ISPRS Journal of Photogrammetry and Remote Sensing 158 (2019) 1–10

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Application of convolutional neural networks for low vegetation filtering from data acquired by UAVs

Wojciech Gruszczyński*, Edyta Puniach, Paweł Ćwiąkała, Wojciech Matwij

AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering, al. Mickiewicza 30, 30-059 Cracow, Poland

ARTICLE INFO

Keywords:
Unmanned aerial vehicle
Digital elevation model
Ground filter
Convolutional neural networks

ABSTRACT

The main advantage of using unmanned aerial vehicles (UAVs) is the relatively low cost of collecting data, especially when using photogrammetry on images of relatively small areas. Additionally, they have high operational flexibility and the results have a high spatial and temporal resolution. To further facilitate the use of UAVs in photogrammetry, we developed an algorithm to filter out points that indicate areas covered in low vegetation (grass, crops) from the generated point cloud. This paper presents a three-layer filtering algorithm based on convolutional neural networks (CNNs) created for this specific purpose. The modular structure of the algorithm makes it easy to expand on and improve. The proposed solution allows errors in the height of digital elevation model (DEM) points caused by the influence of vegetation to be reduced by as much as 60–70% in relation to height errors from the raw data of high grass. At the same time, the solution presented here is practical for low grass because it does not weaken the model. The algorithm significantly reduces the errors in the DEM, as well as the products derived from the DEM.

1. Introduction

In recent years, the use of unmanned aerial vehicles (UAVs) to acquire high resolution spatial data has been the subject of many studies (Agüera-Vegaa et al., 2018; Cook 2017; Ćwiąkała et al., 2018; Kršák et al., 2016; Polat and Uysal, 2018; Rusnák et al., 2018; Salach et al., 2018). The main advantage of using UAVs is their relatively low data acquisition cost, especially when using photogrammetry on relatively small areas. In addition, UAVs have a high operational flexibility compared with manned systems and the results obtained have a better spatial and temporal resolution. However, the data acquired via remote sensing methods must be properly developed to fully utilize their inherent potential. For this reason, the creation of a digital elevation model (DEM) must be preceded by data filtration, so that the model generated is a DEM, and not a digital surface model (DSM).

A DEM represents the actual topography of the earth surface. Point cloud filtration to remove non-ground points is a critical and difficult step when generating the DEM, especially for areas with a very diverse topography. There have been many algorithms developed to filter point clouds. These algorithms can be divided into several categories: interpolation-based, slope-based, segmentation-based, and morphological methods. However, these algorithms have been developed to filter airborne laser scanning (ALS) data, which is significantly different from the data collected using photogrammetry.

The effectiveness of the filters described in the literature depends on several factors, including the topography of the research area, size of the above-ground objects, or the point cloud density (Meng et al., 2010; Yilmaz and Güngör 2018). However, only a few studies have focused on filtering UAV-based photogrammetric point clouds for DEM extraction purposes. Yilmaz and Güngör (2018) tested the effectiveness of UAV-based point cloud filtration using five algorithms originally developed for ALS data filtering. Their analysis showed that the adaptive triangulated irregular network (ATIN) algorithm gave the best results for point cloud filtering, while the progressive morphological 2D (PM2D) yielded slightly worse results. Both algorithms were also the least susceptible to changes in the point cloud density. In other studies (Yilmaz et al., 2018b), seven filtering algorithms, available in both commercial and non-commercial software, were analysed, with cloth simulation filtering (CSF) found to produce the best filtration results. Similar studies were conducted by Zeybek and Sanlıoglu (2019), yielding almost identical conclusions. Zhang et al. (2018) analysed how the DEM accuracy was affected when filtering the photogrammetric point cloud using a standard ALS filter (LASground). One of the conclusions was that all points on a bare-earth surface and in areas covered with grass are likely to be classified as ground points, which negatively affects the DEM accuracy. Tan et al. (2018) focused on the use of UAV-based data

* Corresponding author.
E-mail addresses: wgrusz@agh.edu.pl (W. Gruszczyński), epuniach@agh.edu.pl (E. Puniach), pawelcwi@agh.edu.pl (P. Ćwiąkała), matwij@agh.edu.pl (W. Matwij).

https://doi.org/10.1016/j.isprsjprs.2019.09.014
Received 12 March 2019; Received in revised form 20 September 2019; Accepted 23 September 2019
Available online 27 September 2019

for riverbank monitoring. The authors reported that the progressive morphological (PM) algorithm was not effective for areas covered with dense vegetation because the vegetation points can be incorrectly classified as ground points. For this reason, they developed the improved progressive morphological filter (IPM). The proposed algorithm, along with five other commonly used filtering algorithms, was used to process data collected from UAVs in four research areas. It was found that the IPM gave the best filtration results for most of the analysed datasets.

Recently, a new UAV-based point cloud filtration algorithm based on the integration of an orthophotomap and point cloud was developed (Yilmaz et al., 2018a). The first stage of this algorithm is to classify the orthophotomap using a support vector machine (SVM) and then superimpose it on the generated point cloud to determine the non-ground points. The selected points are then removed from the dataset in the final step of the algorithm. The methodology allows the generation of a DEM with an accuracy of better than half a metre for areas densely covered with trees or other above-ground objects. The DEM of flat areas has an accuracy of approximately 10 cm; however, areas covered with grass are assigned to the ground points class.

Artificial neural networks (Hinton, 1988; Bishop, 2006; LeCun et al., 2015), and in recent years the so-called deep neural networks (DNNs) (Goodfellow et al., 2016), have gained popularity in many tasks related to image classification (Cireşan et al., 2012; Liu et al., 2018). One type of DNN, which is particularly well-suited for this purpose, is the convolutional neural network (CNN) (LeCun et al., 1998; Krizhevsky et al., 2012; Mboga et al., 2017; Shelhamer et al., 2017; Sun et al., 2019).

Artificial neural networks have also been used to generate DEMs based on point clouds. Hu and Yuan (2016) applied a CNN to filter ALS data. In the first step, the point cloud was converted to a grid. Each pixel of the raster has three assigned attributes: the minimum, maximum, and average heights. Over 17 million pre-labelled training samples are then used to train the CNN, which is capable of distinguishing between ground and non-ground points. This method then obtains the exact results but requires a significant amount of labelled data. In addition, the method has a high computational cost, because point-to-image conversion is performed for each point of the cloud. The proposed solution to this problem is to convert all the cloud points into one large multi-channel image (using the elevation, intensity, and the lowest point in the neighbourhood) (Rizaldy et al., 2018). The classification is then performed using the fully convolutional network (FCN). These algorithms cope well with the ALS data classification; however, they focus on segmenting relatively large objects (buildings, trees) from the surface area. The present work focuses on how filtration can be conducted at a more detailed level, i.e., the separation of points representing low vegetation from points located on the bare-earth surface. The ALS data have a significantly lower resolution than photogrammetric data obtained from a UAV. Taking this into account and due to the completely different nature of the recorded data, the algorithms referred to above cannot be directly implemented in the UAV-based point cloud filtration.

