remote sensing

MDPI

Article

IMG2nDSM: Height Estimation from Single Airborne RGB Images with Deep Learning

Savvas Karatsiolis [1,*], Andreas Kamilaris [2,3], Ian Cole [4,5]

[1] CYENS Center of Excellence, Nicosia, Cyprus; s.karatsiolis@cyens.org.cy
[2] CYENS Center of Excellence, Nicosia, Cyprus; a.kamilaris@cyens.org.cy
[3] Dept. of Computer Science, University of Twente, Enschede, The Netherlands; a.kamilaris@utwente.nl
[4] CYENS Center of Excellence, Nicosia, Cyprus; i.cole@cyens.org.cy
[5] University of Cyprus, Nicosia, Cyprus; cole.ian@ucy.ac.cy

* Correspondence: s.karatsiolis@cyens.org.cy

Abstract: Estimating the height of buildings and vegetation in single aerial images is a challenging problem. A task-focused Deep Learning (DL) model which combines architectural features from successful DL models (U-NET and Residual Networks) and learns the mapping from single aerial imagery to a normalized Digital Surface Model (nDSM) is proposed. The model is trained on aerial images whose corresponding DSM and Digital Terrain Maps (DTM) are available and is then used to infer the nDSM of images with no elevation information. The model is evaluated with a dataset covering a large area of Manchester, UK, as well as the 2018 IEEE GRSS Data Fusion Contest LiDAR dataset. The results suggest that the proposed DL architecture is suitable for the task and surpasses other state-of-the-art DL approaches by a large margin.

Keywords: building height estimation; deep learning; digital surface model; object shadow; aerial imagery; LiDAR; convolutional neural networks; remote sensing; digital elevation models.

1. Introduction

Aerial images are widely used in geographic information systems (GIS) for a plethora of interesting tasks, such as: urban monitoring and planning [1–3]; agricultural development [4]; landscape change detection [5–7]; disaster mitigation planning and recovery [8]; as well as aviation [9,10]. However, these images are predominantly two-dimensional (2D) and constitute a poor source of three-dimensional (3D) information, hindering adequate understanding of vertical geometric shapes and relations within a scene. Ancillary 3D information improves the performance of many GIS tasks and facilitates the development of tasks that require geometric analysis of the scene, such as digital twins for smart cities [11] and forest mapping [12]. In such cases, the most popular type of this complementary 3D information is the form of a Digital Surface Model (DSM). The DSM is often obtained with a Light Detection and Ranging Laser Scanner (LiDAR); or an Interferometric Synthetic-Aperture Radar (InSAR), a Structure from Motion (SfM) methodology [13]; or by using stereo image pairs [14]. Structure from motion is a technique for estimating 3D structures from 2D image sequences. The main disadvantages of SfM include the possible deformation of the modeled topography, its over-smoothing effect, the necessity for optimal conditions during data acquisition and the requirement of a ground control point [15]. Like SfM, DSM estimation by stereo image pairs requires difficult and sophisticated acquisition techniques, precise image pairing, and relies on triangulation from pairs of consecutive views. LiDAR sensors can provide accurate height estimations and have recently become affordable [16]. However, LiDAR sensors suffer from poor performance when complex reflective and refractive bodies (such as water) are present and can return irrational values in such cases, especially where there are multiple light paths from reflections in a complex scene. Despite these specific performance issues, LiDAR remains a commonly used technology for the acquisition of DSMs.

A substantial non-technical disadvantage of obtaining aerial DSMs and DTMs of large areas with LiDAR technology is the high cost of the required flight mission. This cost factor can preclude LiDAR acquisition as economically prohibitive. Therefore, elevation estimation from an input image is a compelling idea. However, height estimation from a single image, as with monocular vision in general, is an ill-posed problem: there are infinite possible DSMs that may correspond to a single image. This means that multiple height configurations of a scene may have the same apparent airborne image [17] due to the dimensionality reduction in the mapping of an RGB image to a one-channel height map. Moreover, airborne images frequently pose scale ambiguities that make the inference of geometric relations non-trivial. Consequently, mapping 2D pixel intensities to real-world height values is a challenging task.

