Comparative Analysis of Modern Methods for Surface Type Identification in RGB Image Data

Kateřina Příhodová
Faculty of Economics and Administration
University of Pardubice
Pardubice, Czech Republic
katerina.prihodova@upce.cz

Jakub Jech
Faculty of Economics and Administration
University of Pardubice
Pardubice, Czech Republic
jakub.jech@upce.cz

***Abstract*— The rapid development of drone technology and deep learning algorithms also expands the possibilities of environmental monitoring, e.g., the search and management of water bodies. This study aims to harness these advances for the accurate identification of surface types, with a specific focus on water bodies, in RGB image data. Using a dataset comprised of aerial images captured over the Baroch Pond within a nature reserve in the Czech Republic, this study comparative to evaluate the performance of deep learning models, including U-Net, Pyramid Scene Parsing Network (PSPNet), and DeepLabV3, in classifying surface types. The classification accuracy is slightly over 90% for most deep learning algorithms. These results show the potential of deep learning in this area. And this is key for a number of interested parties, for example, state administration, water resource managers, farmers, and tourism industry.**

Keywords—Surface Type Identification, RGB Image Data, Image Processing Techniques, Deep Learning Algorithms, Pattern Recognition, Very High Spatial Resolution

I. Introduction

The motivation for this research lies in the need for accurate identification of surface types, specifically water bodies, in RGB image data. With the advent of drone technology, the acquisition of aerial images of landscapes, including ponds and lakes, has become increasingly accessible and cost-effective. This advancement presents a unique opportunity to monitor aquatic environments, manage water resources, and support ecological studies with unprecedented accuracy and scale. The ability to accurately distinguish water surfaces from other types of land cover in RGB images captured by drones is essential for applications ranging from flood monitoring and water quality assessment to habitat mapping and conservation efforts. The aim of this article is to identify and compare various deep learning algorithms that are capable of effectively recognising water bodies under various conditions, thereby significantly contributing to environmental monitoring and management strategies.

II. Theoretical Background and Related Work

This section contains the theoretical basis of approaches and review of related work that have been used to identify surface types using RGB image data obtained by an unmanned aerial vehicle (UAV).

A. RGB Image Data

The RGB image data represents the cornerstone of digital imaging, encapsulating visual information through a combination of these three primary colours. This format is particularly relevant for environmental monitoring, as it closely mirrors human visual perception, allowing intuitive analysis of landscapes and water bodies. The peculiarity of RGB data, such as its sensitivity to lighting conditions, poses unique challenges for automatic surface type identification.

B. Mosaicking

For comprehensive surface analysis, individual images captured by the drone are stitched together to form a larger mosaic. This process involves aligning and blending overlapping images, correcting for discrepancies in exposure and colour, and producing a seamless composite image that covers the entire area of interest (AOI). Mosaic generation is crucial to creating a contiguous dataset that accurately represents the geography and types of surface of the study area.

Deep Learning Algorithms

Deep learning, a subset of machine learning, has revolutionised the field of image analysis, offering powerful tools for feature extraction and classification. Convolutional neural networks (CNNs), in particular, have shown exceptional proficiency in recognising complex patterns within images, including the differentiation of surface types [1]. However, CNNs have a far greater variety of uses, e.g., face recognition [2], handwritten Chinese text recognition [3], emotion and speech from gestures and facial expressions [4].

C. Remote sensing

Remote sensing can be defined as the intersection of science and art, the acquisition of information about objects, areas, and phenomena through equipment that is not in contact with them [5]. These devices can be various sensors that collect data. Modern devices for remote sensing are UAVs, which are very well known as drone. The UAV is a small remotely piloted aircraft that provides obtaining local data with very high spatial resolution. UAV flight in low altitude level, then is no need for geometric corrections.

D. Review of Related Work

Identification of surface types, particularly water bodies, within RGB image data is a field that has seen significant interest and development over recent years. This section reviews the relevant literature that has contributed to advancements in image processing techniques, deep learning algorithms, and their application to environmental monitoring and management. The review is structured to highlight key methodologies, compare approaches, and discuss their respective outcomes in the context of surface-type identification.