Over the last year, there were individual attempts to use CNNs to filter and classify the data from UAVs, including DEM extraction. For example, Gevaert et al. (2018) proposed a two-stage criterion to obtain training examples. The first stage uses simple morphological filters to pre-classify the points as ground and non-ground points. The second stage selects information on the geometry of the objects from the first stage and uses radiometric data from the photographs. The combination of these steps allows the selection of appropriate training examples for the CNN. In this case, there is no attempt to filter the points for low vegetation from the bare Earth surface. A quality assessment of the generated DEMs was performed without the presence of ground control points (GCPs) in the field and was based only on DEMs generated using existing and wide-spread algorithms.

Another proposed method that uses a CNN to filter is based on the use of both orthoimage fragments and raw photographs of these fragments to select training examples (Liu and Abd-Elrahman, 2018). The experiments indicated that increases in the amount of input data (especially supplementing data with RGB images obtained from UAV missions) may significantly improve the classification quality. The study reported the results of a detailed vegetation classification based on photogrammetric data collected from a UAV, but recognition of the bare-earth surface was completely omitted.

The aim of the current research was slightly different from those of the studies described above. Our intention was to remove or minimize the impact of any factors that caused a discrepancy in the determined ground heights. The impact of vegetation is one such significant problem owing to the variability of height associated with vegetation and/or the cultivation cycle, which can affect the determination of height values. The motivation for this research was the relatively high ALS cost for small areas rather than the accuracy of the method, which is sufficient for the purpose set. Therefore, our purpose was to achieve an accuracy similar to that achieved by ALS when determining ground heights, but at a much lower cost for small areas. This is important when the study area is small and the frequency of measurements has to be relatively high, which may be the case when monitoring the stability of the terrain surface.

Currently available tools and computing environments allow for the relatively simple implementation and use of CNNs. In this study, MATLAB (version 2018b) with the Deep Learning Toolbox™ from MathWorks was used. All descriptions of the network structures are therefore presented in accordance with the convention adopted in this package.

The aim of the described research was not on increasing the speed of operation, but instead focused on the correctness of the network training to ensure good generalization. Efforts were made to maintain the most pragmatic approach, which allowed the use of a network with very simple processing for the input data. The brightness of pixels in individual RGB channels was not used as input data, with only the point cloud geometry actually used. This approach was geared towards enabling the use of CNNs even with a relatively small training set, which bypassed the problem of lighting and colour changes related to the season or humidity of the area and its vegetation. The methodology used to implement CNNs when generating a DEM only exhausted a small portion of the overall applicability of this approach. However, the promising results indicated that the problem was correctly formulated.

This paper is organized as follows. Section 2 describes the testing site, surveys, and data processing method. Section 3 presents the results of the applied filtering. Section 4 discusses the results against a previously developed algorithm based on local minima (Gruszczyński et al., 2017). Finally, Section 5 summarizes the tests and presents brief conclusions.

2. Materials and methods

The filtration method proposed here uses a CNN to analyse images generated based on the height of points as determined using photogrammetry from a UAV. In the following subsections, the methodology used to collect measurement data (UAV, reference) as well as the data processing used to filter out the points that reflect vegetation are described.

2.1. Datasets

The study used datasets acquired at two different locations in the south of Poland. Data acquired at the village of Łaziska were used to train and validate the network, while data from the village of Jerzmanowice were used for testing (filtering and evaluation using the trained neural networks). In both cases, the measurements included a UAV dataset and reference measurements for the height of points located on the ground mainly using a total station. At the same time, during the measurements, objects were divided into regions with the

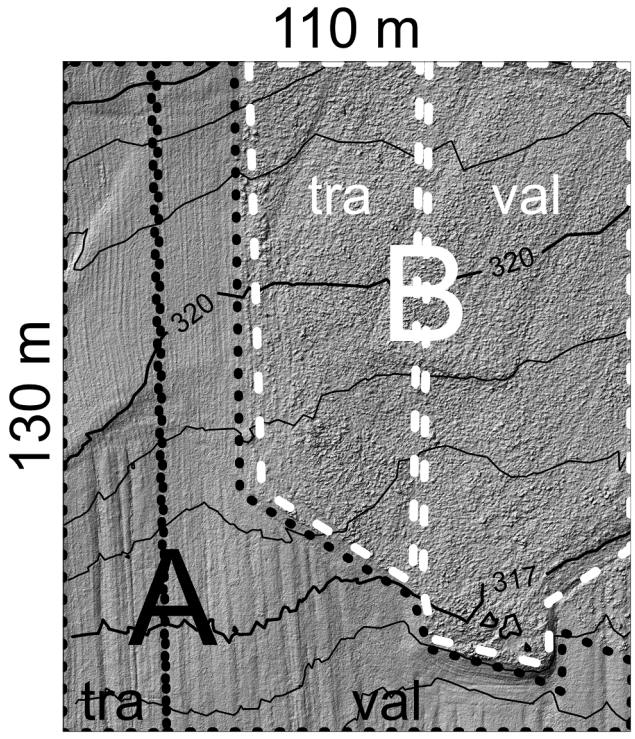

Fig. 1. Division of the Łaziska dataset: A - low grass (mowed), B - high grass (about 60 cm tall), tra - data used for network training, val - data used for training validation.

same type and condition of plant coverage.

The initial data processing focused on creating dense point clouds and determining coordinates for the reference points in a homogeneous coordinate system.

2.1.1. Details of the Łaziska dataset

The total area of the Łaziska dataset was approximately 1.4 ha, 40% of which was used as the training set, with the rest used as the validation set (Fig. 1).

The entire research area in Łaziska (Fig. 1) was measured using two methods: a tacheometric method as the reference and a photogrammetric method (Gruszczyński et al., 2017). All observations were connected to four points in the control network, whose coordinates were determined to within an accuracy of 2–3 mm.

The reference measurements were captured using a Leica Nova MS50 total station and a 360^0 prism mounted on a pole. A total of about 1,700 points were measured to reflect the terrain, with approximate dimensions of 130 × 110 m. It was estimated that the accuracy of the spatial position for any given reference point was about 10 mm.

Photogrammetric measurements of the research area in Łaziska were made using the DJI S900 hexacopter equipped with a Sony ILCE-6000 camera and an SEL35F18 lens. The average flight altitude was 100 m above ground level, which allowed images to be captured with an 11 mm ground sampling distance (GSD). The side and forward overlaps were designed to be at least 50% and 70%, respectively. The UAV survey results did not reveal any errors that would indicate that the overlap values might have been chosen incorrectly. Nine evenly spaced control points were used to appropriately correlate the photogrammetric measurements with the reference data. The data were processed using the Structure from Motion (SfM) algorithm, which was implemented using Agisoft PhotoScan Professional v. 1.1.6.2038. The dense point cloud that was generated, representing both the terrain and its coverage, had a density of 2200 points/m^2.