In contrast to the remote sensing research community, the computer vision (CV) community has shown a significant interest in depth estimation from a single image. Depth perception is known to improve computer vision tasks such as: semantic segmentation [18,19], human pose estimation [20], and image recognition [21–23]; analogous to height estimation improving remote sensing tasks. Prior to the successful application of Deep Learning (DL) for depth prediction, methods such as stereo vision pairing, SfM, and various feature transfer strategies [24] have been used for the task. All these methods require expensive and precise data (pre)processing to deliver quality results. Contrastingly, DL simplifies the process while achieving better performance. Eigen et al. [25] use a multiscale architecture to predict the depth map of a single image. The first component of the architecture is based on the AlexNet DL architecture [23] and produces a coarse depth estimation refined by an additional processing stage. Laina et al. [26] introduce residual blocks into their DL model and use the reverse Huber loss for optimizing the depth prediction. Alhashmin and Wonka [27] use transfer learning from a DenseNet model [28] pretrained on ImageNet [29], which they connect to a decoder using multiple skip connections. By applying multiple losses between the ground truth and the prediction the state-of-the-art in-depth estimation from a single image was achieved. The interest of the CV community on depth estimation originates from the need for better navigation for autonomous agents, space geometry perception and scene understanding, especially in the research fields of robotics and autonomous vehicles. Specifically, regarding monocular depth estimation, AdaBins [30] achieved state-of-the-art performance on KITTI [31] and NYU-Depth v2 [32] datasets by using adaptive bins for depth estimation. Mahjourian et al. [33] use a feature-metric loss for self-supervised learning of depth and ego-motion. Koutilya et al. [34] combine synthetic and real data for unsupervised geometry estimation through a generative adversarial network (GAN) [35] called SharinGAN, which maps both real and synthetic images to a shared domain. SharinGAN achieves state-of-the-art performance on the KITTI dataset.

The main approaches used by researchers in aerial image height estimation based on DL involve: a) training with additional data, b) tackling auxiliary tasks in parallel to depth estimation, c) using deeper models with skip connections between layers, and d) using generative models (such as GANs) with conditional settings.

Alidoost et al. [36] apply a knowledge-based 3D building reconstruction by incorporating additional structural information regarding the buildings in the image, like lines from structure outlines. Mou and Zhu [37] propose an encoder-decoder convolutional architecture called IM2HEIGT that uses a single but provenly functional skip connection from the first residual block to the second last block. They argue that both the use of the residual blocks and the skip connection contribute significantly to the model performance. The advantages of using residual blocks and skip connections are also highlighted in the works of Amirkolaee and Arefi [38] and Liu et al. [17], who also use an encoder-decoder architecture for their IM2ELEVATION model. Liu et al. additionally apply data preprocessing and registration based on mutual information between the optical image and the DSM.

Furthermore, multi-task training proves to be beneficial, especially when height estimation is combined with image segmentation. Srivastava et al. [39] propose joint height estimation and semantic labeling of monocular aerial images with convolutional neural networks (CNNs). Carvalho et al. [40] use multi-task learning for both the height prediction and the semantics of aerial images. A different approach to the height estimation problem uses a generative model that produces the height map of an aerial image given the image as input. This strategy employs the conditional setting of the GAN and performs image-to-image translation, i.e., the model translates an aerial image to a DSM. Ghamisi and Yokoya [41] use this exact approach for their IMG2DSM model. Similarly, Panagiotou et al. [42] estimate the Digital Elevation Models (DEMs) of aerial images.

2. Materials and Methods

This section discusses the datasets used to train and evaluate the model, the technical aspects of the methods and the techniques used in the Deep Learning model. The model's architecture is also presented along with the task-specific design features that make it appropriate for height prediction. For further information on neural networks and Deep Learning, please refer to [43,44].

2.1 Datasets and data pre-processing

Two relatively large datasets are used to develop and evaluate the proposed height prediction model, namely: a Manchester area dataset, compiled by the authors; and an IEEE GRSS data fusion contest dataset. The focus of the Manchester area dataset is on estimating the height of buildings, while the focus of the IEEE GRSS data fusion contest dataset is on estimating the height of all objects in the images.