1) Early Image Processing Techniques

Initial studies in surface-type identification relied heavily on traditional image processing techniques. Methods such as thresholding, colour space transformation, spectral enhancement, and edge detection were commonly used to differentiate between water and land surfaces. For example, by authors [6] was supervised classification one of the most commonly used technique for quantitatively analysing remote sensing image data. Although these approaches laid the foundation for surface identification, they often suffered from limitations in accuracy and generalisability, particularly in complex or variable environmental conditions.

2) Advancements in Deep Learning

The advent of deep learning (2012) marked a significant turning point in the field of artificial intelligence (AI). Convolutional Neural Networks (CNNs) have become the cornerstone of recent studies, offering superior performance in feature extraction and classification tasks.

CNN ResNet proposed by He et al. [7] marked an advance in image classification, showcasing remarkable results on the ImageNet dataset containing millions of images. Another key development milestone is the introduction of the Fully Convolutional Network (FCN) [8], which revolutionised semantic segmentation by replacing fully connected layers in traditional CNNs with deconvolution layers to better capture spatial details.

CNNs are also already used in the field of surface type. For example, include Deep U-Net [9] for pixel-level sea-land segmentation from Google Earth images. And Zhang et al.'s development of a novel rotated feature network to effectively tackle the challenges of multi-orientation in remote sensing object detection [10].

a) U-Net and LSTM for Deforestation Detection

Zhang et al. [11] developed a framework combining U-Net and LSTM for deforestation detection from optical satellite remote sensing images, using the Sentinel-2 dataset covering Guangxi Sanjiang Dong Autonomous County, China. This study introduced a U-Net architecture to extract spatial features while LSTM modelled time dependencies, resulting in an F1 score of 0.715. The method effectively detected changes from forest to bare soil, demonstrating the utility of combining CNNs with recurrent neural network structures for environmental monitoring.

b) DeepLabV3 + for Water Body Semantic Segmentation

Zhang and Zhao [12] utilized DeepLabV3 + enhanced with an Atrous Spatial Pyramid Pooling (ASPP) and an Attentional Feature Fusion Module (AFFM) for semantic segmentation of water bodies in high-resolution remote sensing images. Their improvements addressed issues with grid effects and information loss in traditional ASPP components, achieving better segmentation accuracy. This approach showcases the advancement in handling complex backgrounds and small area water bodies, crucial for precise water resource management.

c) PSPNet for Enhanved Land Cover Classification

In a notable advancement within the realm of environmental monitoring, Li and Guo [13] applied the Pyramid Scene Parsing Network (PSPNet) for precise land cover classification from remote sensing images. Their study, focused on the complex landscapes of the Nyingchi region, utilized PSPNet enhanced with a sophisticated feature integration strategy to differentiate between forested areas, water bodies, urban territories, and agricultural lands with high precision. By incorporating global contextual information and a dense feature pyramid, the model achieved a notable accuracy, highlighting its efficiency in processing mixed pixel environments typical in high-resolution satellite imagery. This application of PSPNet illustrates its capability to improve semantic segmentation tasks, providing crucial data for ecological and urban planning.

Given the remarkable achievements of deep learning in recognition and its proven effectiveness in various applications, this article explores the utilisation of deep learning methodologies for the extraction of water bodies from RGB image data used by the UAV.

III. Methodology

This paper presents a comparative analysis of contemporary methods for identifying surface types in RGB image data. The study focuses on Baroch Pond within a natural reserve near Pardubice, Czech Republic, which spans approximately 30,000 m², see Figure 1. This site, managed by the Pardubice Region Regional Office, is crucial for conservation efforts aimed at protecting grounded pond.

Figure 1 - Position of the Baroch Pond

A. Data Collection

DJI Mavic 3T (enterprise series) was used for the data collection. The characteristics of this UAV are as follows: weight, 920 g, four motors, max. flight time, 45 min. The drone contains a built-in cameras module. This module contains three cameras, which are a wide-angle RGB camera, a tele-RGB camera, and a thermal camera. The main RGB (wide-angle) camera was used for data collection, has 1/2" CMOS (Complementary Metal-Oxide-Semiconductor) sensor with 48 Mpx, objective with 84° angle with aperture f/2,8. The built-in camera module is attached to the drone by a three-axis gimbal [14].