2.1.2. Details of the Jerzmanowice dataset

The total area of the Jerzmanowice dataset used to assess the

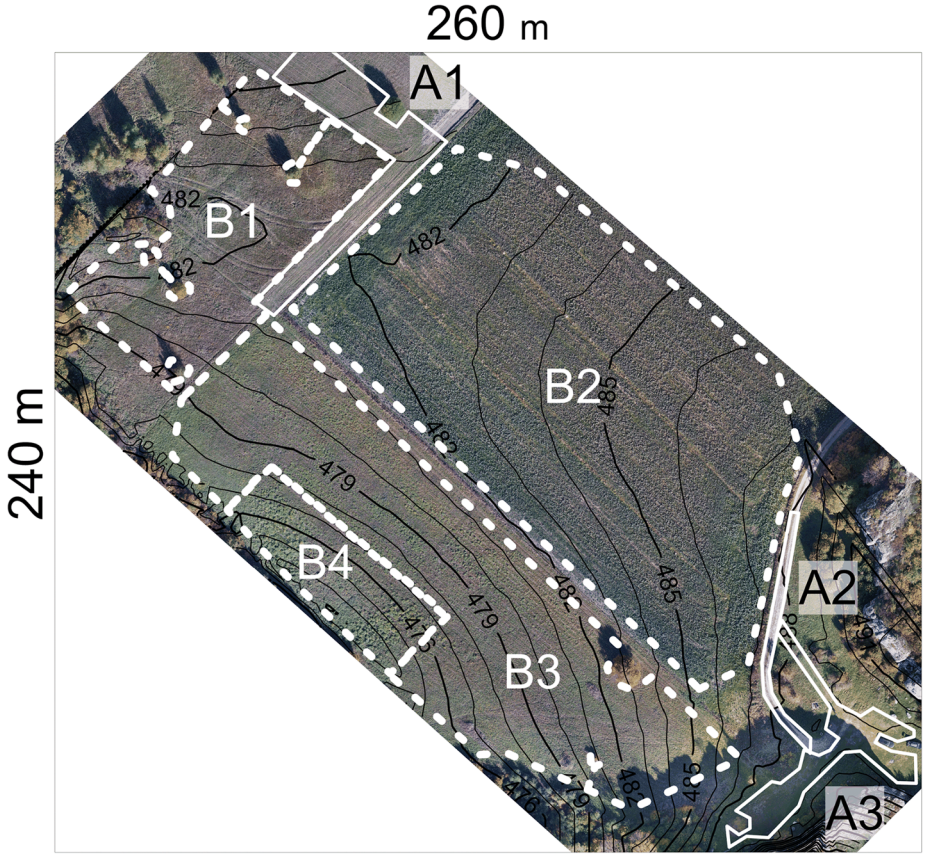

Fig. 2. Division of the Jerzmanowice dataset into regions of groundcover with a uniform type and height: A1 - mowed grass, A2 - paved road, A3 - mowed and trodden grass, B1 - wild meadow (grass height of approximately 40–80 cm), B2 - field covered with young broad bean (height of approximately 60 cm), B3 - meadow (grass height of approximately 10–30 cm), B4 - meadow (grass height of approximately 30–50 cm).

filtering accuracy was approximately 2.7 ha. Fig. 2 presents the division of the area into regions of groundcover with a uniform type and height. The analysis excluded all trees in the measurement area.

Similar to the Łaziska dataset, the research area in Jerzmanowice was characterized using reference measurements and photogrammetry with a UAV. The vast majority of the reference data regarding the terrain were collected using the tacheometric method, and only small parts of the areas covered by high grass were measured with the Global Navigation Satellite Systems real time network (GNSS RTN) technique.

All measurements were connected to four points of the control network established in the research area. Three of them were measured using the GNSS static method with Leica GS16 receivers. The common recording time of the satellite observations at the points was 2.5 h (with reference vectors approximately 20 km long). In addition, angular-line observations were made between all four points of the network using a Leica Nova MS50 total station. The coordinates of the network points were determined through the combined alignment of the angular-linear observations and the GNSS vectors in relation to two reference stations of the SmartNET network. The root mean square (RMS) position error was 3.0 mm and the RMS height error was 4.5 mm.

A Leica Nova MS50 total station and 360^0 prism mounted on a pole were used for terrain reference measurements using the tacheometric method. In this way, approximately 1200 points located primarily in the central part of the area were measured (entire regions for A2, A3, B2, and B4 and large portions of regions A1, B1, B3, and B4). Based on the specifications of the equipment and the measurement conditions, it was estimated that the accuracy of the measurement was approximately 10 mm. In turn, the coordinates of the 950 points located in the regions marked in Fig. 2 as A1, B1, B3 and B4 were measured using the GNSS RTN method. The estimated accuracy of the spatial positions for these measurements was approximately 50 mm. However, in practice, the accuracies for both methods appeared to be better than the rated values (about 1 cm).

Photogrammetric data for the research area in Jerzmanowice were collected using the same equipment configuration as for Łaziska (DJI S900, Sony ILCE-6000 with SEL35F18 lens). The photogrammetric flight took vertical images with a side overlap of not less than 55% and forward overlap of not less than 75% from an altitude of 90 m above ground level. In this way, 331 images were acquired with a 10 mm GSD. The coordinates of the 20 control points were measured using the GNSS RTN method. A number of control points in the central part of the research area were also measured with the Leica Nova MS50 total station to verify the coherence of the GNSS measurements with tacheometric surveying. The development of photogrammetric data was performed using Agisoft Metashape Professional v. 1.5.0.7492. Aerotriangulation was performed at the highest level of accuracy, meaning that the program worked on images that were oversampled by a factor of four. To improve the quality of the image alignment, outliers were removed using the gradual selection tool in the software. Optimization was performed in the next stage, including a re-alignment of the aerotriangulatory block and determination of the camera calibration parameters. The RMS error for the spatial position of the control points was 16.5 mm. At the end of this stage, a dense point cloud was generated at the high detail level, which means that the program algorithm sought to determine spatial coordinates for each group consisting of 4 pixels in the image (2 × 2 pixels). In this way, a point cloud with a resolution of approximately 3200 points/m^2 was generated.

2.2. Data processing

The main part of the processing was performed using the CNN. The data processing was divided into the following three stages:

1. Classification of the point cloud into the high grass (*HG*) or low grass (*LG*) classes. Fragments of cloud considered to belong to the *LG* class were also considered as belonging to the terrain (*Ground*) class.
2. Classification of point cloud fragments that represented the *HG* class for points reflecting the *Ground* class and the remaining points (*NonGround* class). Points found to belong to the *Ground* class were retained, while those belonging to the *NonGround* class were discarded from the point cloud.
3. The determination (regression) and corrections of heights of the points in the cloud for the cloud fragments that were classified simultaneously as the *HG* class and *Ground* class.

Regression (3rd point above) is an additional step that does not necessarily have to be implemented. However, enabling it allows a reduction in systematic errors in the heights determined from the UAV data.

