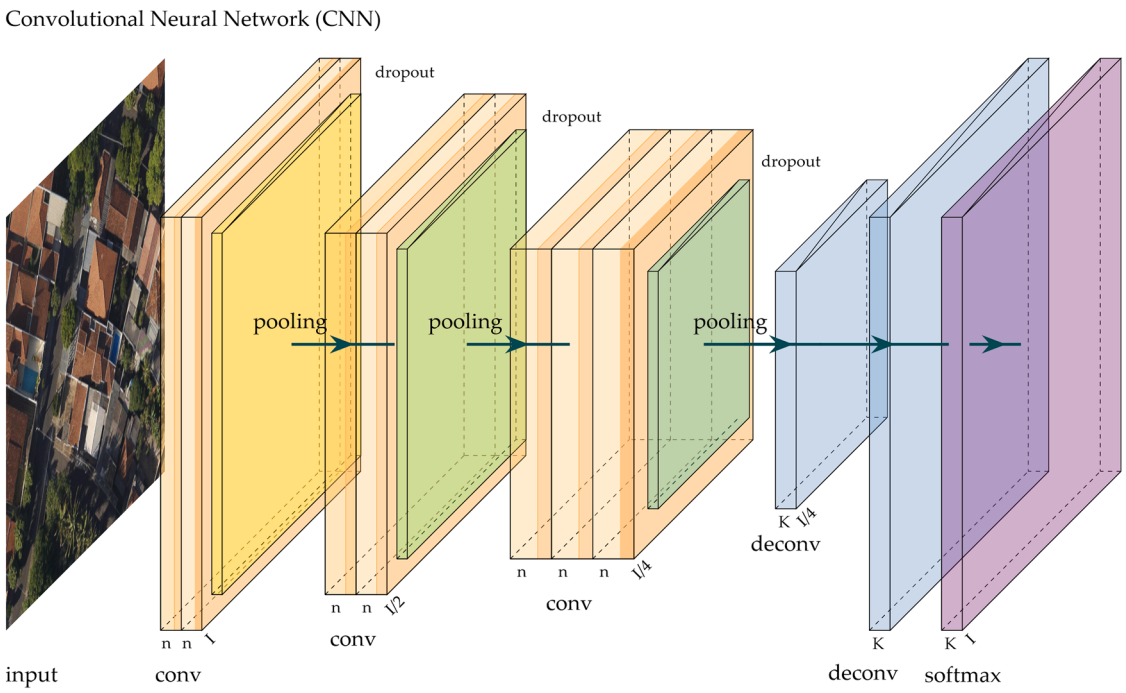


Fig. 3. A CNN type of architecture with convolution and deconvolution layers. This example architecture is formed by convolutional layers, where a dropout layer is added between each conv layer, and a max-pooling layer is adopted each time the convolution window-size is decreased. By the end of it, a deconvolutional layer is used with the same size as the last convolutional, and then it uses information from the previous step to reconstruct the image with its original size. The final layer is of a softmax, where it returns the models' predictions.

2018). For a pixel-based time series analysis aiming to discriminate classes of winter vegetation coverage using SAR Sentinel-1 (Ho Tong Minh et al., 2018), it was verified that RNN models outperformed classical ML approaches. A recent approach (Feng et al., 2020) for accurate vegetation mapping combined multiscale CNN to extract spatial features from UAV-RGB imagery and then fed an attention-based RNN to establish the sequential dependency between multitemporal features. The aggregated spatial-temporal features are used to predict the vegetable category. Such examples with remote sensing data demonstrate the potential in which RNNs are being used. Also, one prominent type of architecture is the CNN-LSTM method (Fig. 4). This network uses convolutional layers to extract important features from the given input image and feed the LSTM. Although few studies implemented this type of network, it should be noted that it serves specific purposes, and its usage, for example, can be valued for multitemporal applications.

As aforementioned, other types of neural networks, aside from CNNs

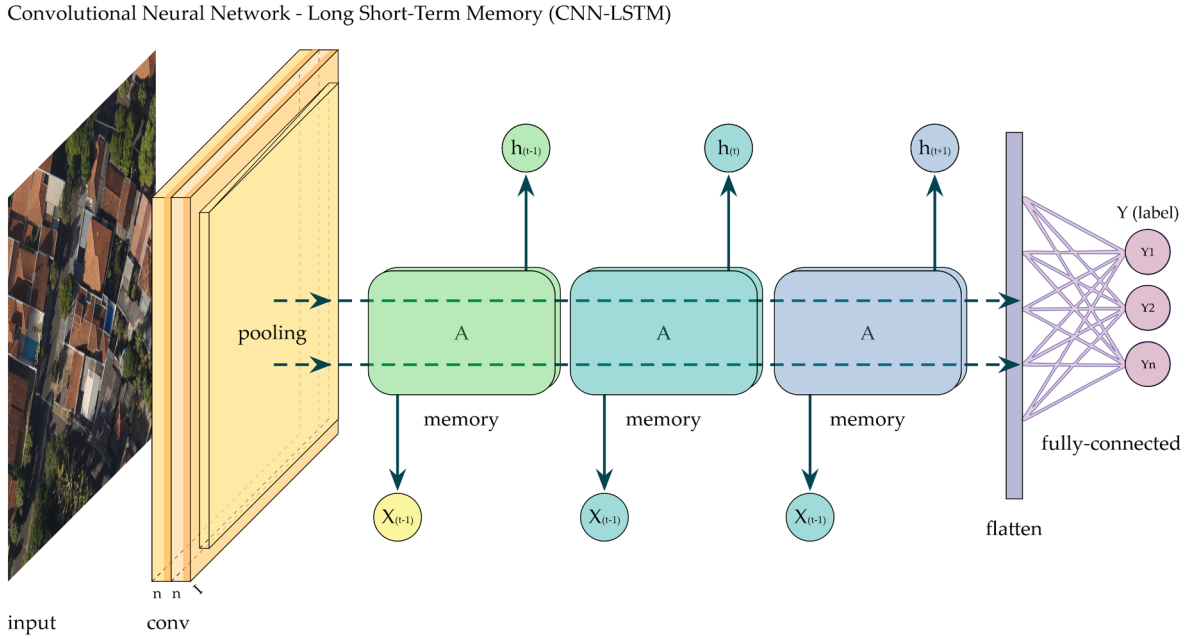


Fig. 4. An example of a neural network based on the CNN-LSTM type of architecture. The input image is processed with convolutional layers, and a max-pooling layer is used to introduce the information to the LSTM. Each memory cell is updated with weights from the previous cell. After this process, one may use a flatten layer to transform the data in an arrangement to be read by a dense (fully-connected) layer, returning a classification prediction, for instance.

and RNNs, are currently being proposed to also deal with an image type of data. GANs are amongst the most innovative unsupervised DL models. GANs are composed of two networks: generative and discriminative, that contest between themselves. The generative network is responsible for extracting features from a particular data distribution of interest, like images, while the discriminative network distinguishes between real (reference or ground truth data) and those data generated by the generative part of GANs (fake data) (Goodfellow et al., 2014; Ma et al., 2019). Recently approaches in the image processing context like the classification of remote sensing images (Lin et al., 2017a) and image-to-image translation problems solution (Isola et al., 2018) adopted GANs as DL model, obtaining successful results.

In short, several DNNs are constantly developed, in both scientific and/or image competition platforms, to surpass existing methods. However, as each year passes, some of these neural networks are often mentioned, remembered, or even improved by novel approaches. A summary of well-known DL methods built in recent years is presented in Fig. 5. A detailed take on this, which we recommend to anyone interested, is found in Khan et al. (2020). Alongside the creations and developments of these and others, researchers observed that higher depth channel exploration, and, as of recently proposed, attention-based feature extraction neural networks, are regarded as some of the most prominent approaches for DL. Initially, most of the proposed supervised DNNs, like CNN and RNN, or CNN-LSTM models, were created to perform and deal with specific issues. Often, these approaches can be grouped into classification tasks, like scene-wise classification, object

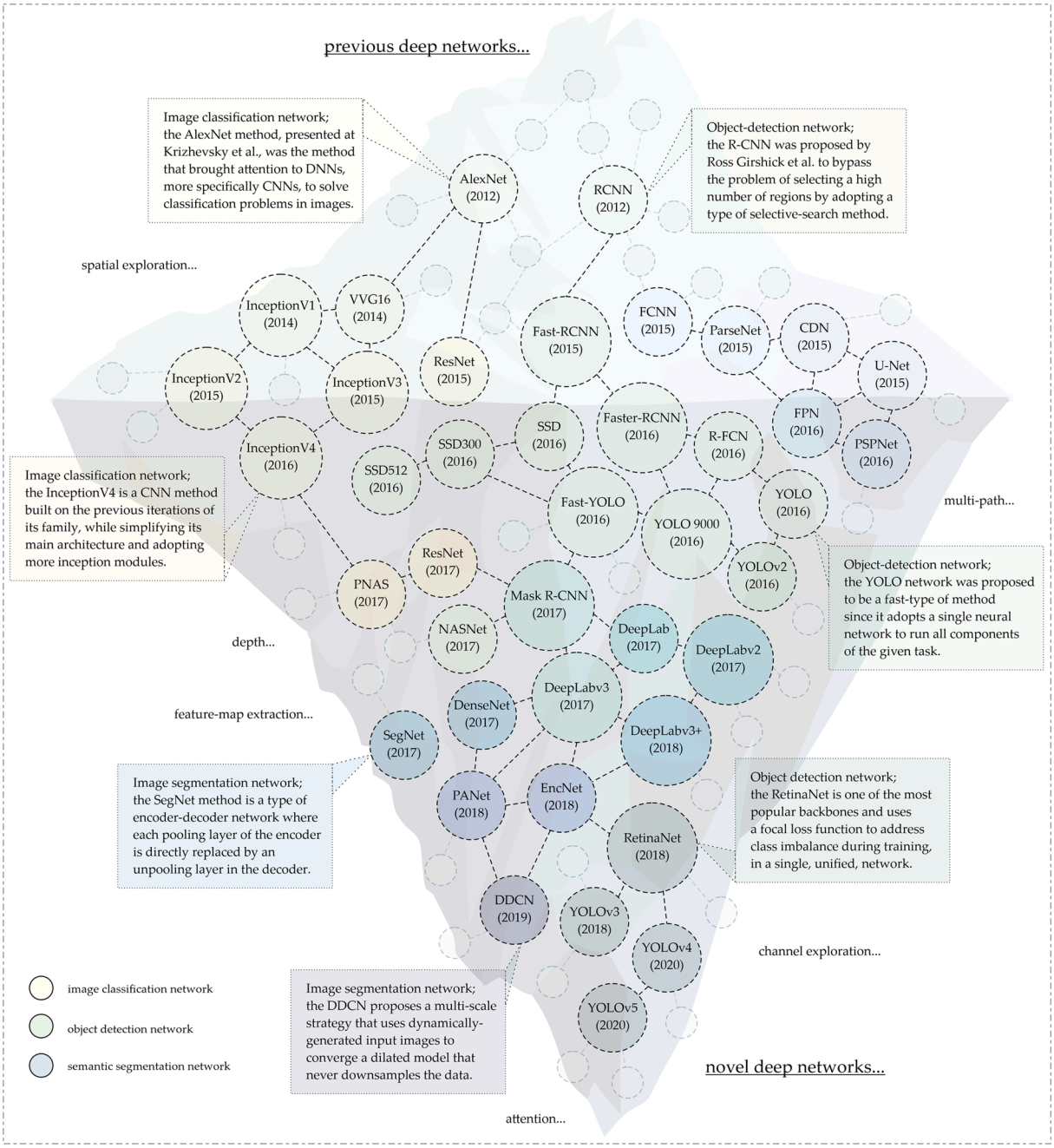


Fig. 5. A DL time-series indicating some popular architectures implemented in image classification (yellowish color), object detection (greenish color), and segmentation (bluish color). These networks often intertwine, and many adaptations have been proposed for them. Although it may appear that most of the DL methods were developed during 2015–2017 annuals, it is important to note that, as some, novel deep networks use most of the already developed methods as backbones, or accompanied from other types of architectures, mainly used as the feature extraction part of a much more complex structure.

detection, semantic and instance segmentation (pixel-wise), and regression tasks.

2.2. Classification and regression approaches

When considering remote sensing data processed with DL-based algorithms, the following tasks can be highlighted: scene-wise classification, semantic and instance segmentation, and object detection. Scene-wise classification involves assigning a class label to each image (or patch), while the object detection task aims to draw bounding boxes around objects in an image (or patch) and labeling each of them according to the class label. Object detection can be considered a more challenging task since it requires to locate the objects in the image and then perform their classification. Another manner to detect objects in an image, instead of drawing bounding boxes, is to draw regions or structures around the boundary of objects, i.e., distinguish the class of the object at the pixel level. This task is known as semantic segmentation. However, in semantic segmentation, it is not possible to distinguish multiple objects of the same category, as each pixel receives one class label (Wu et al., 2020b). To overcome this drawback, a task that combines semantic segmentation and object detection named instance segmentation was proposed to detect multiple objects in pixel-level masks and labeling each mask with a class label (Thoma, 2016; Chen et al., 2016). The instance segmentation, however, consists of a method that, while classifying the image with this pixel-wise approach, is able to individualize objects (Sharma and Mir, 2020).

To produce a deep regression approach, the model needs to be adapted so that the last fully-connected layer of the architecture is changed to deal with a regression problem instead of a common classification one. With this adaptation, continuous values are estimated, differently from classification tasks. In comparison to classification, the regression task using DL is not often used; however, recent publications have shown its potential in remote sensing applications. One approach (Lathuilire et al., 2020) performed a comprehensive analysis of deep regression methods and pointed out that well-known fine-tuned networks, like VGG-16 (Simonyan and Zisserman, 2015) and ResNet-50 (He et al., 2016), can provide interesting results. These methods, however, are normally developed for specific applications, which is a drawback for general-purpose solutions. Another important point is that depending on the application, not always deep regression succeeds. A strategy is to discretize the output space and consider it as a classification solution. For UAV remote sensing applications, the strategy of using well-known networks is in general adopted. Not only VGG-16 and ResNet-50, as investigated by (Lathuilire et al., 2020), but also other networks including AlexNet (Krizhevsky et al., 2012) and VGG-11 have been used. An important issue that could be investigated in future research, depending on the application, is the optimizer. Algorithms with adaptive learning rates such as AdaGrad, RMSProp, AdaDelta (an extension of AdaGrad), and Adam are among the commonly used.

2.2.1. Scene-wise classification, object detection, and segmentation

Scene-wise classification or scene recognition refers to methods that associate a label/theme for one image (or patch) based on numerous images, such as in agricultural scenes, beach scenes, urban scenes, and others (Zou et al., 2015; Ma et al., 2019). Basic DNNs methods were developed for this task, and they are among the most common networks for traditional image recognition tasks. In remote sensing applications, scene-wise classification is not usually applied. Instead, most applications benefit more from object detection and pixel-wise semantic segmentation approaches. For scene-wise classification, the method needs only the annotation of the class label of the image, while other tasks like object detection method needs a drawn of a bounding box for all objects in an image, which makes it more costly to build labeled datasets. For instance or semantic segmentation, the specialist (i.e., the person who performs the annotation or object labeling) needs to draw a mask involving each pixel of the object, which needs more attention and precision in the annotation task, reducing, even more, the availability of datasets. Fig. 6 shows the examples of both annotation approaches (object detection and instance segmentation).