2.1.1 Manchester Area Dataset

The first dataset used to train the model comprises images and LiDAR DEMs (DSMs and DTMs); all from the Trafford area of Manchester, UK. The aerial photography is from Digimap [45] (operated by EDINA [46]). Both the LiDAR DEMs and the RGB images are geospatially aligned according to the Ordnance Survey National Grid reference system, a system using a transverse Mercator projection with a straight-line grid system built over the Airy 1830 ellipsoid. The reference grid system choice is essentially arbitrary as the model uses input image sections small enough for alignment deviations to be insignificant. Furthermore, as the model is trained to build DEMs from the input images, model DEM outputs will be aligned to the image pixel locations. The LiDAR data belongs to the UK Environment Agency [47]. It covers approximately 130 km^2 comprising roughly 8000 buildings. The RGB images have a resolution of 0.25 m by 0.25 m and the LiDAR resolution is 1 m by 1 m. The RGB images and the LiDAR maps were acquired at different dates; hence there are data inconsistencies resulting from new constructions or demolished buildings. Such inconsistencies constitute a barrier to the training of a DL model yet are representative of the real-world problem, especially given that many wide-area LiDAR datasets are compiled as composite images from LiDAR flights on multiple dates. Due to the low LiDAR resolution, this dataset is not appropriate for estimating the height of vegetation; thus, the analysis focuses only on buildings. Since segmentation labels for differentiation of what is vegetation and what is not are not available, a threshold height value of 1.5 m has been used to distinguish buildings. This approach occludes low vegetation and cars, which is desirable since the cars are mobile objects and thus the source of additional inconsistency. Furthermore, vegetation is a highly variable entity that is easily removed from the environment and is not necessary for many applications. The model is trained with the RGB aerial images as input and the normalized DSMs (nDSM = DSM-DTM) as target. nDSMs ignore the altitude information of the terrain and concentrate on the height of the objects. Figure 1 shows examples of different areas from the Manchester dataset and their corresponding ground truth nDSMs. Figure 2 shows the DTM and DSM of Figure 1 (bottom image) and demonstrates some flaws in the specific dataset.