The use of neural networks for data processing requires training to determine the network weights. The CNN training requires both input data and target responses in a process known as supervised learning. The input data for all networks were prepared in the same way as described in Section 2.2.1. Sections 2.2.2–2.2.4 describe the preparation of the training data (target responses) for the networks used in the individual processing stages as well as the structure of the neural networks.

2.2.1. Preparation of input data

Neural network input data should have a clearly defined structure. All input variables, both in the training stage and the subsequent use of a trained network, should have values from a specific closed set or interval. These intervals should be defined at the training stage and should be the same when using trained networks.

The input data were presented as fragments of a greyscale image, with an 8-bit depth. The images were created as follows:

1. The entire point cloud was divided into adjacent cells (squares), with dimensions of 5 × 5 cm. In each cell, the point with the lowest height was selected and attributed as the height of the cell, while the remaining points were discarded. The point cloud processed in this way is almost uniformly dense, except for the single cells that contain no points. Cell values were transcribed into the matrix labelled R.
2. The missing entries in the R matrix were filled using a nearest neighbour interpolation method.
3. A G matrix was created by applying a Gaussian filter to the R matrix. The square, Gaussian filter had a size of 105 × 105 cm (21 × 21 elements), with a standard deviation of 25 cm (5 elements).
4. An M matrix was created as the difference between the R and G matrices:

$$M = R - G$$

5. The M matrix elements were processed into a closed interval and saved as an image I. All values of the M matrix elements were processed in the following manner (pseudocode):

$$N(i, j) = (M(i, j) - \min)/(\max - \min)$$

$$\text{If } N(i, j) < 0 \text{ then } N(i, j) = 0$$

$$\text{If } N(i, j) > 1 \text{ then } N(i, j) = 1$$

$$\min = -0.3 \text{ m}$$

$$\max = +0.3 \text{ m}$$

$$I = \text{round}(255 * N)$$

where

min and max - constants chosen arbitrarily based on the extreme values from the M matrix of the training set,
N - values from the M matrix normalized to the range from 0 to 1,
I - matrix containing values saved in the form of an image, the

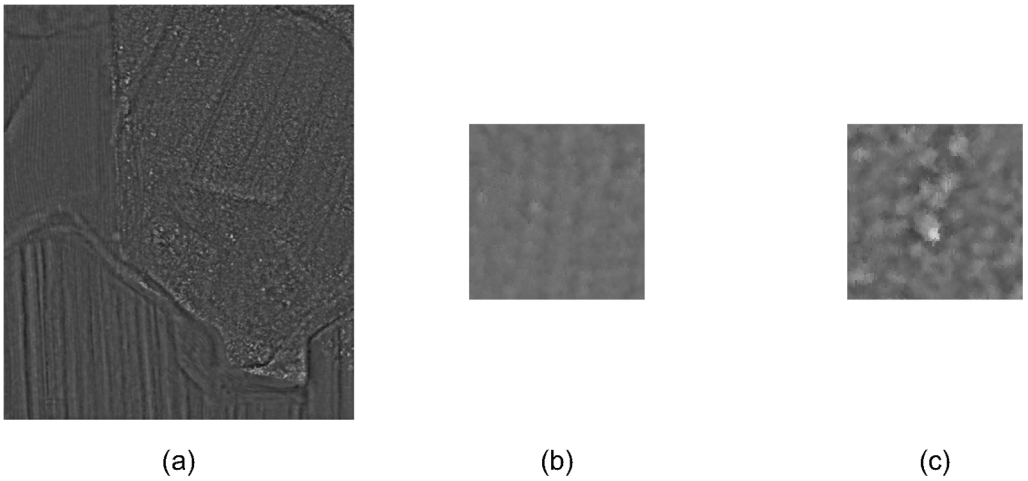

Fig. 3. Full image I for the Łaziska dataset (a) and its exemplary fragments used as the inputs to neural networks from the regions covered by low (b) and high (c) grasses.

fragments of which were used as the inputs to the neural network (Fig. 3).

The network input images were 65 × 65 pixels. For this reason, the actual images recorded on the disk had to be a larger size, i.e. $(65 \times 65) * 2^{0.5}$. As a result, during training when rotating the image by some angle, each network input had a specific value. After rotation, the images were cropped at their centre to fit in a window of 65 × 65 pixels. For all the CNNs, the pixel considered was located in the middle of a given example. This pixel corresponded to the values of the network outputs. Data from the Łaziska site were used for training and validation in all cases, and the data from the Jerzmanowice site were used to evaluate the trained network.

2.2.2. Classification into points with high and low grass

The targets for the *HG* and *LG* classes were assigned based on field observations. The structure of the CNN used to implement this stage of processing is summarized in Table 1. A standard network structure was used. It was chosen to maintain simplicity. Because the processing speed was not a key criterion, and the effectiveness of the classification turned out to be completely satisfactory, the selected network architecture was considered correct for the given task.

The training examples were cropped from the full image by shifting the window vertically and horizontally every 3 pixels (the examples were created for approximately 1/9 pixels). The total number of training and validation examples was around 540,000.

2.2.3. Classification into points located on the ground and others

Only data from the area covered with *high grass* were used to train the CNN classifying points for the *Ground* and *NonGround* classes. The assignment of target classes was based on the reference measurements. The reference surface was created using a linear interpolation between the reference points. If the point of the cloud had a height that was less than 5 cm from the reference model, it was considered to be a point on the ground (*Ground* class). Otherwise, the point was assigned to the *NonGround* class.

Only around 10% of the points from the Łaziska dataset used for training were assigned to the *Ground* class. This caused an even selection of examples from a full picture, resulting in a very uneven number of examples for the classes in the training set. This could lead to difficulties in effective training, because the training set will have a relatively small number of examples that belong to one of the classes. For example, a model that always places the training examples in the *NonGround* class will still have a 90% effectiveness. Therefore, the training was balanced as follows:

- all examples (from the training/validation area) that belonged to the *Ground* class were selected for the training/validation set,
- only elements from every third row and column (from the training/validation area) were considered, and if it belonged to the *NonGround* class the corresponding part of the image was selected for the training/validation set.

As a result, a much more balanced training set, with a ratio of 58%/42% (*Ground/NonGround*), was achieved, while maintaining a limited number of training and validation examples (about 560,000). This increased the probability of selecting an example that belonged to the *Ground* class. However, this disturbed the distribution of probabilities occurring in the data set.

Therefore, when using the trained network for classification, an additional parameter also had to be used, i.e., the minimum probability (threshold) at which the example is considered to belong to the *Ground* class during filtering. By default, when classifying into one of the two available classes, this threshold was set to 0.5. However, if the proportion between the classes in the full set of data was disturbed, the threshold should be set higher. The selection of this threshold value

Table 1
Structure of the network used for the classification of vegetation into the high grass (*HG*) and low grass (*LG*) classes.