Data collection was conducted using a UAV with a planned flight in March 2023 around noon, coinciding with the sun's nadir position. A total of 190 RGB images were captured during the flight. Specification planned flight: The UAV was flown at an altitude of 50 meters with a 70% overlap in both transverse and longitudinal directions.

Several constraints have the data collection process:

- Time Constraint: Data gathering was strategically timed around noon, ensuring optimal sunlight positioning.
- Zonal Constraint: a temperate climate zone, specifically near small bodies of water.
- Weather Constraint: The conditions were maintained close to ideal, and the sunny weather was a critical factor for consistency in data collection.

B. Data Processing and Classification

Data processing, calculations, and visualization tasks were executed on a computer equipped with an AMD Ryzen 7 processor at 3.2 GHz and 40 GB of RAM, which meet the computational needs for the data handling and classification methods utilized. To enhance the performance of deep learning applications, an Nvidia RTX 3070 GPU was employed, which provided processing speeds up to ten times faster than using only the CPU. For data processing and classification tasks, ArcGIS PRO software was utilized.

1) Data Processing

The images collected from the UAV flight were compiled using Pix4Dmapper (version 4.8.4 to create a unified and detailed orthomosaic of the study area through a process called orthorectification. The resulting mosaic is shown in Figure 2.

Figure 2 - Collected RGB Image Data in Single Mosaic

2) Model Training and Image Classification

Deep learning algorithms were selected for the task of image classification. Specifically, models such as U-Net, PSPNet, and DeepLabV3 were utilized because of their relevance and success in comparable applications.

In order to use these deep learning algorithms for classification, transfer learning 3 steps need to be performed:

1. **Data Preparation:** Area of interest about 10% of the original image was randomly selected from the original mosaic to include all surface types. In the area of interest were tiles manually marked and the following polygons (water and other) were added.
2. **Model Pre-Training:** U-Net, PSPNet, DeepLabV3, and models were trained using these labelled tiles with optimized model parameters provided below.

   U-Net: Model arguments: class balancing = false; mix-up = false; focal loss = false; ignore class = null. Data pre-processing: Batch size = 8; Validation = 10 %; Chip size = 224. Backbone model = ResNet-34. Max epochs = 20.

   PSPNet Model arguments: class balancing = false; mix-up = false; focal loss = false; ignore class = null; use unet = True; pyramid size = 1, 2, 3, 6; unet aux loss = False. Data pre-processing: Batch size = 8; Validation = 10 %; Chip size = 224. Backbone model = ResNet-34. Max epochs = 20.

   DeepLabV3: Model arguments: class balancing = false; mix-up = false; focal loss = false; ignore class = null. Data pre-processing: Batch size = 8; Validation = 10 %; Chip size = 224. Backbone model = ResNet-34. Max epochs = 20.
3. **Image Classification:** The image classification was performed using selected algorithms in the area of interest; see Figure 3, which is part of the origin mosaic.

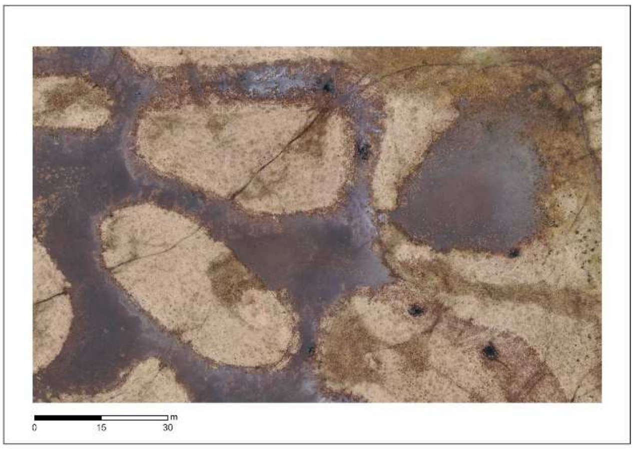

Figure 3 - Chosen Detailed Area of Interest

3) Comparison and verification of models

The results were verified on the original mosaics.