Object detection methods can be described into two mainstream categories: one-stage detectors (or regression-based methods) and two-stage detectors (or region proposal-based methods) (Zhao et al., 2019; Liu et al., 2019; Wu et al., 2020b). The usual two-stage object detection pipeline is to generate region proposals (candidate rectangular bounding boxes) on the feature map. It then classifies each one into an object class label and refines the proposals with a bounding box regression. A widely used strategy in the literature to generate proposals was proposed with the Faster-RCNN algorithm with the Region Proposal Network (RPN) (Zhao et al., 2019). Other state-of-the-art representatives of such algorithms are Cascade-RCNN (Cai and Vasconcelos, 2018), Trident-Net (Li et al., 2019), Grid-RCNN (Lu et al., 2019), Dynamic-RCNN (Zhang et al., 2020b), DetectoRS (Qiao et al., 2020). As for one-stage detectors, they directly make a classification and detect the location of objects without a region proposal classification step. This reduced component achieves a high detection speed for the models but tends to reduce the accuracy of the results. These are known as region-free detectors since they typically use cell grid strategies to divide the image and predict the class label of each one. Besides that, some detectors may serve for both one-stage and two-stage categories.

Object detection-based methods can be described in three components: a) backbone, which is responsible to extract semantic features from images; b) the neck, which is an intermediate component between the backbone and the head components, used to enrich the features obtained by the backbone, and; c) head component, which performs the detection and classification of the bounding boxes.

The backbone is a CNN that receives as input an image and outputs a feature map that describes the image with semantically features. In the DL, the state-of-the-art is composed of the following backbones: VGG (Simonyan and Zisserman, 2015), ResNet (He et al., 2016), ResNeXt (Xie et al., 2017), HRNet (Wang et al., 2020), RegNet (Radosavovic et al., 2020), Res2Net (Gao et al., 2021), and ResNesT (Zhang et al., 2020d). The neck component combines in several scales low-resolution and semantically strong features, capable of detecting large objects, with high-resolution and semantically weak features, capable of detecting small objects, which is done with the lateral and top-down connections of the convolutional layers of the Feature Pyramid Network (FPN) (Lin et al., 2017b), and its variants like PAFPN (Liu et al., 2018) and NAS-FPN (Ghiasi et al., 2019). Although FPN was originally designed to be a two-stage method, the methods' purpose was a manner to use the FPN on single-stage detectors by removing RPN and adding a classification subnet and a bounding box regression subnet. The head component is responsible for the detection of the objects with the softmax classification layer, which produces probabilities for all classes and a regression layer to predict the relative offset of the bounding box positions with the ground truth.

Despite the differences in object detectors (one or two-stage), their universal problem consists of dealing with a large gap between positive samples (foreground) and negative samples (background) during training, i.e. class imbalance problem that can deteriorate the accuracy results (Chen et al., 2020). In these detectors, the candidate bounding boxes can be represented into two main classes: positive samples, which are bounding boxes that match with the ground-truth, according to a metric; and negative samples, which do not match with the ground-truth. In this sense, a non-max suppression filter can be used to refine these dense candidates by removing overlaps to the most promising ones. The Libra-RCNN (Pang et al., 2019), ATSS (Zhang et al., 2019c), Guided Anchoring (Wang et al., 2019), FSAF (Zhu et al., 2019a), PAA (Kim and Lee, 2020), GFL (Li et al., 2020a), PISA (Cao et al., 2020) and VFNet (Zhang et al., 2020c) detectors explore different sampling strategies and new loss metrics to improve the quality of selected positive samples and reduce the weight of the large negative samples.

Another theme explored in the DL literature is the strategy of

Fig. 6. Labeled examples. The first-row consists of a bounding-box type of object detection approach label-example to identify individual tree-species in an urban environment. The second-row is a labeled-example of instance segmentation to detect rooftops in the same environment.

encoding the bounding boxes, which influences the accuracy of the one-stage detectors as they do not use region proposal networks (Zhang et al., 2020c). In this report (Zhang et al., 2020c), the authors represent the bounding boxes like a set of representatives or key-points and find the farthest top, bottom, left, and right points. CenterNet (Duan et al., 2019) detects the object center point instead of using bounding boxes, while CornerNet (Law and Deng, 2020) estimates the top-left corner and the bottom-right corner of the objects. SABL (Wang et al., 2020a) uses a chunk based strategy to discretize horizontally and vertically the image and estimate the offset of each side (bottom, up, left, and right). The VFNet (Zhang et al., 2020c) method proposes a loss function and a star-shaped bounding box (described by nine sampling points) to improve the location of objects.

Regarding semantic segmentation and instance segmentation approaches, they are generally defined as a pixel-level classification problem (Minaee et al., 2020). The main difference between semantic and instance is that the former one is capable to identify pixels belonging to one class but can not distinguish objects of the same class in the image. However, instance segmentation approaches can not distinguish overlapping of different objects, since they are concerned with identifying objects separately. For example, it may be problematic to identify in an aerial urban image the location of the cars, trucks, motorcycle, and the asphalt pavement which consists of the background or region in which the other objects are located. To unify these two approaches, a method was recently proposed in (Kirillov et al., 2019), named panoptic segmentation. With panoptic segmentation, the pixels that are contained in uncountable regions (e.g. background) receive a specific value indicating it.

Considering the success of the RPN method for object detection, some variants of Faster R-CNN were considered to instance segmentation as Mask R-CNN (He et al., 2017), which in parallel to bounding box regression branch add a new branch to predict the mask of the objects (mask generation). The Cascade Mask R-CNN (Cai and Vasconcelos, 2019) and HTC (Chen et al., 2019) extend Mask R-CNN to refine in a cascade manner the object localization and mask estimation. The PointRend (Kirillov et al., 2020) is a point-based method that reformulates the mask generation branch as a rendering problem to iteratively select points around the contour of the object. Regarding semantic segmentation, methods like U-Net (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017), DeepLabV3+ (Chen et al., 2018), and Deep Dual-domain Convolutional Neural Network (DDCN) (Nogueira et al., 2019) have also been regularly used and adapted for recent remote sensing investigations (Nogueira et al., 2020). Another important remote sensing approach that is been currently investigated is the segmentation of objects considering sparse annotations (Hua et al., 2021). Still, as of today, the CGnet (Wu et al., 2020a) and DLNet (Yin et al., 2020) are considered the state-of-art methods for semantic segmentation.

3. Deep learning in UAV imagery

To identify works related to DL in UAV remote sensing applications, we performed a search in the Web of Science (WOS) and Google Scholar databases. WOS is one of the most respected scientific databases and hosts a high number of scientific journals and publications. We conducted a search using the following string in the WOS: (“TS = ((deep learning OR CNN OR convolutional neural network) AND (UAV OR unmanned aerial vehicle OR drone OR RPAS) AND (remote sensing OR

photogrammetry)) AND LANGUAGE: (English) AND Types of Document: (Article OR Book OR Book Chapter OR Book Review OR Letter OR Proceedings Paper OR Review); Indexes = SCI-EXPANDED, SSCI, A% HCI, CPCI-S, CPCI-SSH, ESCI. Stipulated-time = every-years."). We considered DL, but added CNN, as it is one of the main DL-based architectures used in remote sensing applications (Ma et al., 2019). As such, published materials that use these terms in their titles, abstracts or keywords were investigated and included. For such reasons, we opted for this string to achieve a generalist investigation.

We filtered the results to consider only papers that implemented approaches with UAV-based systems. A total of 190 papers were found in the WOS database, where 136 were articles, 46 proceedings, and 10 reviews. An additional search was conducted in the Google Scholar database to identify works not detected in the WOS. We adopted the same combination of keywords in this search. We performed a detailed evaluation of its results and selected only those that, although from respected journals, were not encountered in the WOS search. This resulted in a total of 34 articles, 16 proceedings, and 8 reviews. The entire dataset was composed of 232 articles + proceedings and 18 reviews from scientific journals indexed in those bases. These papers were then organized and revised. Fig. 7 demonstrates the main steps to map this research. The encountered publications were registered only in the last five years (from 2016 to 2021), which indicates how recent UAV-based approaches integrated with DL methods are in the scientific journals.

The review articles gathered at those bases were separated and mostly used in the cloud text analysis of Fig. 1, while the remaining papers (articles and proceedings) were organized according to their category. A total of 283.785 words were analyzed for the word-cloud, as we removed words with less than 5% occurrences to cut lesser-used words unrelated to the theme, and higher than 95% occurrences to remove plain and simple words frequently used in the English language. The published articles and proceedings were divided in terms of DL-based networks (classification: scene-wise classification, segmentation, and object detection and; regression), sensor types (RGB, multispectral, hyperspectral, and LiDAR); and; applications (environmental, urban, and agricultural context). We also provided, in a subsequent section, datasets from previously conducted research for further investigation by novel studies. These datasets were organized and their characteristics were also summarized accordingly.

Most of our research was composed of publications from peer-review publishers in the area of remote sensing journals (Fig. 8). Even though the review articles encountered in the WoS and Google Scholar databases do mention, to some extent, UAV-based applications, none of them were dedicated to it. Towards the end of our paper, we examined state-of-the-art approaches, like real-time processing, data dimensionality reduction, domain adaptation, attention-based mechanisms, few-shot learning, open-set, semi-supervised and unsupervised learning, and others. This information provided an overview of the future opportunities and perspectives on DL methods applied in UAV-based images, where we discuss the implications and challenges of novel approaches.

The 232 papers (articles + proceedings) were investigated through a quantitative perspective, where we evaluated the number of occurrences per journal, the number of citations, year of publication, and location of the conducted applications according to country. We also prepared and organized a sampling portion in relation to the corresponding categories, as previously explained, identifying characteristics like architecture used, evaluation metric approach, task conducted, and type of sensor and mapping context objectives. After evaluating it, we adopted a qualitative approach by revising and presenting some of the applications conducted within the papers (UAV + DL) encountered in the scientific databases, summarizing the most prominent ones. This narrative over

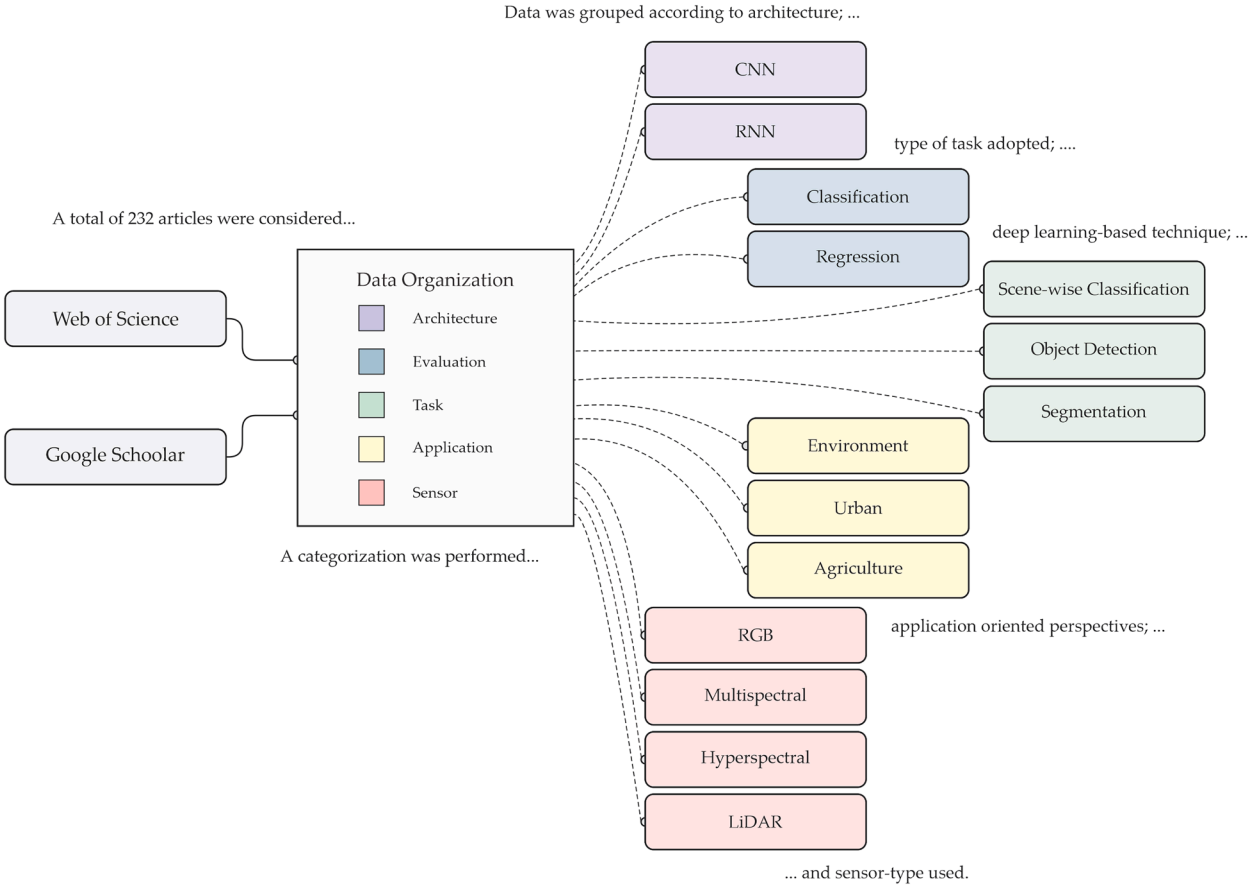


Fig. 7. The schematic procedure adopted to organize the revised material according to their respective categories as proposed in this review.