Layer number	Layer type	Details
1	Image input	65 × 65 × 1, zero centred normalization
2	Convolution	10 units, 5 × 5 × 1 convolutions with stride [1 1] and no padding
3	ReLU	Threshold operation for each element of the input layer, where any value less than zero is set to zero
4	Max pooling	3 × 3 max pooling with stride [2 2] and no padding
5	Fully connected	2 fully connected units
6	Softmax	
7	Classification output	Output *HG* and *LG* classes, cross entropy loss

Table 2
Structure of the network used to classify the points into the *Ground* and the *NonGround* classes.

Layer number	Layer type	Details
1	Image input	65 × 65 × 1, zero centred normalization
2	Convolution	20 units, 5 × 5 × 1 convolutions with stride [1 1] and no padding
3	ReLU	Threshold operation to each element of the input layer where any value less than zero is set to zero
4	Max pooling	3 × 3 max pooling with stride [2 2] and no padding
5	Convolution	20 units, 3 × 3 × 20 convolutions with stride [1 1] and padding [1 1 1 1]
6	ReLU	Threshold operation to each element of the input layer where any value less than zero is set to zero
7	Max pooling	2 × 2 max pooling with stride [1 1] and no padding
8	Fully connected	2 fully connected linear units
9	Softmax	
10	Classification output	Output *Ground* and *NonGround* classes, cross entropy loss

affected the results that were obtained in the following ways:

- at higher thresholds, there were fewer points left as ground points after the cloud filtration, but it was more likely that these points were correctly classified,
- at smaller thresholds, there were more points that were considered to belong to the *Ground* class, but a larger proportion of these points were classified incorrectly.

The threshold value can be treated as a parameter that falls in the range from 0 to 1. The selection of this threshold had to be made when performing the filtering and could be optimized based on the validation set or it could be selected arbitrarily based on the filtering results of a specific set when using the trained networks.

Initially, a network structure identical to that for the *HG* and *LG* classification was chosen for this step. However, the results turned out to be less satisfying than those obtained for the previous classification. Therefore, it was decided to expand the network structure to the one presented in Table 2. This slightly improved the results. Attempts were also made with larger networks, but this did not improve the results. After further attempts, the manner of assigning examples to reference classes and balancing the number of examples in the classes for the training and validation sets was considered to be crucial for network operation. Therefore, it was decided that the structure used met the requirements.

2.2.4. Setting the correction to the point heights classified as Ground

Training was performed on image fragments corresponding to pixels classified as *high grass* for both *Ground* and *NonGround* classes. The CNN was later used to process (for the Jerzmanowice dataset) only examples classified by the network as *Ground*. The structure of the network was selected following the second stage of processing (*Ground/NonGround* classification). There were also attempts to use slightly larger CNN structures, but this did not result in any improvement. Finally, it was decided to use the network with the structure presented in Table 3, because it met the goals set.

Training examples were cropped from the full image by shifting the window every 3 pixels both vertically and horizontally. As a result, the total number of training and validation examples was around 300,000.

3. Results

The Łaziska dataset was used to train and validate the neural networks, while the Jerzmanowice dataset was used to test the trained networks. In Section 3.1, a brief summary of the training results for the best obtained solutions is provided. A wider description of the results is provided for the test dataset (Jerzmanowice), which allows an unbiased assessment of the quality of each filtering step and the entire algorithm.

3.1. Results for the Łaziska dataset

All the results described in this subsection refer to the data used to validate the neural networks, i.e. examples that were not used directly to determine the weights during training. None of the trained CNNs showed any signs of overfitting. The errors and other indicators from the validation and training groups were of a similar magnitude.

3.1.1. Results for the CNN used to classify land cover for low and high grasses with the Łaziska dataset

The results for this network are best summarized by the confusion matrix presented in Table 4. The effectiveness of the neural network was considered to be very high, as 99.3% of the cases were correctly classified. Because the percentage of misclassified cases was small, the possibility that these cases had incorrectly assigned target classes could not be ruled out. Classes were assigned to areas that were considered uniform in the field, and therefore errors resulting from excessive generalizations cannot be excluded at this level of accuracy.

3.1.2. Results for the CNN used to classify HG class points into Ground and NonGround classes for the Łaziska dataset

The confusion matrix for the classification of the validation area fragments into *Ground* and *NonGround* classes for a threshold of 0.5 and 0.9 is shown in Table 5. The accuracy of the classification was much worse than when classifying into the *LG* and *HG* classes, because only 72.1% and 66.5% of the validation cases were correctly classified, respectively, as shown in the tables. This may be due to a weaker

Table 3
Structure of the network used to determine corrections to the point heights.

Layer number	Layer type	Details
1	Image input	65 × 65 × 1, zero centred normalization
2	Convolution	20 units, 5 × 5 × 1 convolutions with stride [1 1] and no padding
3	ReLU	Threshold operation to each element of the input layer where any value less than zero is set to zero
4	Max pooling	3 × 3 max pooling with stride [2 2] and no padding
5	Convolution	20 units, 3 × 3 × 20 convolutions with stride [1 1] and padding [1 1 1 1]
6	ReLU	Threshold operation to each element of the input layer where any value less than zero is set to zero
7	Max pooling	2 × 2 max pooling with stride [1 1] and no padding
8	Fully connected	1 fully connected linear unit
9	Regression output	Mean-squared-error

Table 4
The confusion matrix used to recognize the high grass (*HG*) and low grass (*LG*) classes for the validation group of the Łaziska dataset.

		Target class	
		HG	*LG*
Output class	*HG*	172,214	2091
		52.9%	0.6%
	LG	93	151,235
		0.0%	46.4%

Table 5
The confusion matrix used for the recognition of *Ground* and *NonGround* classes for the validation group of the Łaziska dataset at a threshold of 0.5 and 0.9 (in brackets).

		Target class	
		Ground	*NonGround*
Output class	*Ground*	132,048	70,014
		43.7%	23.1%
		(52218)	(7002)
		(17.3%)	(2.3%)
	NonGround	14,446	85,976
		4.8%	28.4%
		(94276)	(148988)
		(31.2%)	(49.3%)

systematic relationship between the input and output values. At the same time, the target classes were less well defined than in the *LG* and *HG* classification. This was due to imperfections in the reference surface created by the interpolation between points with the actual measured values.

For the validation group, changing the threshold value from 0.5 to 0.9 increased the probability of correct classifications for the *Ground* class from approximately 65% to 88%. At the same time, this change reduced the probability of the correct classification of cases classified as *NonGround* from approximately 86% to 61%. At a threshold of 0.9, almost 96% of cases from the *NonGround* class were classified correctly.

3.1.3. Results for the CNN used to determine corrections to the point heights obtained from the UAV for the Łaziska dataset

Training and validation of the network to determine discrepancies between the heights of the point cloud and the reference data were performed only for the area covered with *HG*. Such corrections (estimated height differences) had to be subtracted from the heights of the point cloud to ensure that they equalled the reference heights. The distribution of the determined values of the correction for the validation set is shown in Fig. 4. The average output value of the network was approximately 11 cm, while the extreme values that occurred in individual cases extended from about −1 cm to over 40 cm.