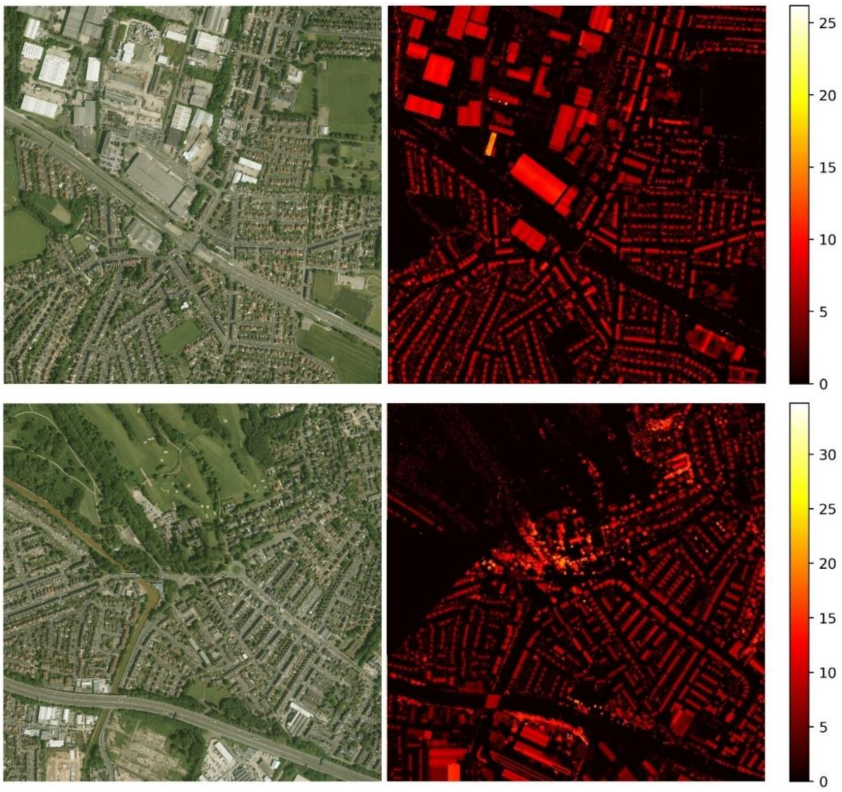


Figure 1. Aerial images from the first dataset (left), the corresponding nDSMs in heat-map format (middle) and the colorbars indicating the color-coding of the nDSM in meters (right). The aerial images on the left of each pair have a size of 4000 × 4000, while the size of the nDSMs is 1000 × 1000. nDSMs are presented at the same size as the aerial images for demonstration reasons. The Figure is best seen in color.

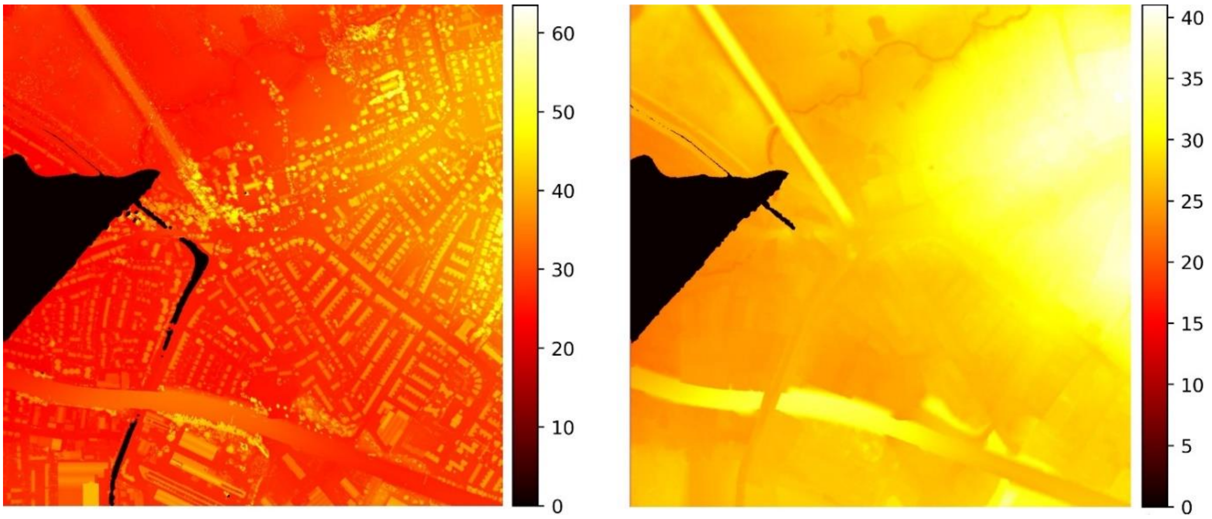


Figure 2. The DSM (left) and the DTM (right) corresponding to the bottom aerial image of Figure 1. The colorbar for each heat-map indicates the color-coding of the DEMs in meters above sea level. Both heat maps have several undetermined or irrational (extremely high or low) values shown in black color. Notably, some of these unexpected values in the DSM map (left) correspond to a river which illustrates a well-known problem of LiDAR measurements near highly reflective and refractive surfaces with multiple light paths. Such erroneous values raise significant problems regarding the training of the model. Thus, they are detected during data preprocessing and excluded from the training data (see Section 2.4). They are also excluded from the validation and test data to avoid inaccurate performance evaluation. Overall, these values roughly comprise 10% of the dataset, but lead to a larger amount of discarded data since any candidate patch containing even a pixel of undetermined or irrational value is excluded from the training pipeline. This figure is best seen in color.