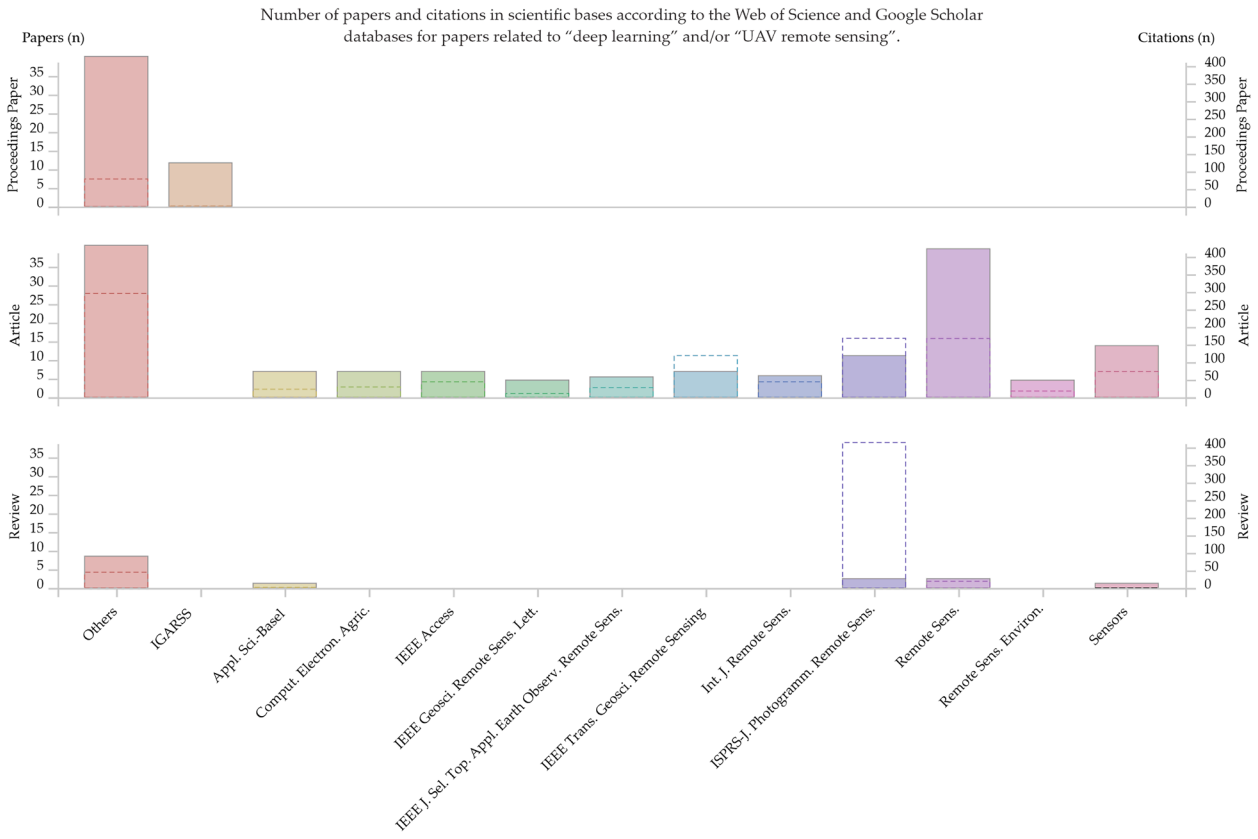


Fig. 8. The distribution of the evaluated scientific material according to data gathered at Web of Science (WOS) and Google Scholar databases. The y-axis on the left represents the number (n) of published papers, illustrated by solid-colored boxes. The y-axis on the right represents the number of citations that these publications, according to peer-review scientific journals, received since their publication, illustrated by dashed-lines of the same color to its corresponding solid-colored box.

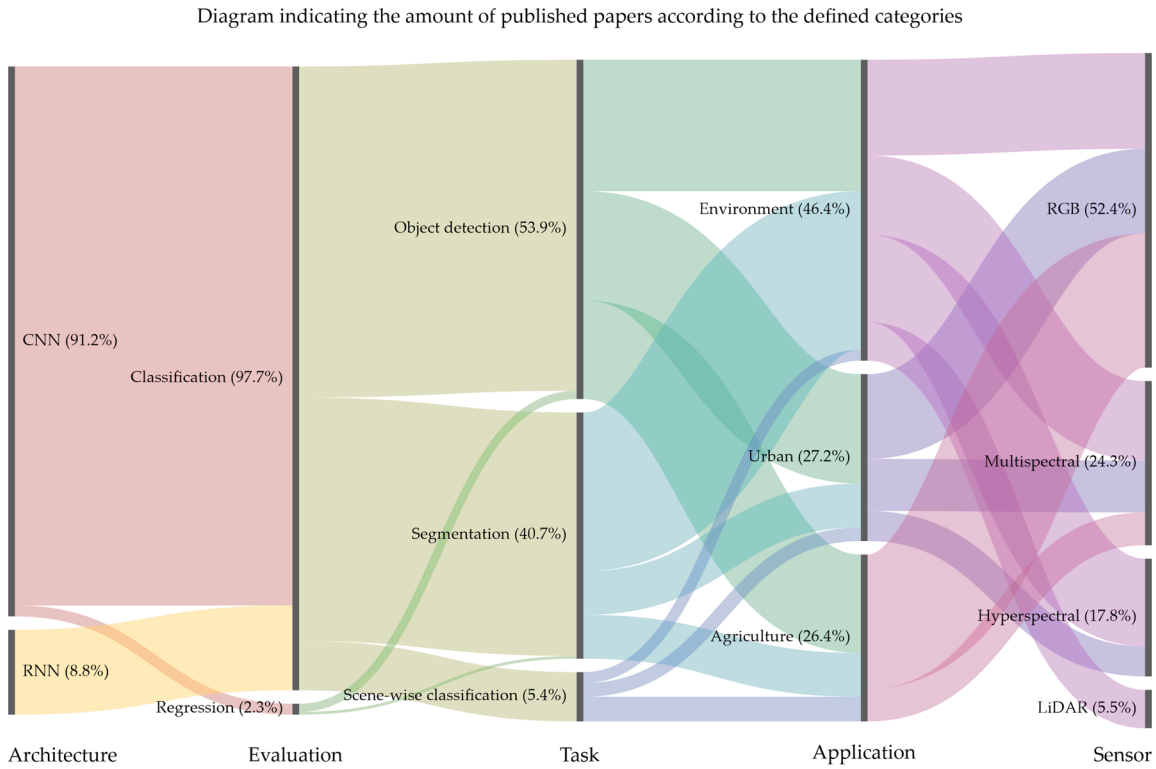


Fig. 9. Diagram describing proceedings and articles according to the defined categories using WOS and Google Scholar datasets.

these applications was separated accordingly to the respective categories related to the mapping context (environmental, urban, and agricultural). Later on, when presenting future perspectives and current trends in DL, we mentioned some of these papers alongside other investigations proposed at computer vision scientific journals that could be potentially used for remote sensing and UAV-based applications.

3.1. Sensors and applications worldwide

In the UAV-based imagery context, several applications were beneficiated from DL approaches. As these networks' usability is increasing throughout different remote sensing areas, researchers are also experimenting with their capability in substituting laborious-human tasks, as well as improving traditional measurements performed by shallow learning or conventional statistical methods. As of recently, several articles and proceedings were published in renowned scientific journals. In general terms, the articles collected at the scientific databases demonstrated a pattern related to its architecture (CNN or RNN), evaluation (classification or regression) approach (object detection, segmentation, or scene-wise classification), type of sensor (RGB, multispectral, hyperspectral or LiDAR) and mapping context (environmental, urban, or agricultural). These patterns can be viewed on a diagram (Fig. 9). The following observations can be extracted from this graphic:

1. The majority of networks in UAV-based applications still rely mostly on CNNs;
2. Even though object detection is the highest type of approach, there has been a lot of segmentation approaches in recent years;
3. Most of the used sensors are RGB, followed by multispectral, hyperspectral, and LiDAR, and;
4. There is an interesting amount of papers published within the environmental context, with forest-type related applications being the most common approach in this category, while both urban and agricultural categories were almost evenly distributed among opted approaches.

The majority of papers published on UAV-based applications implemented a type of CNN (91.2%). Most of these articles used established architectures (Fig. 5 and a small portion proposed their models and compared them against the state-of-the-art networks. In reality, this comparison appears to be a crucial concern regarding recent publications, since it is necessary to ascertain the performance of the proposed method in relation to well-known DL-based models. Still, the popularity of CNNs architecture in remote sensing images is not new, mainly because of reasons already stated in the previous sections. Besides that, even though presented in a small number of articles, RNNs (8.8%), mostly composed of CNN-LSTM architectures, are an emerging trend in this area and appear to be the focus of novel proposals. As UAV systems are capable of operating mostly according to the users' own desires (i.e., can acquire images from multiple dates in a more personalized manner), the same object is viewed through a type of time-progression approach. This is beneficial for many applications that include monitoring of stationary objects, like rivers, vegetation, or terrain slopes, for example.

Although classification (97.7%) tasks are the most common evaluation metrics implemented in these papers, regression (2.3%) is an important estimate and may be useful in future applications. The usage of regression metrics in remote sensing applications is worth it simply because it enables the estimation of continuous data. Applications that could benefit from regression analysis are present in environmental, urban, and agricultural contexts, as in many others, and it is useful to return predictions on measured variables. Classification, on the other hand, is more of a common ground for remote sensing approaches and it is implemented in every major task (object detection; pixel-wise semantic segmentation and scene-wise classification).

The aforementioned DL-based architectures were majorly applied in object detection (53.9%) and image segmentation (40.7%) problems, while (scene-wise) classification (5.4%) were the least common. This preference for object detection may be related to UAV-based data, specifically, since the high amount of detail of an object provided by the spatial resolution of the images is both an advantage and a challenge. It is an advantage because it increases the number of objects to be detected on the surface (thus, more labeled examples), and it is a challenge because it difficulties both the recognition and segmentation of these objects (higher detail implies more features to be extracted and analyzed). Classification (scene-wise), on the other hand, is not as common in remote sensing applications, and image segmentation is often preferred in some applications since assigning a class to each pixel of the image has more benefits for this type of analysis than rather only identifying a scene.

Following it, there is an interesting distribution pattern related to the application context. The data indicated that most of the applications were conducted in the environmental context (46.6%). This context includes approaches that aim to, in a sense, deal with detection and classification tasks on land use and change, environmental hazards and disasters, erosion estimates, wild-life detection, forest tree inventory, monitoring difficult to access regions, as others. Urban and agricultural categories (both 27.2% and 26.4%, respectively) were associated with car and traffic detection, buildings, street, and rooftop extraction, as well as plant counting, plantation-row detection, weed infestation identification, and others. Interestingly, all of the LiDAR data applications were related to environmental mapping, while RGB images were mostly used for urban, followed by the agricultural context. Multispectral and hyperspectral data, however, were less implemented in the urban context in comparison against the other categories. As these categories benefit differently from DL-based methods, a more detailed intake is needed to understand its problems, challenges, and achievements. In the following subsections, we explain these issues and advances while citing some suitable examples from within our search database.

Lastly, another important observation to be made regarding the categorization division used here is that there is a visible dichotomy between the types of sensor used. Most of the published papers in this area evaluating the performance of DL-based networks with RGB sensors (52.4%). This was, respectively, followed by multispectral (24.3%), hyperspectral (17.8%), and LiDAR (5.5%). The preference for RGB sensors in UAV-based systems may be associated with their low-cost and high market availability. As such, published articles may reflect on this, since it is a viable option for practical reasons when considering the replicability of the method. It should be noted that the number of labeled examples in public databases are mostly RGB, which helps improvements and investigation with this type of data. Moreover, data obtained from multispectral, hyperspectral, and LiDAR sensors are used in more specific applications, which contributes to this division.

Most of the object detection applications went on RGB types of data, while segmentation problems were dealt with both RGB, multispectral, hyperspectral, and LiDAR data. A possible explanation for this is that object detection often relies on the spatial, texture, pattern, and shape characteristics of the object in the image, as segmentation approaches are a diverse type of applications, which benefit from the amount of spectral and terrain information provided by these sensors. In object detection, DL-based methods may have potentialized the usage of RGB images, since simpler and traditional methods need additional spectral information to perform it. Also, apart from the spectral information, LiDAR, for example, offers important features of the objects for the networks to learn and refine the edges around them, specifically where their patterns are similar. Regardless, many of these approaches are related to the available equipment and nature of the application itself, so it is difficult to pinpoint a specific reason.

3.2. Environmental mapping

Environmental approaches with DNNs-based methods hold the most

diverse applications with remote sensing data, including UAV-imagery. These applications adopt different sensors simply because of their divergent nature. To map natural habits and their characteristics, studies often relied on methods and procedures specifically related to its goals, and no "universal" approach could be proposed nor discovered. However, although DL-based methods have not reached this type of "universal" approach, they are changing some skepticism by being successfully implemented in the most unique scenarios. Although UAV-based practices still offer some challenges to both classification and regression tasks, DNNs methods are proving to be generally capable of performing such tasks. Regardless, there is still much to be explored.

Several environmental practices could potentially benefit from deep networks like CNNs and RNNs. For example, monitoring and counting wild-life (Barbedo et al., 2020; Hou et al., 2020; Sundaram and Loganathan, 2020), detecting and classifying vegetation from grasslands and heavily-forested areas (Horning et al., 2020; Hamdi et al., 2019), recognizing fire and smoke signals (Alexandra Larsen et al., 2020; Zhang et al., 2019a), analyzing land use, land cover, and terrain changes, which are often implemented into environmental planning and decision-making models (Kussul et al., 2017; Zhang et al., 2020e), predicting and measuring environmental hazards (Dao et al., 2020; Bui et al., 2020), among others. What follows is a brief description of recent material published in the remote sensing scientific journals that aimed to solve some of these problems by integrating data from UAV embedded sensors with DL-based methods.

One of the most common approaches related to environmental remote sensing applications regards land use, land cover, and other types of terrain analysis. A recent study (Giang et al., 2020) applied semantic segmentation networks to map land use over a mining extraction area. Another one, (Al-Najjar et al., 2019), combined information from a Digital Surface Model (DSM) with UAV-based RGB images and applied a type of feature fusion as input for a CNN model. To map coastal regions, an approach (Buscombe and Ritchie, 2018), with RGB data registered at multiple scales, used a CNN in combination with a graphical method named conditional random field (CRF). Another research (Park and Song, 2020), with hyperspectral images in combination between 2D and 3D convolutional layers, was developed to determine the discrepancy of land cover in the assigned land category of cadastral map parcels.

With a semantic segmentation approach, road extraction by a CNN was demonstrated in another investigation (Li et al., 2019b). Another study (Gevaert et al., 2020) investigated the performance of a FCN to monitor household upgrading in unplanned settlements. Terrain analysis is a diversified topic in any type of cartographic scale, but for UAV-based images, in which most data acquisitions are composed by a high-level of detail, DL-based methods are resulting in important discoveries, demonstrating the feasibility of these methods to perform this task. Still, although these studies are proving this feasibility, especially in comparison with other methods, novel research should focus on evaluating the performance of deep networks regarding their domain adaptation, as well as its generalization ability, like using data in different spatial resolutions, multitemporal imagery, etc.

The detection, evaluation, and prediction of flooded areas represents another type of investigation with datasets provided by UAV-embedded sensors. A study (Gebrehiwot et al., 2019) demonstrated the importance of CNNs for the segmentation of flooded regions, where the network was able to separate water from other targets like buildings, vegetation, and roads. One potential application that could be conducted with UAV-based data, but still needs to be further explored, is mapping and predicting regions of possible flooding with a multitemporal analysis, for example. This, as well as many other possibilities related to flooding, water-bodies, and river courses (Carbonneau et al., 2020), could be investigated with DL-based approaches.

For river analysis, an investigation (Zhang et al., 2020f) used a CNN architecture for image segmentation by fusing both the positional and channel-wise attentive features to assist in river ice monitoring. Another study (Jakovljevic et al., 2019) compared LiDAR data with point cloud generated by UAV mapping and demonstrated an interesting approach to DL-based methods applications for point cloud classification and a rapid Digital Elevation Model (DEM) generation for flood risk mapping. One type of application with CNN in UAV data involved measuring hailstones in open areas (Soderholm et al., 2020). For this approach, image segmentation was used in RGB images and returned the maximum dimension and intermediate dimension of the hailstones. Lastly, on this topic, a comparison (Ichim and Popescu, 2020) with CNNs and GANs to segment both river and vegetation areas demonstrated that a type of "fusion" between these networks using a global classifier had an advantage of increasing the efficiency of the segmentation.