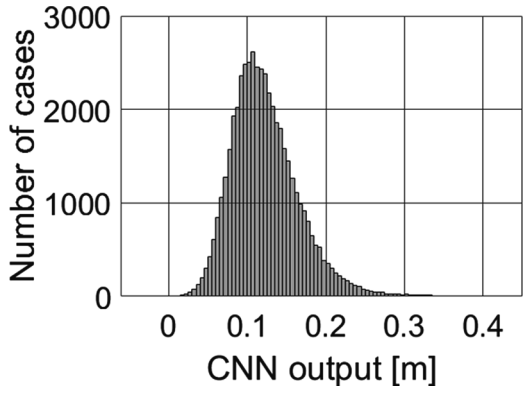

Fig. 4. Histogram of the CNN output values for the validation data group.

Table 6
Classification of regions of the Jerzmanowice dataset into high grass (*HG*) and (*LG*) classes.

Region designation	Land cover	Vegetation height [cm]	*LG* [%]	*HG* [%]
A1	Mowed grass	10	80	20
A2	Paved road	0	59	41
A3	Mowed, trodden grass	5	80	20
B1	Wild meadow	40–80	0	100
B2	Young broad bean (field)	60	0	100
B3	Meadow	10–30	0	100
B4	Meadow	30–50	0	100

The average difference between the reference and determined corrections was approximately −0.5 cm. This value reflects the average systematic error of the values forecasted by the CNN, i.e., the designated corrections to the height. The mean of the absolute values of these differences was approximately 3.5 cm, which better reflects the random errors of the model. This provides the accuracy of the average corrections to the heights of the point cloud. The model's RMS error for the validation set was approximately 4.5 cm.

3.2. Results for the Jerzmanowice dataset

The Jerzmanowice dataset was used to conduct an independent (unbiased) test for both the operation of the trained networks and the entire proposed point cloud processing algorithm.

3.2.1. Classification of high and low grasses for the Jerzmanowice dataset

The regions in the Jerzmanowice dataset marked in Fig. 2 as A1–A3, i.e., with a groundcover height up to 10 cm, were mostly covered with low grass (*LG* class), while the regions marked as B1–B4 were covered with high grass (*HG* class). Table 6 presents the percentage of each class as designated by the CNN for the particular regions, while Fig. 5 shows the classification results for the entire Jerzmanowice dataset. In the figure, fragments classified as *HG* are shown in black and fragments classified as *LG* are left without points, so that the corresponding areas were visible.

Small deviations in the classification of areas were caused by specific inhomogeneities (such as trees and other high obstacles) and not by the improper operation of the network. In the case of the paved road, the fragments that were classified in the *HG* class constituted the area around relatively deep ruts, along with the vegetation located between them.

3.2.2. Classification into Ground and NonGround classes for the Jerzmanowice dataset

To assess the classification accuracy, the heights between the points considered to be in the *Ground* class were interpolated to determine the heights at the points where the reference measurements were made. The *Ground* class included all points from areas classified as *LG* and points from areas classified as *HG* by the CNN (classifying into *Ground*) at a threshold value of 0.9.

Table 7 shows the calculated RMS errors before and after filtration with the use of neural networks and compares them to the RMS error calculated based on local minima algorithm filtration. A comparison of the RMS error values for the raw data confirmed the general correctness of the classification and the greater accuracy of the resulting height model in relation to the raw data. For regions A1 and A2, this improvement was almost negligible. However, for other areas, it ranged from around 30% to 50% of the error value of the raw data. In general, analogous to the situation before filtering, there was a noticeable relationship between the RMS errors and the height of the vegetation

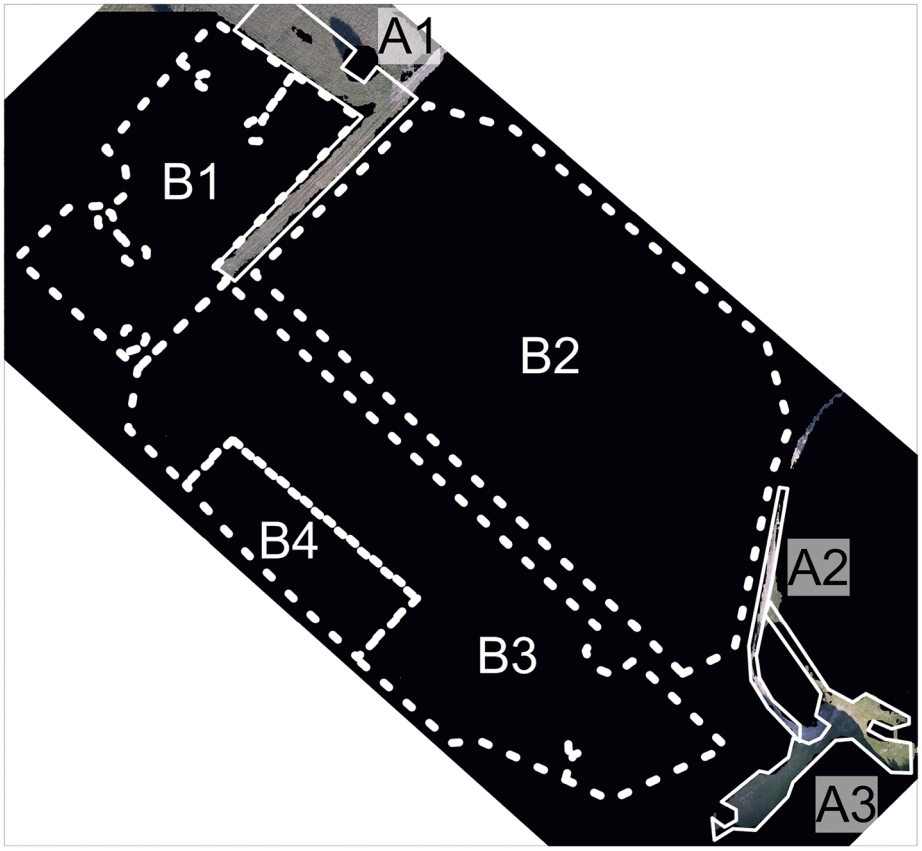

Fig. 5. Points classified by the CNN as high grass (*HG* class - black pixels).

covering a given area. For the areas marked as B1–B4, the RMS errors decreased by approximately 9–13 cm.

A visual analysis of the orthophotomosaic (Fig. 6) for the areas marked B1–B4, which were mainly overgrown with high grass, indicated the correctness of the classification (as compared to Fig. 2). The main designations for the *Ground* class were:

- area B1: paths, ruts, and locally trodden or less covered fragments,
- area B2: furrows between plants,
- area B3: furrows between plants visible on the orthophotomosaic but difficult to see in the field, and well-trodden paths,
- area B4: exposed fragments of soil between taller plants, and a fragment of shorter grass at the north-eastern border of the area,
- ruts on the dirt road between areas B2 and B3.

The errors in the classification mainly occurred in areas in which there were large and sudden changes in the height of the cloud points, i.e., in the direct vicinity of trees or rocky outliers. This was expected because neither the data processing nor the training and validation steps prepared the CNN for such occurrences.