2.1.2 IEEE GRSS Data Fusion Contest Dataset

The Data Fusion 2018 Contest Dataset (DFC2018) [48,49] is part of a set of community data provided by the IEEE Geoscience and Remote Sensing Society (GRSS). The Multispectral LiDAR Classification Challenge data was used herein. The RGB images in the dataset have a $0.05\,m$ by $0.05\,m$ resolution, and the LiDAR resolution is $0.5\,m$ by $0.5\,m$. The data belongs to a $4.172 \times 1.202\,km^2$ area. Given the higher resolution of the RGB images, this dataset is more suitable for estimating vegetation height than the Manchester area dataset. Figure 3 shows example pairs of RGB images and their corresponding nDSMs from this dataset. The ratio between the resolution of the RGB images and the resolution of their corresponding LiDAR scans affects the design of the depth-predicting model. Like in the Manchester area dataset, the model must handle the resolution difference between its input and its output and predict a nDSM that is several times smaller than the RGB image. Since the two datasets have different resolutions between the RGB images and the LiDAR scans, the same model cannot be used for both cases. Consequently, the models differ in their input/output size and the resolution reduction they must apply. For the most part, the models used for the two datasets are very similar, but slight architectural modifications are applied to cope with the resolution difference.

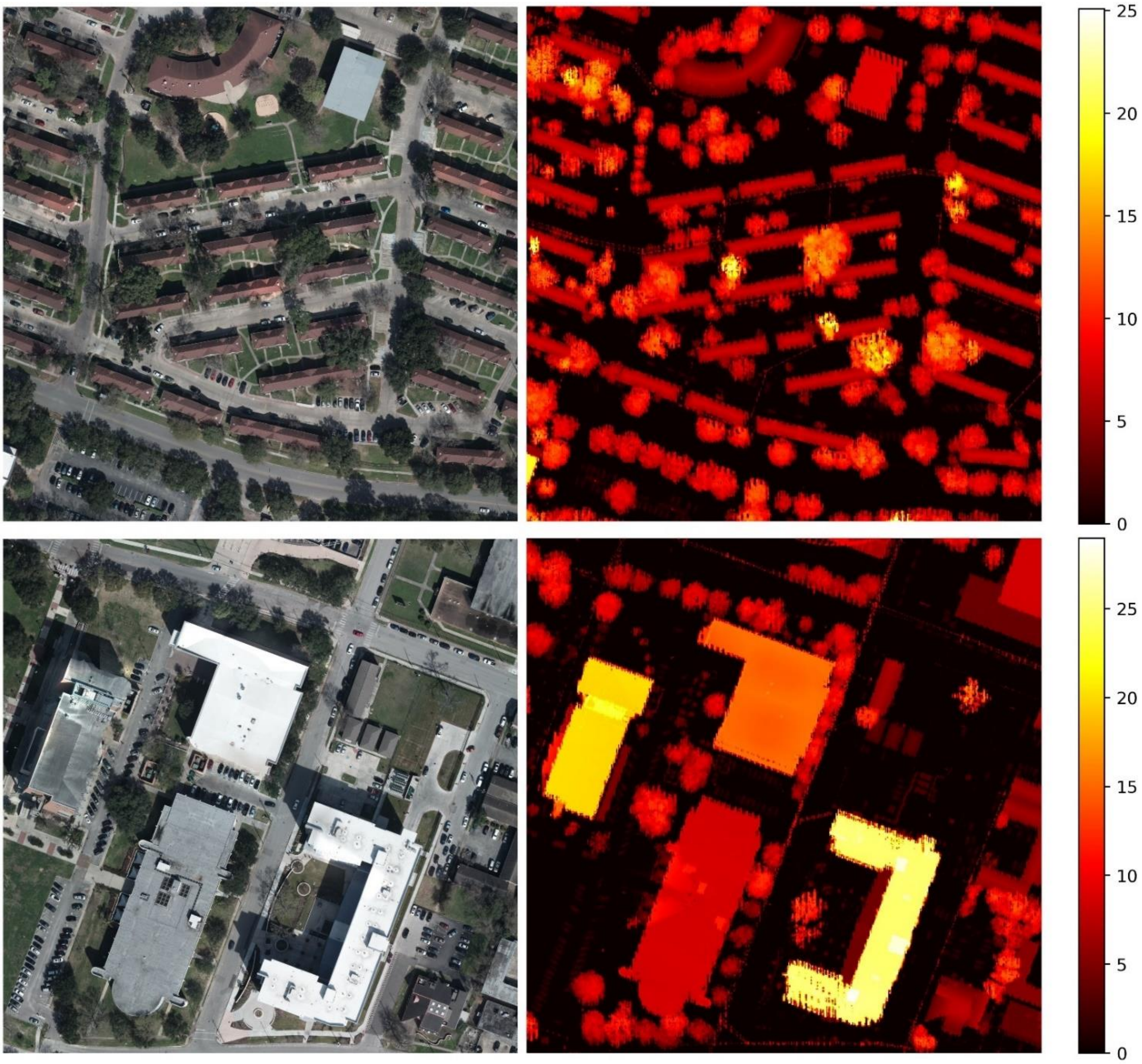


Figure 3. Aerial images from the IEEE GRSS Data Fusion Contest (second dataset), the corresponding nDSMs and the colorbars of the heat maps indicating the color coding in meters. The RGB images on the left of each pair have a size of 5000×5000 pixels, while the size of the nDSMs is 500×500 pixels. The heat maps are shown at the same size as the aerial images for demonstration reasons. This figure is best seen in color.

2.2 Data Preparation

The model operating on the Manchester area dataset uses image patches of size $256 \times 256 \times 3$, while the model operating on the DFC2018 dataset uses image patches of size $520 \times 520 \times 3$. The specific input sizes determine that the former model outputs a map of size 64×64 and the latter has an output of 52×52 since the resolution ratios of the image to LiDAR datasets are *4* and *10*, respectively: every *4* pixels in one aerial image of the Manchester dataset correspond to *1* pixel in the respective nDSM and *10* pixels in one aerial image of the DFC2018 dataset correspond to *1* pixel in the respective nDSM. The 64×64 and the 52×52 sizes of the models' outputs offer a compromise between computational and memory requirements during training and sufficient scenery area consideration when calculating a nDSM, i.e. basing the estimation on several neighboring structures in the input image for achieving better accuracy. Various output sizes were experimented with and it was discovered that predicting larger nDSMs tends to achieve slightly better accuracy at the cost of increased memory usage and computational requirements.