UAV-based forest mapping and monitoring is also an emerging approach that has been gaining the attention of the scientific community and, at some level, governmental bodies. Forest areas often pose difficulties for precise monitoring and investigation, since they can be hard to access and may be dangerous to some extent. In this aspect, images taken from UAV embedded sensors can be used to identify single tree-species in forested environments and compose an inventory. From the papers gathered, multiple types of sensors, RGB, both multi and hyperspectral, and LiDAR, were used for this approach. An application investigated the performance of a 3D-CNN method to classify tree species in a boreal forest, focusing on pine, spruce, and birch trees, with a combination between RGB and hyperspectral data (Nezami et al., 2020).

Single-tree detection and species classification by CNNs were also investigated in (Ferreira et al., 2020) in which three types of palm-trees in the Amazon forest, considered important for its population and native communities, were mapped with this type of approach. Another example (Hu et al., 2020) includes the implementation of a Deep Convolutional Generative Adversarial Network (DCGAN) to discriminate between health diseased pinus-trees in a heavily-dense forested park area. Another recent investigation (Miyoshi et al., 2020) proposed a novel DL method to identify single-tree species in highly-dense areas with UAV- hyperspectral imagery. These and other scientific studies demonstrate how well DL-based methods can deal with such environments.

Although the majority of approaches encountered at the databases of this category relate to tree-species mapping, UAV-acquired data were used for other applications in these natural environments. A recent study (Zhang et al., 2020a) proposed a method based on semantic segmentation and scene-wise classification of plants in UAV-based imagery. The method bases itself on a CNN that classifies individual plants by increasing the image scale while integrating features learned from small scales. This approach is an important intake in multi-scale information fusion. Also related to vegetation identification, multiple CNNs architectures were investigated in (Hamylton et al., 2020) to detect between plants and non-type of plants with UAV-based RGB images achieving interesting performance.

Another application aside from vegetation mapping involves wild-life identification. Animal monitoring in open spaces and grasslands is also something that received attention as DL-based object detection and semantic segmentation methods are providing interesting outcomes. A paper by (Kellenberger et al., 2018) covers this topic and discusses, with practical examples, how CNNs may be used in conjunction with UAV-based images to recognize mammals in the African Savannah. This study relates the challenges related to this task and proposes a series of suggestions to overcome them, focusing mostly on imbalances in the labeled dataset. The identification of wild-life, also, was not only performed in terrestrial environments, but also in marine spaces, where a recent publication (Gray et al., 2019) implemented a CNN-based semantic segmentation method to identify cetacean species, mainly blue, humpback, and minke whales, in the ocean. These studies not only demonstrate that such methods can be highly accurate at different tasks but also imply the potential of DL approaches for UAVs in the current literature.

3.3. Urban mapping

For urban environments, many DL-based proposals with UAV data have been presented in the literature in the last years. The high-spatial-resolution easily provided by UAV embedded sensors are one of the main reasons behind its usage in these areas. Object detection and instance segmentation methods in those images are necessary to individualize, recognize, and map highly-detailed targets. Thus, many applications rely on CNNs and, in small cases, RNNs (CNN-LSTM) to deal with them. Some of the most common examples encountered in this category during our survey are the identification of pedestrians, car and traffic monitoring, segmentation of individual tree-species in urban forests, detection of cracks in concrete surfaces and pavements, building extraction, etc. Most of these applications were conducted with RGB type of sensors, and, in a few cases, spectral ones.

The usage of RGB sensors is, as aforementioned, a preferred option for small-budget experiments, but also is related to another important preference of CNNs, and that is that features like pixel-size, form, and texture of an object are essential to its recognition. In this regard, novel experiments could compare the performance of DL-based methods with RGB imagery with other types of sensors. As low-budget systems are easy to implement in larger quantities, many urban monitoring activities could benefit from such investigations. In urban areas, the importance of UAV real-time monitoring is relevant, and that is one of the current objectives when implementing such applications.

The most common practices on UAV-based imagery in urban environments with DL-based methods involve the detection of vehicles and traffic. Car identification is an important task to help urban monitoring and may be useful for real-time analysis of traffic flow in those areas. It is not an easy task, since vehicles can be occluded by different objects like buildings and trees, for example. A recent approach using RGB video footage obtained with UAV, as presented in (Zhang et al., 2019b), used an object detection CNN for this task. They also dealt with differences in traffic monitoring to motorcycles, where a frame-by-frame analysis enabled the neural network to determine if the object in the image was a person (pedestrian) or a person riding a motorcycle since differences in its pattern and frame-movement indicated it. Regarding pedestrian traffic, an approach with thermal cameras presented by (de Oliveira and Wehrmeister, 2018) demonstrated that CNNs are appropriate to detect persons with different camera rotations, angles, sizes, translation, and scale, corroborating the robustness of its learning and generalization capabilities.

Another important survey in those areas is the detection and localization of single-tree species, as well as the segmentation of their canopies. Identifying individual species of vegetation in urban locations is an important requisite for urban-environmental planning since it assists in inventorying species and providing information for decision-making models. A recent study (Santos et al., 2019) applied object detection methods to detect and locate tree-species threatened by extinction. Following their intentions, a research (Torres et al., 2020) evaluated semantic segmentation neural networks to map endangered tree-species in urban environments. While one approach aimed to recognize the object to compose an inventory, the other was able to identify it and return important metrics, like its canopy-area for example. Indeed, some proposals that were implemented in a forest type of study could also be adopted in urban areas, and this leaves an open field for future research that intends to evaluate DL-based models in this environment. Urban areas pose different challenges for tree monitoring, so these applications need to consider their characteristics.

DL-based methods have also been used to recognize and extract infrastructure information. An interesting approach demonstrated by (Boonpook et al., 2021), based on semantic segmentation methods, was able to extract buildings in heavily urbanized areas, with unique architectural styles and complex structures. Interestingly enough, a combination of RGB with a DSM improved building identification, indicating that the segmentation model was able to incorporate appropriate information related to the objects' height. This type of combinative approach, between spatial-spectral data and height, may be useful in other identification and recognition approaches. Also regarding infrastructure, another possible application in urban areas is the identification and location of utility poles (Gomes et al., 2020). This application, although being of rather a specific example, is important to maintain and monitor the conditions of poles regularly. These types of monitoring in urban environments is something that benefits from DL-based models approaches, as it tends to substitute multiple human inspection tasks. Another application involves detecting cracks in concrete pavements and surfaces (Bhowmick et al., 2020). Because some regions of civil structures are hard to gain access to UAV-based data with object detection networks may be useful to this task, returning a viable real-life application.

Another topic that is presenting important discoveries relates to land cover pixel segmentation in urban areas, as demonstrated by (Benjdira et al., 2019a). In this investigation, an unsupervised domain adaptation method based on GANs was implemented, working with different data from UAV-based systems, while being able to improve image segmentation of buildings, low vegetation, trees, cars, and impervious surfaces. As aforementioned, GANs or DCGANs are quickly gaining the attention of computer vision communities due to their wide area of applications and the way they function by being trained to differentiate between real and fake data (Goodfellow et al., 2014). Regardless, its usage in UAV-based imagery is still underexplored, and future investigations regarding not only land change and land cover but also other types of applications' accuracies may be improved with them. Nonetheless, apart from differences in angles, rotation, scales, and other UAV-based imagery-related characteristics, diversity in urban scenarios is a problem that should be considered by unsupervised approaches. Therefore, in the current state, DL-based networks still may rely on some supervised manner to guide image processing, specifically regarding domain shift factors.

3.4. Agricultural mapping

Precision agriculture applications have been greatly benefited from the integration between UAV-based imagery and DL methods in recent scientific investigations. The majority of issues related to these approaches involve object detection and feature extraction for counting plants and detecting plantation lines, recognizing plantation-gaps, segmentation of plant species and invasive species such as weeds, phenology, and phenotype detection, and many others. These applications offer numerous possibilities for this type of mapping, especially since most of these tasks are still conducted manually by human-vision inspection. As a result, they can help precision farming practices by returning predictions with rapid, unbiased, and accurate results, influencing decision-making for the management of agricultural systems.

Regardless, although automatic methods do provide important information in this context, they face difficult challenges. Some of these include similarity between the desired plant and invasive plants, hard-to-detect plants in high-density environments (i.e. presenting small spacing between plants and lines), plantation-lines that do not follow a straight-path, edge-segmentation in mapping canopies with conflicts between shadow and illumination, and many others. Still, novel investigations aim to achieve a more generative capability to these networks in dealing with such problems. In this sense, approaches that implement methods in more than one condition or plantation are being the main focus of recent publications. Thus, varied investigation scenarios are currently being proposed, with different types of plantations, sensors, flight-altitudes, angles, spatial and spectral divergences, dates, phenological-stages, etc.

An interesting approach that has the potential to be expanded to different orchards was used in (Apolo-Apolo et al., 2020). There, a low-altitude flight approach was adopted with side-view angles to map yield by counting fruits with the CNN-based method. Counting fruits is not

something entirely new in DL-based approaches, some papers demonstrated the effectiveness of bounding-box and point-feature methods to extract it (Biffi et al., 2021; Tian et al., 2019b; Kang and Chen, 2020) aside from several differences in occlusion, lightning, fruit size, and image corruption.

Today's deep networks demonstrate high potential in yield-prediction, as some applications are adapted to CNN architectures mainly because of its benefits in image processing. One of which includes predicting pasture-forage with only RGB images (Castro et al., 2020). Another interesting example in crop-yield estimates is presented by (Nevavuori et al., 2020), where a CNN-LSTM was used to predict yield with a spatial multitemporal approach. There the authors implemented this structure since RNNs are more appropriate to learn from temporal data, while a 3D-CNN was used to process and classify the image. Although used less frequently than CNNs in the literature, there is emerging attention to LSTM architectures in precision agriculture approaches, which appear to be an appropriate intake for temporal monitoring of these areas.

Nonetheless, one of the most used and beneficiated approaches in precision agriculture with DL-based networks is counting and detecting plants and plantation lines. Counting plants is essential to produce estimates regarding production rates, as well as, by geolocating it, determine if a problem occurred during the seedling process by identifying plantation-gaps. In this regard, plantation-lines identification with these gaps is also a desired application. Both object detection and image segmentation methods were implemented in the literature, but most approaches using image semantic segmentation algorithms rely on additional procedures, like using a blob detection method (Kitano et al., 2019), for example. These additional steps may not always be desirable, and to prove the generality capability of one model, multiple tests at different conditions should be performed.

For plantation-line detection, segmentations are currently being implemented and often used to assist in more than one information extraction. In (Osco et al., 2021) semantic segmentation methods were applied in UAV-based multispectral data to extract canopy areas and was able to demonstrate which spectral regions were more appropriate to it. A recent application with UAV-based data was also proposed in (Osco et al., 2020a), where a CNN model is presented to simultaneously count and detect plants and plantation-lines. This model is based on a confidence map extraction and was an upgraded version from previous research with citrus-tree counting (Osco et al., 2020b). This CNN works by implementing some convolutional layers, a Pyramid Pooling Module (PPM) (Zhao et al., 2017), and a Multi-Stage Module (MSM) with two information branches that, concatenated at the end of the MSM processes, shares knowledge learned from one to another. This method ensured that the network learned to detect plants that are located at a plantation-line, and understood that a plantation-line is formed by linear conjunction of plants. This type of method has also been proved successful in dealing with highly-dense plantations. Another research (Ampatzidis and Partel, 2019) that aimed to count citrus-trees with a bounding-box-based method also returned similar accuracies. However, it was conducted in a sparse plantation, which did not impose the same challenges faced at (Osco et al., 2020b; Osco et al., 2020a). Regardless, to deal with highly dense scenes, feature extraction from confidence maps appears to be an appropriate approach.

However, agricultural applications do not always involve plant counting or plantation-line detection. Similar to wild-animal identification as included in other published studies (Kellenberger et al., 2018; Gray et al., 2019), there is also an interest in cattle detection, which is still an onerous task for human-inspection. In UAV-based imagery, some approaches included DL-based bounding-boxes methods (Barbedo et al., 2019), which were also successfully implemented. DNNs used for this task are still underexplored, but published investigations (Rivas et al., 2018) argue that one of the main reasons behind the necessity to use DL methods is based on occurrences of changes in terrain (throughout the seasons of the year) and the non-uniform distribution of the animals throughout the area. On this matter, one interesting approach should involve the usage of real-time object detection on the flight. This is because it is difficult to track animal movement, even in open areas such as pastures, when a UAV system is acquiring data. Another agricultural application example refers to the monitoring offshore aquaculture farms using UAV-underwater color imagery and DL models to classify them (Bell et al., 2020). These examples reveal the widespread variety of agriculture problems that can be attended with the integration of DL models and UAV remote sensing data.

Lastly, a field yet to be also explored in the literature is the identification and recognition of pests and disease indicators in plants using DL-based methods. Most recent approaches aimed to identify invasive species, commonly named "weeds", in plantation-fields. In a demonstration with unsupervised data labeling, (Dian Bah et al., 2018) evaluated the performance of a CNN-based method to predict weeds in the plantation lines of different crops. This pre-processing step to automatically generate labeled data, which is implemented outside the CNN model structure, is an interesting approach. However, others prefer to include a "one-step" network to deal with this situation, and different fronts are emerging in the literature. Unsupervised domain adaptation, in which the network extracts learning features from new unviewed data, is one of the most current aimed models.

A recent publication (Li et al., 2020b) proposed it to recognize and count in-field cotton-boll status identification. Regardless, with UAV-based data examples, this is still an issue. As for disease detection, a study (Kerkech et al., 2020) investigated the use of image segmentation for vine-crops with multispectral images, and was able to separate visible symptoms (RGB), infrared symptoms (i.e. when considering only the infrared band) and in an intersection between visible and infrared spectral data. Another interesting example regarding pests identification with UAV-based image was demonstrated in (Tetila et al., 2020) where superpixel image samples of multiple pest species were considered, and activation filters used to recognize undesirable visual patterns implemented alongside different DL-based architectures.

4. Publicly available UAV-based datasets

As mentioned, one of the most important characteristics of DL-based methods is that they tend to increase their learning capabilities as a number of labeled examples are used to train a network. In most of the early approaches to remote sensing data, CNNs were initialized with pre-trained weights from publicly available image repositories over the internet. However, most of these repositories are not from data acquired with remote sensing platforms. Still, there are some known aerial repositories with labeled examples, which were presented in recent years, such as the DOTA (Xia et al., 2018), UAVDT (Du et al., 2018), VisDrone (Zhu et al., 2019), WHU-RS19 (Sheng et al., 2012), RSSCN7 (Zou et al., 2015), RSC11 (Zhao et al., 2016), Brazilian Coffee Scene (Penatti et al., 2015) datasets. These and others are gaining notoriety in UAV-based applications and could be potentially used to pre-train or benchmark DL methods. These datasets not only serve as an additional option to start a network but also may help in novel proposals to be compared against the evaluated methods.

Since there is a still scarce amount of labeled examples with UAV-acquired data, specifically in multispectral and hyperspectral data, we aimed to provide UAV-based datasets in both urban and rural scenarios for future research to implement and compare the performance of novel DL-based methods with them. Table 1 summarizes some of the information related to these datasets, as well as indicates recent publications in which previously conducted approaches were implemented, as well as the results achieved on them. They are available on the following webpage, which is to be constantly updated with novel labeled datasets from here on: Geomatics and Computer Vision/Datasets

Table 1
UAV-based datasets that are publically available from previous research.