3.2.3. Corrections of the point cloud height for the Jerzmanowice dataset

Only examples that were simultaneously placed in the *HG* and *Ground* classes were subject to correction. For areas marked A1–A3, the corrections resulted in a slight increase in the RMS error values, while for areas marked B1–B4 the errors were reduced by up to nearly 50% relative to the values without corrections (Table 8). This processed data allowed the generation of a model that was consistent with the reference data, as measured by the RMS error, with all the analysed areas below 12 cm.

4. Discussion

To assess the overall quality of the proposed data processing algorithm, its results were compared to the results of a previously proposed heuristic filtering algorithm based on local minima (Gruszczyński et al., 2017). Comparing both algorithms (Table 7) indicated that before the height correction was introduced, both algorithms provided similar RMS errors. However, after correcting the heights in the CNN filtered model following the method proposed here, the RMS error values were significantly lower than those obtained with the previously proposed algorithm.

Table 7
RMS errors of the terrain height determined from the cloud points of the Jerzmanowice dataset.

Region designation	Vegetation height [cm]	RMS error raw data [cm]	RMS error after CNN filtration [cm]	RMS error after CNN correction [cm]	RMS error local minima [cm]
A1	10	7	7	7	6
A2	0	3	2	3	4
A3	5	5	4	4	5
B1	40–80	31	20	12	21
B2	60	32	17	9	18
B3	10–30	19	10	7	13
B4	30–50	31	18	12	20

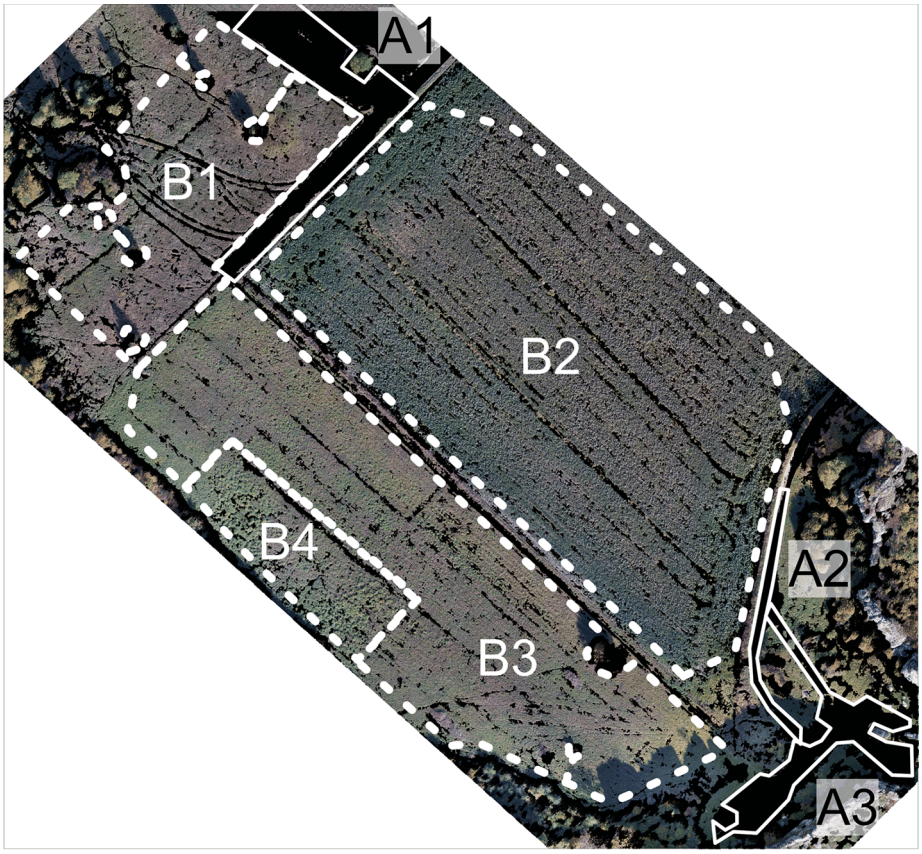

Fig. 6. Points of the Jerzmanowice dataset classified as terrain (*Ground* class - black pixels).

When the DEM was generated, one of the most important parameters was the even distribution of the points. The filtration described in the previous section resulted in the creation of areas without *Ground* points. Thus, the heights in these areas were determined solely from interpolation. It was decided that to compare the overall results of both algorithms, the largest radius of a circle that could be fit such that there were no points in the *Ground* class was determined. The largest radius of inaccessibility (LRI) determined in this way indicated the size of the areas from the model without height data as determined from the filtered points. In Table 8, the values of these LRI are compared for the CNN-based and local minima algorithms. The LRI value for the CNN-based algorithm depends on the threshold value adopted for the *Ground* and *NonGround* classification. Table 8 also shows the values of these radii for the adopted threshold of 0.9. In this case, for areas marked B1–B4, the LRI using the CNN-based algorithm was twice as large as that for the algorithm based on local minima. Thus, the algorithm based on local minima gave a much more even distribution of ground points. For smaller threshold values, the LRIs decreased, which indicated that this metric could be applied to select the appropriate threshold value.

Table 8
The largest radius of a circle without points in the *Ground* class.

Region designation	Vegetation height [cm]	Max radius CNNs [m]	Max radius minima [m]
A1	10	1.65	1.80
A2	0	0.65	1.25
A3	5	1.00	1.35
B1	40–80	3.05	1.55
B2	60	2.55	1.55
B3	10–30	3.65	1.80
B4	30–50	2.55	1.65

Dividing the proposed algorithm into several neural networks enabled an assessment of which tasks caused the most problems. An additional advantage is the possibility of introducing various solutions based on the results of previously used networks and heuristic algorithms.

While the classification using the CNNs for high and low grasses gave accurate results, there was still room for significant improvement in the subsequent stages. The key problems were the high-density reference data and/or the way the input data was formulated. The problem of high-resolution reference heights could be solved by surveying the areas before and after the grass is mowed. It was then assumed that the measurements of the low-mowed grass gave a height sufficiently close to the ground. A solution for non-mining (stable height) areas with low intensity agriculture, i.e., meadows, could be to use archived ALS data that is available at a low price. In general, extending the set of examples used during training to different types of crops and land cover will extend the applicability and improve the network's ability to generalize. For the input data formulation, after the expansion of the training dataset, the use of RGB data instead of, or in addition to, the height data needs to be tested.

Processing variants that slightly differed from those presented were also tested to determine their effect on the resulting classification accuracy. First, training examples (input images) with smaller dimensions, i.e., 17 × 17 pixels and 33 × 33 pixels, were tested. For examples with 17 × 17 pixels, the results obtained were worse (i.e., the accuracy of classification into *HG* and *LG* for the validation set was 94.7%). However, the examples with dimensions of 33 × 33 pixels had results that were close (i.e., the accuracy of classification into *HG* and *LG* for the validation set was 99.0%) to those presented (at 65 × 65 pixels). At the same time, the training time and subsequent operation of the correspondingly smaller CNNs were much shorter. A deeper analysis of the size of the input data could allow the determination of the optimal

dimensions for a given dataset.