A comparison of the performance of classification models will be made using selected metric: Classification Accuracy (Acc); Definition see Formula 1.

$$Acc = \frac{Numer\ of\ correct\ predictions}{Total\ number\ of\ predictions} \qquad (1)$$

IV. RESULTS

Comparative analysis was conducted using several modern deep learning models to identify surface types in RGB image data collected over Baroch Pond. The models U-Net, PSPNet (Pyramid Scene Parsing Network), and DeepLabV3 were used. These models were selected for their relevance and established efficacy in similar applications. The classification accuracy for these deep learning algorithms was over 90%.

Figure 4 represents digitalised image data by manual identification of land cover surfaces. It is a graphical representation of reference data.

Figure 5 to Figure 7, shows the result of classification by selected deep learning models, with successful classification. For models U-net, PSPNet, and DeepLabV3 similar classification accuracy (Acc), as detailed in Table 1.

All deep learning classification models from the successful group were verified on the origin mosaic. Table 2 shows the result of the classification accuracy verified on the origin mosaic. Figure 8 shows classify origin mosaic by PSPNet, as the best classification model for full range of source data.

These results underscore the potential of deep learning to contribute significantly to environmental monitoring and management, particularly in the identification and management of water bodies.

Table 1 - Classification Accuracy on Area of Interest

Model	**Acc [%]**
U-Net	91,2
PSPNet	91,8
Deep Lab V3	92,4

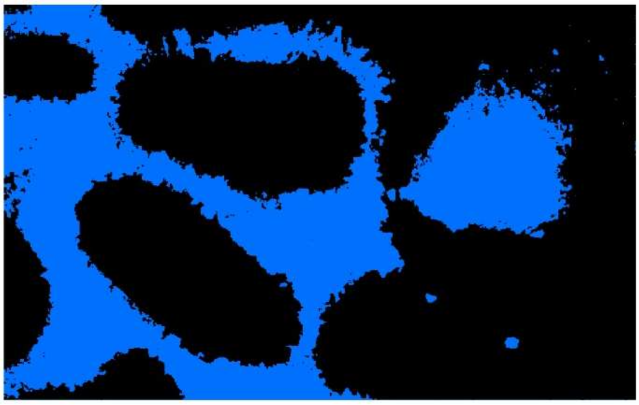

Figure 4 - Manual Identification (Reference Data)