Reference	Task	Target	Sensor	$GSD_{(cm)}$	Best Method	Result
(Santos et al., 2019)	Detection	Trees	RGB	0.82	RetinaNet	AP = 92.64%
(Torres et al., 2020)	Segmentation	Trees	RGB	0.82	FC-DenseNet	F1 = 96.0%
(Osco et al., 2021)	Segmentation	Citrus	Multispectral	12.59	DDCN	F1 = 94.4%
(Osco et al., 2020a)	Detection	Citrus	RGB	2.28	(Osco et al., 2020a)	F1 = 96.5%
(Osco et al., 2020a)	Detection	Corn	RGB	1.55	(Osco et al., 2020a)	F1 = 87.6%
(Osco et al., 2020b)	Detection	Citrus	Multispectral	12.59	(Osco et al., 2020b)	F1 = 95.0%

5. Perspectives in deep learning with UAV data

There is no denying that DL-based methods are a powerful and important tool to deal with the numerous amounts of data daily produced by remote sensing systems. What follows in this section is a short commentary on the near perspectives of one of the most emerging fields in the DL and remote sensing communities that could be implemented with UAV-based imagery. These topics, although individually presented here, have the potential to be combined, as already performed in some studies, contributing to the development of novel approaches.

In general, DL architectures require low resolution input images (e.g., 512 × 512 pixels). High resolution images are generally scaled to the size required for processing. However, UAVs have the advantage of capturing images in higher resolution than most other types of sensing platforms aside from proximal sensing, and the direct application of traditional architectures may not take advantage of this feature. As such, processing images with DL while maintaining high resolution in deeper layers is a challenge to be explored. In real-time applications, such as autonomous navigation, this processing must be fast, which opens up a range of research related to reducing the complexity of architectures while preserving accuracy. Regarding DL, recently, some CNN architectures that try to maintain high resolution in deeper layers, such as HRNet, have been proposed (Kannojia and Jaiswal, 2018). These novel architectures can really take advantage of the high resolution from UAV images compared to commonly available orbital data.

To summarize, the topics addressed in this section compose some of the hot topics in the computer vision community, and the combination of them with remote sensing data can contribute to the development of novel approaches in the context of UAV mapping. In this regard, it is important to emphasize that not only these topics are currently being investigated by computer vision research, but that they also are being fastly implemented in multiple approaches aside from remote sensing. As other domains are investigated, novel ways of improving and adapting these networks can be achieved. Future studies in remote sensing communities, specifically on UAV-based systems, may benefit from these improvements and incorporate them into their applications.

5.1. Real-time processing

Most of the environmental, urban, and agricultural applications presented in this study can benefit from real-time responses. Although UAV and DL-based combinations speed up the processing pipeline, these algorithms are highly computer-intensive. Usually, they do require post-processing in data centers or dedicated Graphics Processing Units (GPUs) machines. Although DL is considered a fast method to extract information from data after its training, it still bottlenecks real-time applications mainly because of the number of layers intrinsic to the DL methods architecture. Research groups, especially from the IoT industry/academy, race to develop real-time DL methods because of it. The approach usually goes in two directions: developing faster algorithms and developing dedicated GPU processors.

DL models use 32-bit floating points to represent the weights of the neural network. A simple strategy known as quantization reduces the amount of memory required by DL models representing the weights, using 16, 8, or even 1 bit instead of 32-bits floating points. A 32-bit full precision ResNet-18 (He et al., 2016) achieves 89.2% top-5 accuracy on the ImageNet dataset (ImageNet, 2018), while the ResNet-18 (He et al., 2016) ported to XNOR-Net achieves 73.2% top-5 accuracy in the same dataset. The quantization goes beyond weights, in all network components, while the literature reports activation functions and gradient optimizations quantized methods. The survey conducted in (Guo, 2018) gives an important overview of quantization methods. Also, knowledge distillation (Hinton et al., 2015) is another example of a training model using a smaller network, where a larger "teacher" network guides the learning process of a smaller "student" network.

Another strategy to develop fast DL models is to design layers with fewer parameters that are still capable of retaining predictive performance. MobileNets (Howard et al., 2017) and its variants are a good example of this idea. In specific tasks, such as object detection, it is possible to develop architectural enhancements for this approach, such as the Context Enhanced Module (CEM) and the Spatial Attention Module (SAM) (Qin et al., 2019). When considering even smaller computational power, it is possible to find DL running on microcontroller units (MCU) where the memory and computational power are 3–4 orders of magnitude smaller than mobile phones.

On hardware, the industry has already developed embedded AI platforms that run DL algorithms. NVIDIA's Jetson is amongst the most popular choices and a survey (Mittal, 2019) of studies using the Jetson platform and its applications demonstrate it. Also, a broader survey on this theme, that considers GPU, ASIC, FPGA, and MCUs of AI platforms, can be read in (Imran et al., 2020). Regardless, research in the context of UAV remote sensing is quite limited, and there is a gap that can be fulfilled by future works. Several applications can be benefited by this technology, including, for example, agricultural spraying UAV, which can recognize different types of weeds in real-time, and simultaneously use the spray. Other approaches may include real-time monitoring of trees in both urban and forest environments, as well as the detection of other types of objects that benefit from a rapid intake.

5.2. Dimensionality reduction

Due to recent advances in capture devices, hyperspectral images can be acquired even in UAVs. These images consist of tens to hundreds of spectral bands that can assist in the classification of objects in a given application. However, two main issues arise from the high dimensionality: i) the bands can be highly correlated, and ii) the excessive increase in the computational cost of DL models. High-dimensionality could invoke a problem known as the Hughes phenomenon, which is also known as the curse of dimensionality, i.e., when the accuracy of a classification is reduced due to the introduction of noise and other implications encountered in hyperspectral or high-dimensional data (Hennessy et al., 2020). Regardless, hyperspectral data may pose an hindrance for the DL-based approaches accuracies, thus being an important issue to be considered in remote sensing practices. The classic approach to address high dimensionality is by applying a Principal Component Analysis (PCA) (Licciardi et al., 2012).

Despite several proposals, PCA is generally not applied in conjunction with DL, but as a pre-processing step. Although this method may be one of the most known approaches to reduce dimensionality when dealing with hyperspectral data, different intakes were already

presented in the literature. A novel DL approach, implemented with UAV-based imagery, was demonstrated by Miyoshi et al. (2020). There, the authors proposed a one-step approach, conducted within the networks' architecture, to consider a combination of bands of a hyperspectral sensor that were highly related to the labeled example provided in the input layer at the initial stage of the network. Another investigation (Vaddi and Manoharan, 2020) combines a band selection approach, spatial filtering, and CNN to simultaneously extract the spectral and spatial features. Still, the future perspective to solve this issue appears to be a combination of spectral band selection and DL methods in an end-to-end approach. Thus, both selection and DL methods can exchange information and improve results. This can also contribute to understanding how DL operates with these images, which was slightly accomplished at Miyoshi et al. (2020).

5.3. Domain adaptation and transfer learning

The training steps of DL models are generally carried out on images captured in a specific geographical region, in a short-time period, or on single capture equipment (also known as domains). When the model is used in practice, it is common for spectral shifts to occur between the training and test images due to differences in acquisition, geographic region, atmospheric conditions, among others (Tuia et al., 2016). Domain adaptation is a technique for adapting models trained in a source domain to a different, but still related, target domain. Therefore, domain adaptation is also viewed as a particular form of transfer learning (Tuia et al., 2016). On the other hand, transfer learning (Zhuang et al., 2020; Tan et al., 2018) does include applications in which the characteristics of the domain's target space may differ from the source domain.

A promising research line for domain adaptation and transfer learning is to consider GANs (Goodfellow et al., 2014; Elshamli et al., 2017). For example, (Benjdira et al., 2019b) proposed the use of GANs to convert an image from the source domain to the target domain, causing the source images to mimic the characteristics of the images from the target domain. Recent approaches seek to align the distribution of the source and target domains, although they do not consider direct alignment at the level of the problem classes. Approaches that are attentive to class-level shifts may be more accurate, as the category-sensitive domain adaptation proposed by (Fang et al., 2019). Thus, these approaches reduce the domain shift related to the quality and characteristics of the training images and can be useful in practice for UAV remote sensing.

5.4. Attention-based mechanisms

Attention mechanisms aim to highlight the most valuable features or image regions based on assigning different weights for them in a specific task. It is a topic that has been recently applied in remote sensing, providing significant improvements. As pointed out by (Xu et al., 2018), high-resolution images in remote sensing provide a large amount of information and exhibit minor intra-class variation while it tends to increase. These variations and a large amount of information make extraction of relevant features more difficult, since traditional CNNs process all regions with the same weight (relevance). Attention mechanisms, such as the one proposed by (Xu et al., 2018), are useful tools to focus the feature extraction in discriminative regions of the problem, be it image segmentation (Ding et al., 2021; Su et al., 2019; Zhou et al., 2020), scene-wise classification (Zhu et al., 2019b; Li et al., 2020c), or object detection (Li et al., 2019; Li et al., 2020c), as others.

Besides, (Su et al., 2019) argue that when remote sensing images are used, they are generally divided into patches for training the CNNs. Thus, objects can be divided into two or more sub-images, causing the discriminative and structural information to be lost. Attention mechanisms can be used to aggregate learning by focusing on relevant regions that describe the objects of interest, as presented in (Su et al., 2019), through a global attention upsample module that provides global context and combines low and high-level information. Recent advances in computer vision were achieved with attention mechanisms for classification (e.g., Vision Transformer (Dosovitskiy et al., 2020) and Data-efficient Image Transformers (Touvron et al., 2020)) and in object detection (e.g., DETR (Carion et al., 2020)) that have not yet been fully evaluated in remote sensing applications. Some directions also point to the use of attention mechanisms directly in a sequence of image patches (Dosovitskiy et al., 2020; Touvron et al., 2020). These new proposals can improve the results already achieved in remote sensing data, just as they have advanced the results on the traditional image datasets in computer vision (e.g., ImageNet (ImageNet, 2018)).

5.5. Few-shot learning

Although recent materials demonstrated the feasibility of DL-based methods for multiple tasks, they still are considered limited in terms of high generalization. This occurs when dealing with the same objects in different geographical areas or when new object classes are considered. Traditional solutions require retraining the model with a robust labeled dataset for the new area or object. Few-shot learning aims to cope with situations in which few labeled datasets are available. A recent study (Li et al., 2020), in the context of scene classification, pointed out that few-shot methods in remote sensing are based on transfer learning and meta-learning. Meta-learning can be more flexible than transfer learning, and when applied in the training set to extract meta-knowledge, contributes significantly to few-shot learning in the test set. An interesting strategy to cope with large intraclass variation and interclass similarity is the implementation of the attention mechanism in the feature learning step, as previously described. The datasets used in the (Li et al., 2020) study were not UAV-based; however, the strategy can be explored in UAV imagery.

In the context of UAV remote sensing, there are few studies on few-shot learning. Recently, an investigation (Karami et al., 2020) aimed for the detection of maize plants using the object detection method CenterNet. The authors adopted a transfer learning strategy using pre-trained models from other geographical areas and dates. Fewer images (in total, 150 images), when compared to the previous training (with 600 images), from the new area were used for fine-tuning the model. Based on the literature survey, there is a research-gap to be further explored in the context of object detection using few-shot learning in UAV remote sensing. The main idea behind this is to consider less labeled datasets for training, which may help in some remote applications where data availability is scarce or presents few occurrences.

5.6. Semi-supervised learning and unsupervised learning

With the increasing availability of remote sensing images, the labeling task for supervised training of DL models is expensive and time-consuming. Thus, the performance of DL models is impacted due to the lack of large amount of labeled training images. Efforts have been made to consider unlabeled images in training through unsupervised (unlabeled images only) and semi-supervised (labeled and unlabeled images) learning. In remote sensing, most semi-supervised or unsupervised approaches are based on transfer learning, which usually requires a supervised pre-trained model (Liu and Qin, 2020). In this regard, a recent study (Kang et al., 2020) proposed a promising approach for unlabeled remote sensing images that define spatial augmentation criteria for relating close sub-images. Regardless, this is still an underdeveloped practice with UAV-based data and should be investigated in novel approaches.

Future perspectives point to the use of contrastive loss (Bachman et al., 2019; Tian et al., 2019a; Hjelm et al., 2019; He et al., 2020) and clustering-based approaches (Caron et al., 2018; Caron et al., 2021). Recent publications have shown interesting results with the use of contrastive loss that has not yet been fully evaluated in remote sensing. For example, (He et al., 2020) proposed an approach based on

contrastive loss that surpassed the performance of its supervised pre-trained counterpart. As for clustering-based methods, they often group images with similar characteristics (Caron et al., 2018). On this matter, a research (Caron et al., 2018) presented an approach that groups the data while reinforcing the consistency between the cluster assignments produced for a pair of images (same images with two augmentations). An efficient and effective way to use a large number of unlabeled images can considerably improve the performance, mainly related to the generalizability of the models.

5.7. Multitask learning

Multitask learning aims to perform multiple tasks simultaneously. Several advantages are mentioned in (Crawshaw, 2020), including fast learning and the minimization of overfitting problems. Recently, in the context of UAV remote sensing, there were some important researches already developed. A study (Wang et al., 2021) proposed a method to conduct three tasks (semantic segmentation, height estimation, and boundary detection), which also considered boundary attention modules. Another research (Osco et al., 2020a) simultaneously detecting plants and plantation lines in UAV-based imagery. The proposed network benefited from the contributions of considering both tasks in the same structure, since the plants must, essentially belong to a plantation line. In short, improvements occurred in the detection task when line detection was considered at the same time. This approach can be further explored in several UAV-based remote sensing applications.

5.8. Open-set

The main idea of an open-set is to deal with unknown or unseen classes during the inference in the testing set (Bendale and Boult, 2016). As the authors mention, recognition in real-world scenarios is "open-set", different from neural networks' nature, which is in a "close-set". Consequently, the testing set is classified considering only the classes used during the training. Therefore, unknown or unseen classes are not rejected during the test. There are few studies regarding open-set in the context of remote sensing. Regarding semantic segmentation of aerial imagery, a study by (da Silva et al., 2020) presented an approach considering the open-set context. There, an adaptation of a close-set semantic segmentation method, adding a probability threshold after the softmax, was conducted. Later, a post-processing step based on morphological filters was applied to the pixels classified as unknown to verify if they are inside pixels or from borders. Another interesting approach is to combine open-set and domain adaptation methods, as proposed by (Adayel et al., 2020) in the remote sensing context.

5.9. Photogrammetric processing

Although not as developed as other practices, DL-based methods can be adopted for processing and optimizing the UAV photogrammetric processing task. This process aims to generate a dense point cloud and an orthomosaic, and it is based on Structure-from-Motion (SfM) and Multi-View Stereo (MVS) techniques. In SfM, the interior and exterior orientation parameters are estimated, and a sparse point cloud is generated. A matching technique between the images is applied in SfM. A recent survey on image matching (Ma et al., 2021) concluded that this thematic is still an open problem and pointed out the potential of DL is this task. The authors mentioned that DL techniques are mainly applied to feature detection and description, and further investigations on feature matching can be explored. Finally, they pointed out that a promising direction is the customization of modern feature matching techniques to attend SfM.