Second, the results of the networks trained for different resolutions of the generated data were tested, and consequently were also tested for various training examples, i.e., every 1 and every 3 pixels. The applied solution that balanced the number of examples representing different classes gave the best results, while also maintaining a reasonable training time. Therefore, the decision of which data generation method to use was made primarily based on a pragmatic approach.

Third, different standard deviations of the Gaussian filter were tested with a range of up to 1 m when generating the data (item 3 Section 2.2.1). However, no major changes were observed in the generated examples or results of the network training. It was ultimately assumed that the adopted standard deviation of the Gaussian filter (at the level of 25 cm) allowed both the incorporation of local noise in the heights of the cloud points around the analysed point and a quick reaction to the actual changes in terrain geometry.

5. Conclusions

The use of CNNs with the proposed input data formulation enabled the correct classification of areas as low or high grass in nearly all cases, and therefore there is no need to improve the filtering stage. In contrast the classification of the cloud points as ground or non-ground points using the CNN gave promising results, but requires further processing or training using a more extensive dataset. In this study, the CNN used to classify points as ground or non-ground gave similar DEM accuracies to those obtained using an algorithm based on local minima. The introduction of a correction set by the CNN to the height of the cloud points located in areas overgrown with high grass resulted in significant reductions in the systematic deviations of the height of the cloud points from the true values. As a consequence of this additional step, the CNN-based filtering algorithm finally obtained a much better accuracy than results based on a local minima.

Declaration of Competing Interest

None.

Acknowledgements

This research was funded by the AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering [grant number 16.16.150.545]

References

Agüera-Vegaa, F., Carvajal-Ramíreza, F., Martínez-Carricondoa, P., Sánchez-Hermosilla López, J., Mesas-Carrascosab, F.J., García-Ferrerb, A., Pérez-Porrasb, F.J., 2018. Reconstruction of extreme topography from UAV structure from motion photogrammetry. Measurement 121, 127–138.

Bishop, Ch.M., 2006. Pattern Recognition and Machine Learning. Springer Science +Business Media, LLC.

Cireşan, D., Meier, U., Schmidhuber, J., 2012. Multi-column deep neural networks for image classification. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Cook, K.L., 2017. An evaluation of the effectiveness of low-cost UAVs and structure from motion for geomorphic change detection. Geomorphology 278, 195–208.

Ćwiąkała, P., Kocierz, R., Puniach, E., Nędzka, M., Mamczarz, K., Niewiem, W., Wiącek, P., 2018. Assessment of the possibility of using unmanned aerial vehicles (UAVs) for the documentation of hiking trails in alpine areas. Sensors 18 (1), 81.

Gevaert, C.M., 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.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Gruszczyński, W., Mawij, W., Ćwiąkała, P., 2017. Comparison of low-altitude UAV photogrammetry with terrestrial laser scanning as data-source methods for terrain covered in low vegetation. ISPRS J. Photogramm. Remote Sens. 126, 168–179.

Hinton, G.E., 1988. Neural network architectures for artificial intelligence. American Association for Artificial Intelligence Menlo Park, CA.

Hu, X., Yuan, Y., 2016. Deep-learning-based classification for DTM extraction from ALS point cloud. Remote Sens. 8 (9), 730.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Adv. Neural Inform. Process. Syst. 1097–1105.

Kršák, B., Blišťan, P., Pauliková, A., Puškárová, P., Kovanic, L., Palková, J., Zeliznaková, V., 2016. Use of low-cost UAV photogrammetry to analyze the accuracy of a digital elevation model in a case study. Measurement 91, 276–287.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-Based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Liu, T., Abd-Elrahman, A., 2018. Deep convolutional neural network training enrichment using multi-view object-based analysis of unmanned aerial systems imagery for wetlands classification. ISPRS J. Photogramm. Remote Sens. 139, 154–170.

Liu, Y., Ren, Q., Geng, J., Ding, M., Li, J., 2018. Efficient patch-wise semantic segmentation for large-scale remote sensing images. Sensors 18 (10), 3232.

Mboga, N., Persello, C., Bergado, J., Stein, A., 2017. Detection of informal settlements from VHR Images using convolutional neural networks. Remote Sens. 9 (11), 1106.

Meng, X., Currit, N., Zhao, K., 2010. Ground filtering algorithms for airborne LiDAR data: a review of critical issues. Remote Sens. 2 (3), 833–860.

Polat, N., Uysal, M., 2018. An experimental analysis of digital elevation models generated with lidar data and UAV photogrammetry. J. Indian Soc. Remote Sens. 46 (7), 1135–1142.

Rizaldy, A., Persello, C., Gevaert, C.M., Oude Elberink, S.J., 2018. Fully convolutional networks for ground classification from lidar point clouds. ISPRS annals of the photogrammetry. Remote Sens. Spatial Inf. Sci. 4 (2), 231–238.

Rusnák, M., Sládek, J., Kidová, A., Lehotský, M., 2018. Template for high-resolution river landscape mapping using UAV technology. Measurement 115, 139–151.

Salach, A., Bakuła, K., Pilarska, M., Ostrowski, W., Górski, K., Kurczyński, Z., 2018. Accuracy assessment of point clouds from LiDAR and dense image matching acquired using the UAV platform for DTM Creation. ISPRS Int. J. Geo-Inf. 7 (9), 342.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (4), 640–651.

Sun, G., Huang, H., Zhang, A., Li, F., Zhao, H., Fu, H., 2019. Fusion of multiscale convolutional neural networks for building extraction in very high-resolution images. Remote Sens. 11 (3), 227.

Tan, Y., Wang, S., Xu, B., Zhang, J., 2018. An improved progressive morphological filter for UAV-based photogrammetric point clouds in river bank monitoring. ISPRS J. Photogramm. Remote Sens. 146, 421–429.

Yilmaz, C.S., Güngör, O., 2018. Comparison of the performances of ground filtering algorithms and DTM generation from a UAV-based point cloud. Geocarto Int. 33 (5), 522–537.

Yilmaz, V., Konakoglu, B., Serifoglu, C., Güngör, O., Gökalp, E., 2018a. Image classification-based ground filtering of point clouds extracted from UAV-based aerial photos. Geocarto Int. 33 (3), 310–320.

Yilmaz, C.S., Yilmaz, V., Güngör, O., 2018b. Investigating the performances of commercial and non-commercial software for ground filtering of UAV-based point clouds. Int. J. Remote Sens. 39 (15–16), 5016–5042.

Zeybek, M., Sanlıoglu, I., 2019. Point cloud filtering on UAV based point cloud. Measurement 133, 99–111.

Zhang, Z., Gerke, M., Vosselman, G., Yang, M.Y., 2018. Filtering photogrammetric point clouds using standard lidar filters towards DTM generation. ISPRS annals of the photogrammetry. Remote Sens. Spatial Inform. Sci. 4 (2), 319–326.