2.3 Model Description

In this section, technical aspects of the methods and techniques used in the proposed DL model are discussed. The model's architecture is presented, together with the task-specific design features that make it appropriate for depth prediction. The authors have called the model presented herein 'IMG2nDSM' because it maps an aerial photography image to a nDSM.

The proposed architecture shares some similarity with semantic segmentation models, where the model must predict the label of each pixel in an image and thus partition it into segments. The segmentation may have the size of the input image or a scaled-down size. In this study, instead of labels, the real values corresponding to the elevation of each pixel are predicted in a down-scaled version of each RGB image. As with the semantic segmentation task, several DL models are suitable for learning the task of predicting the nDSM values. A popular DL architecture, the U-Net model [50], was chosen due to its efficiency and effectiveness on tasks based on pixel-level manipulations like semantic segmentation [51–53]; and as its architectural scheme easily combines with other DL techniques to introduce task-specific enhancements.

The U-Net architecture has been implemented with residual blocks both in the encoder and the decoder mechanisms. Specifically, three types of residual blocks are used, as shown in Figure 4:

- A typical residual block (RBLK),
- A down-sampling residual block (DRBLK),
- An up-sampling residual block (URBLK).

A typical residual block contains two convolutional layers at the data path and a convolutional layer with a kernel size of one at the residual connection path. The down-sampling residual block differs in the stride used at the first convolutional layer and the skip connection. Using a higher stride at these convolutions, the previous feature maps are downscaled by a factor s (here, $s = 2$) at the first convolutional layer of the block and the skip connection, which results in smaller feature maps. The up-sampling residual block uses sub-pixel convolutional up-scaling [54] in the first layer of the block. Sub-pixel upscaling is performed in two steps, with the first step calculating a representation comprising feature maps of size $h \times w \times s^2 c$, where s is the up-scaling factor, and $h \times w \times c$ is the size of the input feature maps. The second step of the process applies a *reshape* operation on the feature maps and produces a representation containing feature maps of size $2h \times 2w \times c$. The skip connection of the up-scaling residual block also applies sub-pixel up-scaling. The detailed architecture of the residual blocks is shown in Figure 4.