Regarding DL for UAV image matching, there is a lack of work indicating a potential for future exploration. In the UAV photogrammetric process, DL also can be used in filtering the DSM, which is essential to generate high-quality orthoimages. Previous work (Gevaert et al., 2018) showed the potential of using DL to filter the DSM and generate the DTM. Further investigations are required in this thematic, mainly considering UAV data. Besides, another task that can be beneficiated by DL is the color balancing between images when generating orthomosaic from thousands of images, corresponding to extensive areas.

6. Conclusions

DL is still considered up to the time of writing, a "black-box" type of solution for most of the problems, although novel research is advancing in minimizing this notion at considerable proportions. Regardless, in the remote sensing domain, it already provided important discoveries on most of its implementations. Our literature revision has focused on the application of these methods in UAV-based image processing. In this sense, we structured our study to offer more of a comprehensive approach to the subject while presenting an overview of state-of-the-art techniques and perspectives regarding its usage. As such, we hope that this literature revision may serve as an inclusive survey to summarize the UAV applications based on DNNs. Thus, in the evaluated context, this review concludes that:

1. In the context of UAV remote sensing, most of the published materials are based on object detection methods and RGB sensors; however, some applications, as in precision agriculture and forest-related, benefit from multi/hyperspectral data;
2. There is a need for additional labeled public available datasets obtained with UAVs to be used to train and benchmark the networks. In this context, we contributed by providing a repository with some of our UAV datasets in both agricultural and environmental applications;
3. Even though CNNs are the most adopted architecture, other methods based on CNN-LSTMs and GANs are gaining attention in UAV remote sensing and image applications, and future UAV remote sensing works may benefit from their inclusion;
4. DL, when assisted by GPU processing, can provide fast inference solutions. However there is still a need for further investigation regarding real-time processing using embedded systems on UAVs, and, lastly;
5. Some promising thematics, such as open-set, attention-based mechanisms, few shot and multitask learning can be combined and provide novel approaches in the context of UAV remote sensing; also, these thematics can contribute significantly to the generalization capacity of the DNNs.

Funding

This research was funded by CNPq (p: 433783/2018-4, 310517/2020-6, 314902/2018-0, 304052/2019-1 and 303559/2019-5), FUNDECT (p: 59/300.066/2015) and CAPES PrInt (p: 88881.311850/2018-01). The authors acknowledge the support of the UFMS (Federal University of Mato Grosso do Sul) and CAPES (Finance Code 001).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge Nvidia Corporation for the donation of the Titan X and V graphic cards.

References

Adayel, R., Bazi, Y., Alhichri, H., Alajlan, N., 2020. Deep open-set domain adaptation for cross-scene classification based on adversarial learning and pareto ranking. Remote Sens. 12, 1716. https://doi.org/10.3390/rs12111716.

Ado, T., Hruka, J., Pdua, L., Bessa, J., Peres, E., Morais, R., Sousa, J.J., 2020. Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry. Remote Sens. 12 https://doi.org/10.3390/rs9111110.

Alexandra Larsen, A., Hanigan, I., Reich, B.J., Qin, Y., Cope, M., Morgan, G., Rappold, A. G., 2020. A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication. J. Exposure Sci. Environ. Epidemiol. 31, 170–176.

Al-Najjar, H.A.H., Kalantar, B., Pradhan, B., Saeidi, V., Halin, A.A., Ueda, N., Mansor, S., 2019. Land cover classification from fused dsm and uav images using convolutional neural networks. Remote Sens. 11. https://doi.org/10.3390/rs11121461 https://www.mdpi.com/2072-4292/11/12/1461.

Ampatzidis, Y., Partel, V., 2019. UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. Remote Sens. 11 https://doi.org/10.3390/rs11040410.

Aparna, Bhatia, Y., Rai, R., Gupta, V., Aggarwal, N., Akula, A., 2019. Convolutional neural networks based potholes detection using thermal imaging. J. King Saud Univ. Comput. Inform. Sci. doi: https://doi.org/10.1016/j.jksuci.2019.02.004. URL https://www.sciencedirect.com/science/article/pii/S1319157818312837.

Apolo-Apolo, O.E., Martínez-Guanter, J., Egea, G., Raja, P., Pérez-Ruiz, M., 2020. Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. Eur. J. Agron. 115, 126030. https://doi.org/10.1016/j.eja.2020.126030.

Audebert, N., Le Saux, B., Lefevre, S., 2019. Deep learning for classification of hyperspectral data: A comparative review. IEEE Geosci. Remote Sens. Mag. 7, 159–173. https://doi.org/10.1109/MGRS.2019.2912563 arXiv:1904.10674.

Bachman, P., Hjelm, R.D., Buchwalter, W., 2019. Learning representations by maximizing mutual information across views. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc. pp. 15535–15545. https://proceedings.neurips.cc/paper/2019/file/ddf354219aac374f1d40b7e760ee5bb7-Paper.pdf.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615 arXiv:1511.00561.

Ball, J.E., Anderson, D.T., Chan, C.S., 2017. A comprehensive survey of deep learning in remote sensing: Theories, tools and challenges for the community. arXiv 11. doi: 10.1117/1.jrs.11.042609, arXiv:1709.00308.

Barbedo, J.G.A., Koenigkan, L.V., Santos, T.T., Santos, P.M., 2019. A study on the detection of cattle in UAV images using deep learning. Sensors (Switzerland) 19, 1–14. https://doi.org/10.3390/s19245436.

Barbedo, J.G.A., Koenigkan, L.V., Santos, P.M., Ribeiro, A.R.B., 2020. Counting cattle in uav images-dealing with clustered animals and animal/background contrast changes. Sensors 20. https://doi.org/10.3390/s20072126 https://www.mdpi.com/1424-8220/20/7/2126.

Bell, T.W., Nidzieko, N.J., Siegel, D.A., Miller, R.J., Cavanaugh, K.C., Nelson, N.B., Griffith, M., 2020. The utility of satellites and autonomous remote sensing platforms for monitoring offshore aquaculture farms: A case study for canopy forming kelps. Front. Mar. Sci.

Bendale, A., Boult, T.E., 2016. Towards open set deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), p. 14.

Benjdira, B., Bazi, Y., Koubaa, A., Ouni, K., 2019a. Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images. Remote Sens. 11 https://doi.org/10.3390/rs11111369 arXiv:1905.03198.

Benjdira, B., Bazi, Y., Koubaa, A., Ouni, K., 2019b. Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images. Remote Sens. 11. https://doi.org/10.3390/rs11111369 https://www.mdpi.com/2072-4292/11/11/1369.

Bhowmick, S., Nagarajaiah, S., Veeraraghavan, A., 2020. Vision and deep learning-based algorithms to detect and quantify cracks on concrete surfaces from UAV videos. Sensors (Switzerland) 20, 1–19. https://doi.org/10.3390/s20216299.

Biffi, L.J., Mitishita, E., Liesenberg, V., Dos Santos, A.A., Gonçalves, D.N., Estrabis, N.V., Silva, J.d.A., Osco, L.P., Ramos, A.P.M., Centeno, J.A.S., Schimalski, M.B., Rufato, L., Neto, S.L.R., Junior, J.M., Gonçalves, W.N., 2021. Article atss deep learning-based approach to detect apple fruits. Remote Sens. 13, 1–23. doi: 10.3390/rs13010054.

Bithas, P.S., Michailidis, E.T., Nomikos, N., Vouyioukas, D., Kanatas, A.G., 2019. A survey on machine-learning techniques for UAV-based communications. Sensors (Switzerland) 19, 1–39. https://doi.org/10.3390/s19235170.

Boonpook, W., Tan, Y., Xu, B., 2021. Deep learning-based multi-feature semantic segmentation in building extraction from images of UAV photogrammetry. Int. J. Remote Sens. 42, 1–19. https://doi.org/10.1080/01431161.2020.1788742.

Bui, D.T., Tsangaratos, P., Nguyen, V.T., Liem, N.V., Trinh, P.T., 2020. Comparing the prediction performance of a deep learning neural network model with conventional machine learning models in landslide susceptibility assessment. CATENA 188, 104426. https://doi.org/10.1016/j.catena.2019.104426 http://www.sciencedirect.com/science/article/pii/S0341816219305685.

Buscombe, D., Ritchie, A.C., 2018. Landscape classification with deep neural networks. Geosciences 8. https://doi.org/10.3390/geosciences8070244. https://www.mdpi.com/2076-3263/8/7/244.

Cai, Z., Vasconcelos, N., 2018. Cascade r-cnn: Delving into high quality object detection. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6154–6162. doi: 10.1109/CVPR.2018.00644.

Cai, Z., Vasconcelos, N., 2019. Cascade r-cnn: high quality object detection and instance segmentation. IEEE Trans. Pattern Anal. Mach. Intell.

Cao, Y., Chen, K., Loy, C.C., Lin, D., 2020. Prime sample attention in object detection. In: IEEE Conference on Computer Vision and Pattern Recognition, p. 9.

Carbonneau, P.E., Dugdale, S.J., Breckon, T.P., Dietrich, J.T., Fonstad, M.A., Miyamoto, H., Woodget, A.S., 2020. Adopting deep learning methods for airborne RGB fluvial scene classification. Remote Sens. Environ. 251 https://doi.org/10.1016/j.rse.2020.112107.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (Eds.), Computer Vision – ECCV 2020. Springer International Publishing, Cham, pp. 213–229.

Caron, M., Bojanowski, P., Joulin, A., Douze, M., 2018. Deep clustering for unsupervised learning of visual features. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision – ECCV 2018. Springer International Publishing, Cham, pp. 139–156.

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A., 2021. Unsupervised learning of visual features by contrasting cluster assignments. arXiv:2006.09882.

Castro, W., Junior, J.M., Polidoro, C., Osco, L.P., Gonçalves, W., Rodrigues, L., Santos, M., Jank, L., Barrios, S., Valle, C., Simeão, R., Carromeu, C., Silveira, E., Jorge, L.A.d. C., Matsubara, E., 2020. Deep learning applied to phenotyping of biomass in forages with uav-based rgb imagery. Sensors (Switzerland) 20, 1–18. doi: 10.3390/s20174802.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2016. Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv: 1412.7062.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40, 834–848. https://doi.org/10.1109/TPAMI.2017.2699184 arXiv:1606.00915.

Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Shi, J., Ouyang, W., Loy, C.C., Lin, D., 2019. Hybrid task cascade for instance segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, p. 10.

Chen, J., Wu, Q., Liu, D., Xu, T., 2020. Foreground-background imbalance problem in deep object detectors: A review. In: 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 285–290. https://doi.org/10.1109/MIPR49039.2020.00066.

Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. 117, 11–28. https://doi.org/10.1016/j.isprsjprs.2016.03.014, arXiv:1603.06201.

Cheng, G., Han, J., Lu, X., 2017. Remote sensing image scene classification: Benchmark and state of the art. arXiv.

Crawshaw, M., 2020. Multi-task learning with deep neural networks: A survey. arXiv: 2009.09796.

Dao, D.V., Jaafari, A., Bayat, M., Mafi-Gholami, D., Qi, C., Moayedi, H., Phong, T.V., Ly, H.B., Le, T.T., Trinh, P.T., Luu, C., Quoc, N.K., Thanh, B.N., Pham, B.T., 2020. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. CATENA 188, 104451. https://doi.org/10.1016/j.catena.2019.104451 http://www.sciencedirect.com/science/article/pii/S0341816219305934.

da Silva, C.C.V., Nogueira, K., Oliveira, H.N.d., Santos, A., 2020. Towards open-set semantic segmentation of aerial images. In: 2020 IEEE Latin American GRSS ISPRS Remote Sensing Conference (LAGIRS), pp. 16–21. https://doi.org/10.1109/LAGIRS48042.2020.9165597.

de Oliveira, D.C., Wehrmeister, M.A., 2018. Using deep learning and low-cost rgb and thermal cameras to detect pedestrians in aerial images captured by multirotor uav. Sensors (Switzerland) 18. https://doi.org/10.3390/s18072244.

Dian Bah, M., Hafiane, A., Canals, R., 2018. Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. Remote Sens. 10, 1–22. https://doi.org/10.3390/rs10111690.

Ding, L., Tang, H., Bruzzone, L., 2021. Lanet: Local attention embedding to improve the semantic segmentation of remote sensing images. IEEE Trans. Geosci. Remote Sens. 59, 426–435. https://doi.org/10.1109/TGRS.2020.2994150.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv: 2010.11929.

dos Santos, A.A., Marcato Junior, J., Araújo, M.S., Di Martini, D.R., Tetila, E.C., Siqueira, H.L., Aoki, C., Eltner, A., Matsubara, E.T., Pistori, H., Feitosa, R.Q., Liesenberg, V., Gonçalves, W.N., 2019. Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVS. Sensors (Switzerland) 19, 1–11. https://doi.org/10.3390/s19163595.

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., Tian, Q., 2019. CenterNet: Keypoint triplets for object detection. In: Proceedings of the IEEE International Conference on Computer Vision 2019-October. https://doi.org/10.1109/ICCV.2019.00667 arXiv: 1904.08189.

Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., Zhang, W., Huang, Q., Tian, Q., 2018. The unmanned aerial vehicle benchmark: Object detection and tracking. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11214 LNCS, 375–391. doi: 10.1007/978-3-030-01249-6_23.

Elshamli, A., Taylor, G.W., Berg, A., Areibi, S., 2017. Domain adaptation using representation learning for the classification of remote sensing images. IEEE J. Sel.