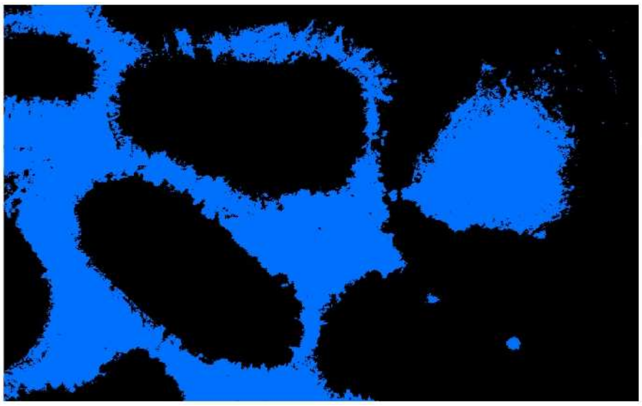

Figure 5 - Results Achieved by U-Net

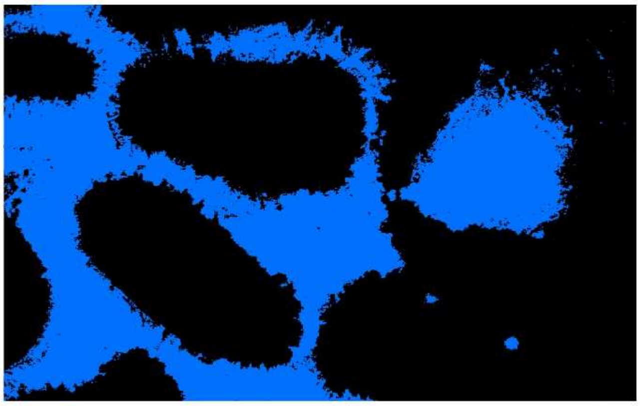

Figure 6 - Results Achieved by PSPNet

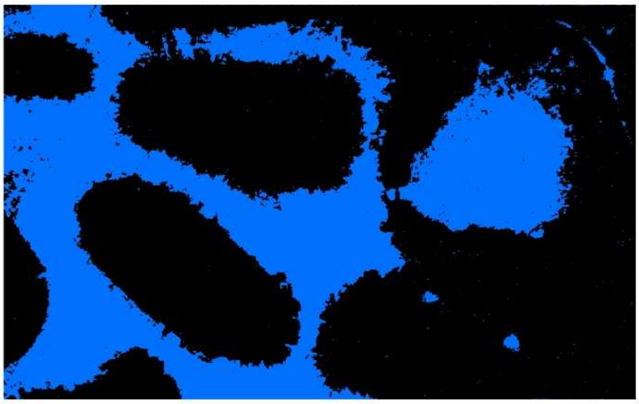

Figure 7 - Results Achieved by DeepLabV3

Table 2 - Classification Accuracy on Origin Data

Model	**Acc [%]**
U-Net	88,98
PSPNet	89,98
Deep Lab V3	89,57

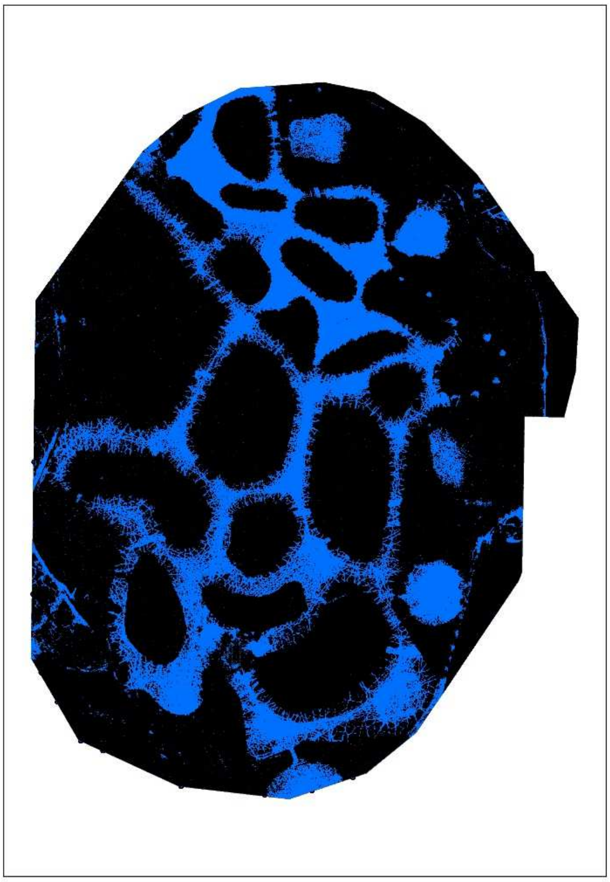

Figure 8 - Classification of Origin Mosaic

V. Discusion

This article has undertaken a comparison of deep learning algorithms for identifying surface types, with a particular focus on water bodies. The water bodies were classified on the RGB data that was acquired using the UAV. Through the comparison of various deep learning algorithms, we have strived to highlight advances in environmental monitoring and management strategies. This section discusses the achievements and limitations encountered in this endeavour and proposes directions for future research.

A. Discussion of Achievements and Limitations

Research has shown that with proper application of deep learning algorithms (U-Net, PSPNet, and DeepLabV3) particularly good surface type classification results can be achieved in RGB images.

However, the research also encountered several limitations. One of the main challenges lies in the inherent variability of the natural environment, which can significantly affect the reliability of RGB image data. Factors such as seasonal changes and the presence of algae or sediment in the water can change the appearance of surfaces and make identification difficult. In addition, relying on large, annotated datasets to train deep learning models presents logistical and resource challenges.