A very similar model for both datasets is used, with minor changes regarding the input patch size and the size of the output prediction. The Manchester area dataset has an RGB image over nDSM resolution ratio equal to 4, so the neural network dealing with this dataset reduces the input size from $256 \times 256 \times 4$ to output size 64×64. On the other hand, the DFC2018 dataset has an RGB image over depth map resolution ratio equal to 10, so the neural network dealing with this dataset reduces the input size from 520 × 520 × 4 to output size 52 × 52.

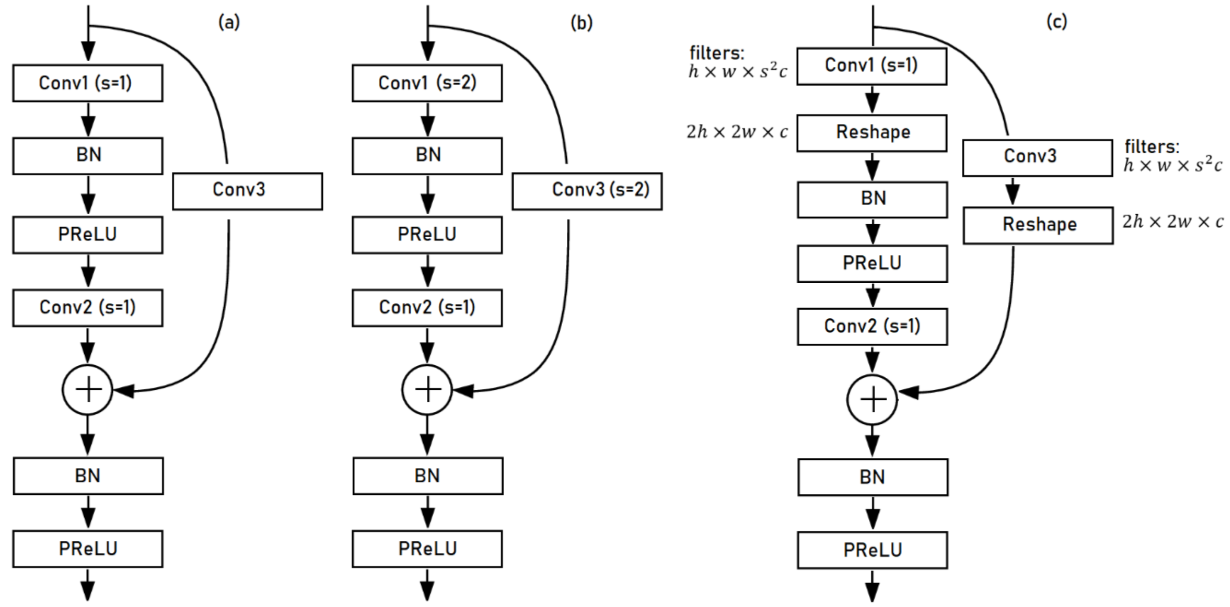


Figure 4. The architecture of the three types of residual blocks used in the proposed models: (a) The typical residual block (RBLK) (b) The down-sampling residual block (DRBLK) uses a stride of two at both the first convolutional layer and the skip connection. (c) The up-sampling residual block (URBLK) uses sub-pixel up-scaling at the first convolution and the skip connection. BN stands for batch normalization [55], PReLU for parametric ReLU, and s is the stride of the convolutional layer.