Top. Appl. Earth Observ. Remote Sens. 10, 4198–4209. https://doi.org/10.1109/JSTARS.2017.2711360.
Fang, B., Kou, R., Pan, L., Chen, P., 2019. Category-sensitive domain adaptation for land cover mapping in aerial scenes. Remote Sens. 11 https://doi.org/10.3390/rs11222631 https://www.mdpi.com/2072-4292/11/22/2631.
Feng, Q., Yang, J., Liu, Y., Ou, C., Zhu, D., Niu, B., Liu, J., Li, B., 2020. Multi-temporal unmanned aerial vehicle remote sensing for vegetable mapping using an attention-based recurrent convolutional neural network. Remote Sens. 12 https://doi.org/10.3390/rs12101668.
Ferreira, M.P., de Almeida, D.R.A., Papa, D.d.A., Minervino, J.B.S., Veras, H.F.P., Formighieri, A., Santos, C.A.N., Ferreira, M.A.D., Figueiredo, E.O., Ferreira, E.J.L., 2020. Individual tree detection and species classification of Amazonian palms using UAV images and deep learning. Forest Ecol. Manage. 475, 118397. URL https://doi.org/10.1016/j.foreco.2020.118397, doi: 10.1016/j.foreco.2020.118397.
Gao, S.H., Cheng, M.M., Zhao, K., Zhang, X.Y., Yang, M.H., Torr, P., 2021. Res2net: A new multi-scale backbone architecture. IEEE Trans. Pattern Anal. Mach. Intell. 43, 652–662. https://doi.org/10.1109/TPAMI.2019.2938758.
Gebrehiwot, A., Hashemi-Beni, L., Thompson, G., Kordjamshidi, P., Langan, T.E., 2019. Deep convolutional neural network for flood extent mapping using unmanned aerial vehicles data. Sensors 19. https://doi.org/10.3390/s19071486 https://www.mdpi.com/1424-8220/19/7/1486.
Gevaert, C., Persello, C., Nex, F., Vosselman, G., 2018. A deep learning approach to dtm extraction from imagery using rule-based training labels. ISPRS J. Photogramm. Remote Sens. 142, 106–123. https://doi.org/10.1016/j.isprsjprs.2018.06.001 http://www.sciencedirect.com/science/article/pii/S0924271618301643.
Gevaert, C.M., Persello, C., Sliuzas, R., Vosselman, G., 2020. Monitoring household upgrading in unplanned settlements with unmanned aerial vehicles. Int. J. Appl. Earth Obs. Geoinf. 90, 102117. https://doi.org/10.1016/j.jag.2020.102117.
Ghiasi, G., Lin, T.Y., Le, Q.V., 2019. Nas-fpn: Learning scalable feature pyramid architecture for object detection, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7036–7045.
Giang, T.L., Dang, K.B., Toan Le, Q., Nguyen, V.G., Tong, S.S., Pham, V.M., 2020. U-net convolutional networks for mining land cover classification based on high-resolution uav imagery. IEEE Access 8, 186257–186273. https://doi.org/10.1109/ACCESS.2020.3030112.
Gomes, M., Silva, J., Gonçalves, D., Zamboni, P., Perez, J., Batista, E., Ramos, A., Osco, L., Matsubara, E., Li, J., Junior, J.M., Gonçalves, W., 2020. Mapping utility poles in aerial orthoimages using atss deep learning method. Sensors (Switzerland) 20, 1–14. https://doi.org/10.3390/s20216070.
Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. arXiv:1406.2661.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
Gray, P.C., Bierlich, K.C., Mantell, S.A., Friedlaender, A.S., Goldbogen, J.A., Johnston, D. W., 2019. Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. Methods Ecol. Evol. 10, 1490–1500. https://doi.org/10.1111/2041-210X.13246.
Guo, Y., 2018. A survey on methods and theories of quantized neural networks. arXiv preprint arXiv:1808.04752.
Hamdi, Z.M., Brandmeier, M., Straub, C., 2019. Forest damage assessment using deep learning on high resolution remote sensing data. Remote Sens. 11, 1–14. https://doi.org/10.3390/rs11171976.
Hamylton, S., Morris, R., Carvalho, R., Roder, N., Barlow, P., Mills, K., Wang, L., 2020. Evaluating techniques for mapping island vegetation from unmanned aerial vehicle (UAV) images: Pixel classification, visual interpretation and machine learning approaches. Int. J. Appl. Earth Obs. Geoinf. 89, 102085. https://doi.org/10.1016/j.jag.2020.102085.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-December, 770–778. https://doi.org/10.1109/CVPR.2016.90 arXiv:1512.03385.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
He, K., Gkioxari, G., Dollr, P., Girshick, R., 2017. Mask r-cnn. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.322.
He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9726–9735. https://doi.org/10.1109/CVPR42600.2020.00975.
Hennessy, A., Clarke, K., Lewis, M., 2020. Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability. Remote Sens. 12, 113. https://doi.org/10.3390/rs12010113.
Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.
Hjelm, D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., Bengio, Y., 2019. Learning deep representations by mutual information estimation and maximization. In: ICLR 2019, ICLR. p. 24.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 https://doi.org/10.1162/neco.1997.9.8.1735.
Horning, N., Fleishman, E., Ersts, P.J., Fogarty, F.A., Wohlfeil Zillig, M., 2020. Mapping of land cover with open-source software and ultra-high-resolution imagery acquired with unmanned aerial vehicles. Remote Sens. Ecol. Conserv. 6, 487–497. https://doi.org/10.1002/rse2.144.
Hossain, M.D., Chen, D., 2019. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. ISPRS J. Photogramm. Remote Sens. 150, 115–134. https://doi.org/10.1016/j.isprsjprs.2019.02.009.
Ho Tong Minh, D., Ienco, D., Gaetano, R., Lalande, N., Ndikumana, E., Osman, F., Maurel, P., 2018. Deep recurrent neural networks for winter vegetation quality mapping via multitemporal sar sentinel-1. IEEE Geosci. Remote Sens. Lett. 15, 464–468. doi: 10.1109/LGRS.2018.2794581.
Hou, J., He, Y., Yang, H., Connor, T., Gao, J., Wang, Y., Zeng, Y., Zhang, J., Huang, J., Zheng, B., Zhou, S., 2020. Identification of animal individuals using deep learning: A case study of giant panda. Biol. Conserv. 242, 108414. https://doi.org/10.1016/j.biocon.2020.108414 http://www.sciencedirect.com/science/article/pii/S000632071931609X.
Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
Hua, Y., Marcos, D., Mou, L., Zhu, X.X., Tuia, D., 2021. Semantic segmentation of remote sensing images with sparse annotations. IEEE Geosci. Remote Sens. Lett.
Hu, G., Yin, C., Wan, M., Zhang, Y., Fang, Y., 2020. Recognition of diseased Pinus trees in UAV images using deep learning and AdaBoost classifier. Biosyst. Eng. 194, 138–151. https://doi.org/10.1016/j.biosystemseng.2020.03.021.
Ichim, L., Popescu, D., 2020. Segmentation of vegetation and flood from aerial images based on decision fusion of neural networks. Remote Sens. 12 https://doi.org/10.3390/rs12152490 https://www.mdpi.com/2072-4292/12/15/2490.
Ienco, D., Gaetano, R., Dupaquier, C., Maurel, P., 2017. Land cover classification via multitemporal spatial data by deep recurrent neural networks. IEEE Geosci. Remote Sens. Lett. 14, 1685–1689. https://doi.org/10.1109/LGRS.2017.2728698.
ImageNet, 2018. Imagenet object localization challenge. https://www.kaggle.com/c/imagenet-object-localization-challenge.
Imran, H.A., Mujahid, U., Wazir, S., Latif, U., Mehmood, K., 2020. Embedded development boards for edge-ai: A comprehensive report arXiv preprint arXiv:2009.00803.
Isola, P., Zhu, J.Y., Zhou, T., Efros, A., 2018. Image-to-image translation with conditional adversarial networks.
Jakovljevic, G., Govedarica, M., Alvarez-Taboada, F., Pajic, V., 2019. Accuracy assessment of deep learning based classification of lidar and uav points clouds for dtm creation and flood risk mapping. Geosciences 9. https://doi.org/10.3390/geosciences9070323 https://www.mdpi.com/2076-3263/9/7/323.
Jia, S., Jiang, S., Lin, Z., Li, N., Xu, M., Yu, S., 2021. A survey: Deep learning for hyperspectral image classification with few labeled samples. Neurocomputing 448, 179–204. https://doi.org/10.1016/j.neucom.2021.03.035 https://www.sciencedirect.com/science/article/pii/S0925231221004033.
Kang, H., Chen, C., 2020. Fast implementation of real-time fruit detection in apple orchards using deep learning. Comput. Electron. Agric. 168, 105108. https://doi.org/10.1016/j.compag.2019.105108.
Kang, J., Fernandez-Beltran, R., Duan, P., Liu, S., Plaza, A.J., 2020. Deep unsupervised embedding for remotely sensed images based on spatially augmented momentum contrast. IEEE Trans. Geosci. Remote Sens. 1–13 https://doi.org/10.1109/TGRS.2020.3007029.
Kannojia, S.P., Jaiswal, G., 2018. Effects of Varying Resolution on Performance of CNN based Image Classification An Experimental Study. Int. J. Comput. Sci. Eng. 6, 451–456. https://doi.org/10.26438/ijcse/v6i9.451456.
Karami, A., Crawford, M., Delp, E.J., 2020. Automatic plant counting and location based on a few-shot learning technique. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 13, 5872–5886. https://doi.org/10.1109/JSTARS.2020.3025790.
Kellenberger, B., Marcos, D., Tuia, D., 2018. Detecting mammals in uav images: Best practices to address a substantially imbalanced dataset with deep learning. Remote Sens. Environ. 216, 139–153. https://doi.org/10.1016/j.rse.2018.06.028 http://www.sciencedirect.com/science/article/pii/S0034425718303067.
Kerkech, M., Hafiane, A., Canals, R., 2020. Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach. Comput. Electron. Agric. 174 https://doi.org/10.1016/j.compag.2020.105446.
Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S., 2020. A survey of the recent architectures of deep convolutional neural networks, vol. 53. Springer, Netherlands. https://doi.org/10.1007/s10462-020-09825-6 arXiv:1901.06032.
Khelifi, L., Mignotte, M., 2020. Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review and Meta-Analysis. IEEE Access 8, 126385–126400. https://doi.org/10.1109/ACCESS.2020.3008036 arXiv:2006.05612.
Kim, K., Lee, H.S., 2020. Probabilistic anchor assignment with iou prediction for object detection. In: European Conference on Computer Vision (ECCV), p. 22.
Kirillov, A., He, K., Girshick, R., Rother, C., Dollr, P., 2019. Panoptic segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9396–9405. https://doi.org/10.1109/CVPR.2019.00963.
Kirillov, A., Wu, Y., He, K., Girshick, R., 2020. Pointrend: Image segmentation as rendering, in. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), p. 10.
Kitano, B.T., Mendes, C.C.T., Geus, A.R., Oliveira, H.C., Souza, J.R., 2019. Corn Plant Counting Using Deep Learning and UAV Images. IEEE Geosci. Remote Sens. Lett. 1–5 https://doi.org/10.1109/lgrs.2019.2930549.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. Curran Associates Inc., Red Hook, NY, USA, pp. 1097–1105.
Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14, 778–782. https://doi.org/10.1109/LGRS.2017.2681128.

Lathuilire, S., Mesejo, P., Alameda-Pineda, X., Horaud, R., 2020. A comprehensive analysis of deep regression. IEEE Trans. Pattern Anal. Mach. Intell. 42, 2065–2081. https://doi.org/10.1109/TPAMI.2019.2910523.

Law, H., Deng, J., 2020. CornerNet: Detecting Objects as Paired Keypoints. Int. J. Comput. Vision 128, 642–656. https://doi.org/10.1007/s11263-019-01204-1 arXiv:1808.01244.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Licciardi, G., Marpu, P.R., Chanussot, J., Benediktsson, J.A., 2012. Linear versus nonlinear pca for the classification of hyperspectral data based on the extended morphological profiles. IEEE Geosci. Remote Sens. Lett. 9, 447–451. https://doi.org/10.1109/LGRS.2011.2172185.

Li, Y., Zhang, H., Xue, X., Jiang, Y., Shen, Q., 2018. Deep learning for remote sensing image classification: A survey. Wiley Interdiscipl. Rev. Data Min. Knowl. Discov. 8, 1–17. https://doi.org/10.1002/widm.1264.

Li, C., Xu, C., Cui, Z., Wang, D., Zhang, T., Yang, J., 2019. Feature-attentioned object detection in remote sensing imagery. In: 2019 IEEE International Conference on Image Processing (ICIP), pp. 3886–3890. https://doi.org/10.1109/ICIP.2019.8803521.

Li, Y., Chen, Y., Wang, N., Zhang, Z., 2019. Scale-aware trident networks for object detection. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6053–6062. https://doi.org/10.1109/ICCV.2019.00615.

Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., Benediktsson, J.A., 2019a. Deep learning for hyperspectral image classification: An overview. IEEE Trans. Geosci. Remote Sens. 57, 6690–6709. https://doi.org/10.1109/TGRS.2019.2907932 arXiv:1910.12861.

Li, Y., Peng, B., He, L., Fan, K., Li, Z., Tong, L., 2019b. Road extraction from unmanned aerial vehicle remote sensing images based on improved neural networks. Sensors (Switzerland) 19. https://doi.org/10.3390/s19194115.

Li, L., Han, J., Yao, X., Cheng, G., Guo, L., 2020. Dla-matchnet for few-shot remote sensing image scene classification. IEEE Trans. Geosci. Remote Sens. 1–10 https://doi.org/10.1109/TGRS.2020.3033336.

Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J., Yang, J., 2020a. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. arXiv preprint arXiv:2006.04388.

Li, Y., Cao, Z., Lu, H., Xu, W., 2020b. Unsupervised domain adaptation for in-field cotton boll status identification. Comput. Electron. Agric. 178, 105745. https://doi.org/10.1016/j.compag.2020.105745 http://www.sciencedirect.com/science/article/pii/S0168169920306517.

Li, Y., Huang, Q., Pei, X., Jiao, L., Shang, R., 2020c. Radet: Refine feature pyramid network and multi-layer attention network for arbitrary-oriented object detection of remote sensing images. Remote Sens. 12 https://doi.org/10.3390/rs12030389 https://www.mdpi.com/2072-4292/12/3/389.

Lin, D., Fu, K., Wang, Y., Xu, G., Sun, X., 2017a. Marta gans: Unsupervised representation learning for remote sensing image classification. IEEE Geosci. Remote Sens. Lett. 14, 2092–2096. https://doi.org/10.1109/LGRS.2017.2752750.

Lin, T., Dollr, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017b. Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944. https://doi.org/10.1109/CVPR.2017.106.

Liu, W., Qin, R., 2020. A multikernel domain adaptation method for unsupervised transfer learning on cross-source and cross-region remote sensing data classification. IEEE Trans. Geosci. Remote Sens. 58, 4279–4289. https://doi.org/10.1109/TGRS.2019.2962039.

Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., 2018. Path aggregation network for instance segmentation, in. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), p. 11.

Liu, L., Ouyang, W., Wang, X., Fieguth, W.P., Chen, J., Liu, X., Pietikinen, M., 2019. Deep learning for generic object detection: A survey. Int. J. Comput. Vision 261–318.

Lu, X., Li, B., Yue, Y., Li, Q., Yan, J., 2019. Grid R-CNN plus: Faster and better. CoRR abs/1906.05688. http://arxiv.org/abs/1906.05688, arXiv:1906.05688.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015 http://www.sciencedirect.com/science/article/pii/S0924271619301108.

Ma, J., Jiang, X., Fan, A., Jiang, J., Yan, J., 2021. Image matching from handcrafted to deep features: A survey. Int. J. Comput. Vision 129, 23–79. https://doi.org/10.1007/s11263-020-01359-2.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2020. Image segmentation using deep learning: A survey arXiv:2001.05566.

Mittal, S., 2019. A survey on optimized implementation of deep learning models on the nvidia jetson platform. J. Syst. Architect. 97, 428–442.

Miyoshi, G.T., Arruda, M.d.S., Osco, L.P., Marcato Junior, J., Gonalves, D.N., Imai, N.N., Tommaselli, A.M.G., Honkavaara, E., Gonalves, W.N., 2020. A novel deep learning method to identify single tree species in uav-based hyperspectral images. Remote Sens. 12. doi: 10.3390/rs12081294. URL https://www.mdpi.com/2072-4292/12/8/1294.

Nevavuori, P., Narra, N., Linna, P., Lipping, T., 2020. Crop yield prediction using multitemporal UAV data and spatio-temporal deep learning models. Remote Sens. 12, 1–18. https://doi.org/10.3390/rs12234000.

Nezami, S., Khoramshahi, E., Nevalainen, O., Plnen, I., Honkavaara, E., 2020. ree species classification of drone hyperspectral and rgb imagery with deep learning convolutional neural networks. Remote Sens. 12 https://doi.org/10.3390/rs12071070.

Nogueira, K., Dalla Mura, M., Chanussot, J., Schwartz, W.R., Dos Santos, J.A., 2019. Dynamic multicontext segmentation of remote sensing images based on convolutional networks. IEEE Trans. Geosci. Remote Sens. 57, 7503–7520. https://doi.org/10.1109/TGRS.2019.2913861 arXiv:1804.04020.