B. Suggestions for Further Research and Potential Directions

Given the achievements and limitations mentioned above, several avenues for future research emerge. First, more robust models that can be generalised under different environmental conditions are needed. This can include the integration of multispectral or hyperspectral imaging data that can provide additional information beyond what is visible in the RGB spectrum.

Furthermore, it is necessary to create large data sets that contain, among other things, time data. Incorporating temporal data could improve the identification of water bodies by accounting for seasonal fluctuations and long-term changes in the environment.

VI. Conclusion

In conclusion, comparative analysis of modern methods for surface type identification in RGB image data has shed light on the potential and challenges of using deep learning for environmental monitoring. Although significant progress has been made, the variability of natural environments and the limitations of current technologies highlight the need for ongoing research. By addressing these challenges and exploring new methodologies, we can continue to advance our understanding and management of the world's precious water resources. And this is key for a number of interested parties, for example, state administration, water resource managers, farmers, and tourism industry.

Acknowledgement

The authors thank the University of Pardubice, Students grant competition, SGS_2024_17 project for the support.

References

[1] C. Ma, J. Li, Z. Wang, X. Yi and L. Li, "Remote Sensing Image Recognition Method Based on Faster R-CNN," 2020 *International Conference on Computer Engineering and Application (ICCEA),* Guangzhou, China, 2020, pp. 869-872, doi: 10.1109/ICCEA50009.2020.00191.

[2] Y. Luo, J. Wu, Z. Zhang, H. Zhao and Z. Shu, "Design of Facial Expression Recognition Algorithm Based on CNN Model," *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)*, Shenyang, China, 2023, pp. 580-583, doi: 10.1109/ICPECA56706.2023.10075779.

[3] S. Wang, L. Chen, L. Xu, W. Fan, J. Sun and S. Naoi, "Deep Knowledge Training and Heterogeneous CNN for Handwritten Chinese Text Recognition," *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Shenzhen, China, 2016, pp. 84-89, doi: 10.1109/ICFHR.2016.0028.

[4] H. Avula, and A. S. Pillai, "CNN based Recognition of Emotion and Speech from Gestures and Facial Expressions," *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, Coimbatore, India, 2022, pp. 1360-1365, doi: 10.1109/ICECA55336.2022.10009316.

[5] T. Lillesand, R. W. Kiefer and J. Chipman, Remote sensing and image interpretation, 7th ed., John Wiley & Sons, 2015.

[6] J. A. Richards, Remote Sensing Digital Image Analysis: An Introduction. 5th ed. Springer, Berlin, Heidelberg, 2015, 494 pp. ISBN 978-3-642-30061-5, doi: 10.1007/978-3-642-30062-2.

[7] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[8] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 3431-3440, doi: 10.1109/CVPR.2015.7298965.

[9] R. Li *et al*., "DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 3954-3962, Nov. 2018, doi: 10.1109/JSTARS.2018.2833382

[10] K. Zhou, Z. Zhang, C. Gao and J. Liu, "Rotated Feature Network for Multiorientation Object Detection of Remote-Sensing Images,"

in *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 1, pp. 33-37, Jan. 2021, doi: 10.1109/LGRS.2020.2965629.

[11] J. Zhang, Z. Wang, L. Bai, G. Song, J. Tao and L. Chen, "Deforestation Detection Based on U-Net and LSTM in Optical Satellite Remote Sensing Images," *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Brussels, Belgium, 2021, pp. 3753-3756, doi: 10.1109/IGARSS47720.2021.9554689.

[12] X. Zhang and S. Zhao, "Semantic Segmentation of Water Body in High-Resolution Remote Sensing Images Based on DeepLabV3+," 2023 3rd International Conference on Electronic Information Engineering and Computer Science (EIECS), Changchun, China, 2023, pp. 1061-1064, doi: 10.1109/EIECS59936.2023.10435442.

[13] Z. Li and Y. Guo, "Semantic segmentation of landslide images in Nyingchi region based on PSPNet network," *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*, Changsha, China, 2020, pp. 1269-1273, doi: 10.1109/ICISCE50968.2020.00256.

[14] DJI - Official Website (n.d.), DJI Official, https://www.dji.com/, 1.2