Nogueira, K., Machado, G.L., Gama, P.H., da Silva, C.C., Balaniuk, R., dos Santos, J.A., 2020. Facing erosion identification in railway lines using pixel-wise deep-based approaches. Remote Sens. 12, 1–21. https://doi.org/10.3390/rs12040739.

Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S., 2018. Activation functions: Comparison of trends in practice and research for deep learning. arXiv preprint arXiv:1811.03378.

Osco, L.P., dos Santos de Arruda, M., Gonalves, D.N., Dias, A., Batistoti, J., de Souza, M., Gomes, F.D.G., Ramos, A.P.M., de Castro Jorge, L.A., Liesenberg, V., Li, J., Ma, L., Junior, J.M., Gonalves, W.N., 2020a. A cnn approach to simultaneously count plants and detect plantation-rows from uav imagery. arXiv:2012.15827.

Osco, L.P., de Arruda, M.d.S., Marcato Junior, J., da Silva, N.B., Ramos, A.P.M., Moryia, É.A.S., Imai, N.N., Pereira, D.R., Creste, J.E., Matsubara, E.T., Li, J., Gonçalves, W.N., 2020b. A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing 160, 97–106. URL https://doi.org/10.1016/j.isprsjprs.2019.12.010, doi:10.1016/j.isprsjprs.2019.12.010.

Osco, L.P., Nogueira, K., Marques Ramos, A.P., Faita Pinheiro, M.M., Furuya, D.E.G., Gonalves, W.N., de Castro Jorge, L.A., Marcato Junior, J., dos Santos, J.A., 2021. Semantic segmentation of citrus-orchard using deep neural networks and multispectral uav-based imagery. Precision Agric. https://doi.org/10.1007/s11119-020-09777-5.

Pang, J., Chen, K., Shi, J., Feng, H., Ouyang, W., Lin, D., 2019. Libra R-CNN: Towards balanced learning for object detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2019-June, 821–830. https://doi.org/10.1109/CVPR.2019.00091 arXiv:1904.02701.

Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A., 2019. Deep learning classifiers for hyperspectral imaging: A review. ISPRS J. Photogramm. Remote Sens. 158, 279–317. https://doi.org/10.1016/j.isprsjprs.2019.09.006.

Park, S., Song, A., 2020. Discrepancy analysis for detecting candidate parcels requiring update of land category in cadastral map using hyperspectral uav images: A case study in jeonju, south korea. Remote Sens. 12 https://doi.org/10.3390/rs12030354 https://www.mdpi.com/2072-4292/12/3/354.

Penatti, O.A., Nogueira, K., Dos Santos, J.A., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2015-October, pp. 44–51. doi:10.1109/CVPRW.2015.7301382.

Petersson, H., Gustafsson, D., Bergström, D., 2017. Hyperspectral image analysis using deep learning - A review. In: 2016 6th International Conference on Image Processing Theory, Tools and Applications, IPTA, 10.1109/IPTA.2016.7820963.

Qiao, S., Chen, L.C., Yuille, A., 2020. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. arXiv preprint arXiv:2006.02334.

Qin, Z., Li, Z., Zhang, Z., Bao, Y., Yu, G., Peng, Y., Sun, J., 2019. Thundernet: Towards real-time generic object detection on mobile devices. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 6718–6727.

Radosavovic, I., Kosaraju, R., Girshick, R., He, K., Dollar, P., 2020. Designing network design spaces. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, pp. 10425–10433.

Rivas, A., Chamoso, P., González-Briones, A., Corchado, J.M., 2018. Detection of cattle using drones and convolutional neural networks. Sensors (Switzerland) 18, 1–15. https://doi.org/10.3390/s18072048.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9351, 234–241. doi:10.1007/978-3-319-24574-4_28, arXiv:1505.04597.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Netw. 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003 http://www.sciencedirect.com/science/article/pii/S0893608014002135.

Sharma, V., Mir, R.N., 2020. A comprehensive and systematic look up into deep learning based object detection techniques: A review. Comput. Sci. Rev. 38, 100301. https://doi.org/10.1016/j.cosrev.2020.100301.

Sheng, G., Yang, W., Xu, T., Sun, H., 2012. High-resolution satellite scene classification using a sparse coding based multiple feature combination. Int. J. Remote Sens. 33, 2395–2412. https://doi.org/10.1080/01431161.2011.608740.

Signoroni, A., Savardi, M., Baronio, A., Benini, S., 2019. Deep learning meets hyperspectral image analysis: A multidisciplinary review. J. Imag. 5 https://doi.org/10.3390/jimaging5050052 https://www.mdpi.com/2313-433X/5/5/52.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations, p. 14.

Soderholm, J.S., Kumjian, M.R., McCarthy, N., Maldonado, P., Wang, M., 2020. Quantifying hail size distributions from the sky – application of drone aerial photogrammetry. Atmospheric. Meas. Tech. 13, 747–754. https://doi.org/10.5194/amt-13-747-2020 https://amt.copernicus.org/articles/13/747/2020/.

Su, Y., Wu, Y., Wang, M., Wang, F., Cheng, J., 2019. Semantic segmentation of high resolution remote sensing image based on batch-attention mechanism. In: IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 3856–3859. https://doi.org/10.1109/IGARSS.2019.8898198.

Sundaram, D.M., Loganathan, A., 2020. FSSCaps-DetCountNet: fuzzy soft sets and CapsNet-based detection and counting network for monitoring animals from aerial images. J. Appl. Remote Sens. 14, 1–30. https://doi.org/10.1117/1.JRS.14.026521.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A survey on deep transfer learning. In: International conference on artificial neural networks. Springer, pp. 270–279.

Tetila, E.C., Machado, B.B., Menezes, G.K., Da Silva Oliveira, A., Alvarez, M., Amorim, W.P., De Souza Belete, N.A., Da Silva, G.G., Pistori, H., 2020. Automatic

Recognition of Soybean Leaf Diseases Using UAV Images and Deep Convolutional Neural Networks. IEEE Geosci. Remote Sens. Lett. 17, 903–907. https://doi.org/10.1109/LGRS.2019.2932385.
Thoma, M., 2016. A survey of semantic segmentation. arXiv:1602.06541.
Tian, Y., Krishnan, D., Isola, P., 2019a. Contrastive multiview coding. CoRR abs/1906.05849. http://arxiv.org/abs/1906.05849, arXiv:1906.05849.
Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., Liang, Z., 2019b. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Comput. Electron. Agric. 157, 417–426. https://doi.org/10.1016/j.compag.2019.01.012.
Torres, D.L., Feitosa, R.Q., Happ, P.N., La Rosa, L.E.C., Junior, J.M., Martins, J., Bressan, P.O., Gonçalves, W.N., Liesenberg, V., 2020. Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery. Sensors (Switzerland) 20, 1–20. https://doi.org/10.3390/s20020563.
Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jgou, H., 2020. Training data-efficient image transformers & distillation through attention arXiv:2012.12877.
Tsagkatakis, G., Aidini, A., Fotiadou, K., Giannopoulos, M., Pentari, A., Tsakalides, P., 2019. Survey of deep-learning approaches for remote sensing observation enhancement. Sensors (Switzerland) 19, 1–39. https://doi.org/10.3390/s19183929.
Tuia, D., Persello, C., Bruzzone, L., 2016. Domain adaptation for the classification of remote sensing data: An overview of recent advances. IEEE Geosci. Remote Sens. Mag. 4, 41–57. https://doi.org/10.1109/MGRS.2016.2548504.
Vaddi, R., Manoharan, P., 2020. Cnn based hyperspectral image classification using unsupervised band selection and structure-preserving spatial features. Infrared Phys. Technol. 110, 103457. https://doi.org/10.1016/j.infrared.2020.103457 http://www.sciencedirect.com/science/article/pii/S1350449520305053.
Wang, J., Chen, K., Yang, S., Loy, C.C., Lin, D., 2019. Region proposal by guided anchoring. In: IEEE Conference on Computer Vision and Pattern Recognition, p. 12.
Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., Xiao, B., 2020. Deep high-resolution representation learning for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 1–1 https://doi.org/10.1109/TPAMI.2020.2983686.
Wang, J., Zhang, W., Cao, Y., Chen, K., Pang, J., Gong, T., Shi, J., Loy, C.C., Lin, D., 2020a. Side-aware boundary localization for more precise object detection. In: European Conference on Computer Vision (ECCV), p. 21.
Wang, S., Zhou, J., Lei, T., Wu, H., Zhang, X., Ma, J., Zhong, H., 2020b. Estimating land surface temperature from satellite passive microwave observations with the traditional neural network, deep belief network, and convolutional neural network. Remote Sens. 12 https://doi.org/10.3390/RS12172691.
Wang, Y., Ding, W., Zhang, R., Li, H., 2021. Boundary-aware multitask learning for remote sensing imagery. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 14, 951–963. https://doi.org/10.1109/JSTARS.2020.3043442.
Wu, T., Tang, S., Zhang, R., Cao, J., Zhang, Y., 2020a. Cgnet: A light-weight context guided network for semantic segmentation. IEEE Trans. Image Process. 30, 1169–1179.
Wu, X., Sahoo, D., Hoi, S.C., 2020b. Recent advances in deep learning for object detection. Neurocomputing 396, 39–64. https://doi.org/10.1016/j.neucom.2020.01.085.
Xavier Prochaska, J., Cornillon, P.C., Reiman, D.M., 2021. Deep learning of sea surface temperature patterns to identify ocean extremes. Remote Sens. 13, 1–18. https://doi.org/10.3390/rs13040744.
Xia, G.S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 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, pp. 3974–3983. https://doi.org/10.1109/CVPR.2018.00418 arXiv:1711.10398.
Xie, S., Girshick, R., Dollr, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5987–5995. https://doi.org/10.1109/CVPR.2017.634.
Xu, R., Tao, Y., Lu, Z., Zhong, Y., 2018. Attention-mechanism-containing neural networks for high-resolution remote sensing image classification. Remote Sens. 10 https://doi.org/10.3390/rs10101602 https://www.mdpi.com/2072-4292/10/10/1602.
Yao, C., Luo, X., Zhao, Y., Zeng, W., Chen, X., 2018. A review on image classification of remote sensing using deep learning. In: 2017 3rd IEEE International Conference on Computer and Communications, ICCC 2017 2018-Janua, 1947–1955. https://doi.org/10.1109/CompComm.2017.8322878.
Yin, M., Yao, Z., Cao, Y., Li, X., Zhang, Z., Lin, S., Hu, H., 2020. Disentangled non-local neural networks.
Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716. https://doi.org/10.1016/j.rse.2020.111716.
Yuan, X., Shi, J., Gu, L., 2021. A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Syst. Appl. 169, 114417. https://doi.org/10.1016/j.eswa.2020.114417.
Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. 4, 22–40. https://doi.org/10.1109/MGRS.2016.2540798.
Zhang, G., Wang, M., Liu, K., 2019a. Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China. Int. J. Disaster Risk Sci. 10, 386–403. https://doi.org/10.1007/s13753-019-00233-1.
Zhang, H., Liptrott, M., Bessis, N., Cheng, J., 2019b. Real-time traffic analysis using deep learning techniques and UAV based video. 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2019, 1–5. doi:10.1109/AVSS.2019.8909879.
Zhang, S., Chi, C., Yao, Y., Lei, Z., Li, S.Z., 2019c. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. arXiv preprint arXiv:1912.02424.
Zhang, C., Atkinson, P.M., George, C., Wen, Z., Diazgranados, M., Gerard, F., 2020a. Identifying and mapping individual plants in a highly diverse high-elevation ecosystem using UAV imagery and deep learning. ISPRS J. Photogramm. Remote Sens. 169, 280–291. https://doi.org/10.1016/j.isprsjprs.2020.09.025.
Zhang, H., Chang, H., Ma, B., Wang, N., Chen, X., 2020b. Dynamic R-CNN: Towards high quality object detection via dynamic training. arXiv preprint arXiv:2004.06002.
Zhang, H., Wang, Y., Dayoub, F., Sünderhauf, N., 2020c. Varifocalnet: An iou-aware dense object detector arXiv preprint arXiv:2008.13367.
Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., Li, M., Smola, A., 2020d. Resnest: Split-attention networks. arXiv:2004.08955.
Zhang, X., Han, L., Han, L., Zhu, L., 2020e. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? Remote Sensing 12. https://www.mdpi.com/2072-4292/12/3/417, doi:10.3390/rs12030417.
Zhang, X., Jin, J., Lan, Z., Li, C., Fan, M., Wang, Y., Yu, X., Zhang, Y., 2020f. ICENET: A semantic segmentation deep network for river ice by fusing positional and channel-wise attentive features. Remote Sens. 12, 1–22. https://doi.org/10.3390/rs12020221.
Zhao, L., Tang, P., Huo, L., 2016. Feature significance-based multibag-of-visual-words model for remote sensing image scene classification. J. Appl. Remote Sens. 10, 1–21. https://doi.org/10.1117/1.JRS.10.035004.
Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. arXiv:1612.01105.
Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X., 2019. Object detection with deep learning: A review. IEEE Trans. Neural Netw. Learn. Syst. 30, 3212–3232. https://doi.org/10.1109/tnnls.2018.2876865.
Zheng, Z., Lei, L., Sun, H., Kuang, G., 2020. A Review of Remote Sensing Image Object Detection Algorithms Based on Deep Learning. In: 2020 IEEE 5th International Conference on Image, Vision and Computing, ICIVC 2020, 34–43. https://doi.org/10.1109/ICIVC50857.2020.9177453.
Zhou, D., Wang, G., He, G., Long, T., Yin, R., Zhang, Z., Chen, S., Luo, B., 2020. Robust building extraction for high spatial resolution remote sensing images with self-attention network. Sensors 20. https://doi.org/10.3390/s20247241 https://www.mdpi.com/1424-8220/20/24/7241.
Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q., 2020. A comprehensive survey on transfer learning. Proc. IEEE 109, 43–76.
Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36. https://doi.org/10.1109/MGRS.2017.2762307.
Zhu, P., Wen, L., Du, D., Bian, X., Ling, H., Hu, Q., Nie, Q., Cheng, H., Liu, C., Liu, X., Ma, W., Wu, H., Wang, L., Schumann, A., Brown, C., Lagani, R., 2019. VisDrone-DET2018: The Vision Meets Drone Object Detection in Image Challenge Results, vol. 1. Springer, Cham. doi: 10.1007/978-3-030-11021-5.
Zhu, C., He, Y., Savvides, M., 2019a. Feature selective anchor-free module for single-shot object detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2019-June, 840–849. https://doi.org/10.1109/CVPR.2019.00093 arXiv:1903.00621.
Zhu, R., Yan, L., Mo, N., Liu, Y., 2019b. Attention-based deep feature fusion for the scene classification of high-resolution remote sensing images. Remote Sens. 11. https://doi.org/10.3390/rs11171996 https://www.mdpi.com/2072-4292/11/17/1996.
Zou, Q., Ni, L., Zhang, T., Wang, Q., 2015. Deep learning based feature selection for remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. 12, 2321–2325. https://doi.org/10.1109/LGRS.2015.2475299.
Zou, Q., Ni, L., Zhang, T., Wang, Q., 2015. Remote Sensing Scene Classification. IEEE Trans. Geosci. Remote Sens. Lett. 12, 2321–2325.