The input/output sizes of the models are a compromise between having a manageable input size in terms of computational requirements and having a sufficient output map size and performance. The few differences between the two models are necessary for applying the different reduction factors between the input and the output of the two datasets as dictated by the RGB/nDSM resolution ratio of each dataset. Specifically, the number of layers, the number of channels and the kernel sizes of the convolutional layers of the model trained on the DFC2018 dataset are different from the ones used in the model trained on the Manchester dataset. This is due to the requirement of a larger resolution reduction. However, these changes are applied at the initial and the last layers of the model to maintain the U-NET scheme unaltered. Figures 5 and 6 show the detailed architectures of both models and the size of the feature maps after each layer.

The model dealing with the Manchester area dataset has 164 layers (including concatenation layers and residual blocks' addition layers) and consists of approximately 125 M trainable parameters. The model dealing with the DFC2018 dataset has 186 layers (including concatenation layers and residual blocks' addition layers) but has fewer parameters to handle the higher memory requirements during training due to the larger input size. Precisely, it consists of 104 M trainable parameters. The only differences with the model used for the Manchester area dataset are a) the addition of some convolutional layers with "valid" padding to achieve the correct output size and b) the reduction of the parameters of the convolutional layers.

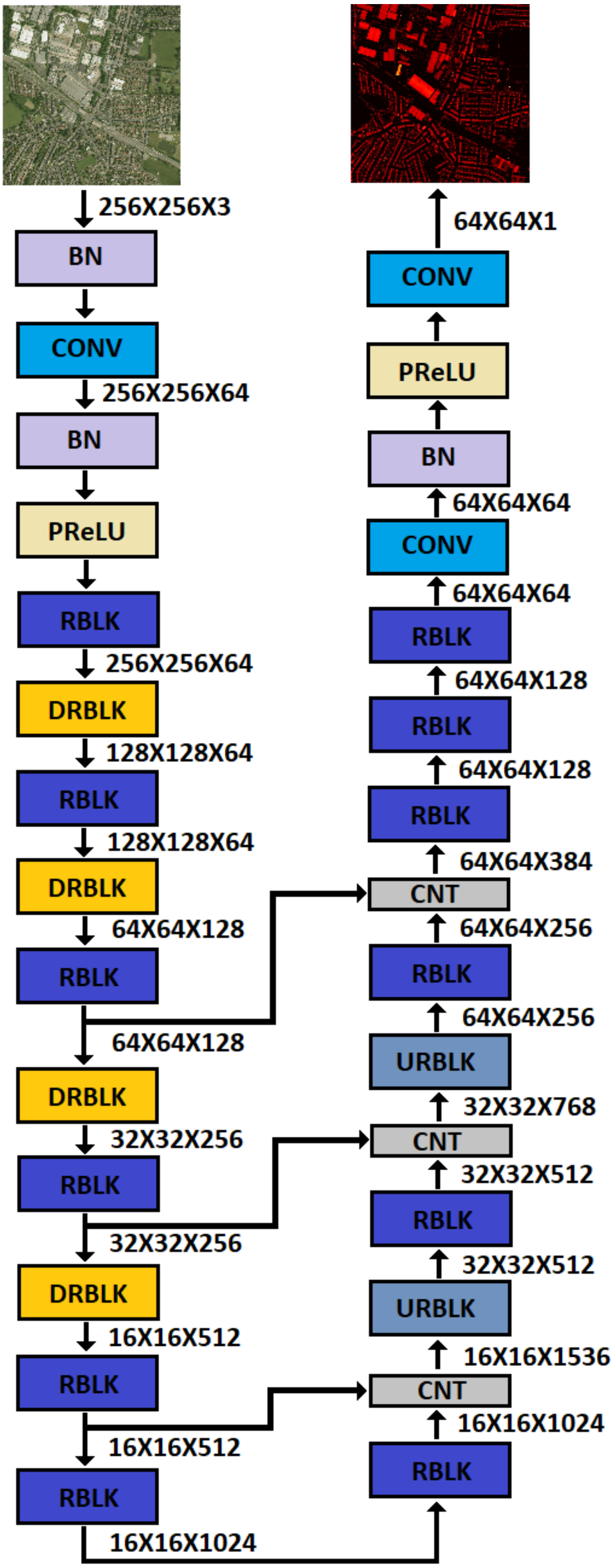
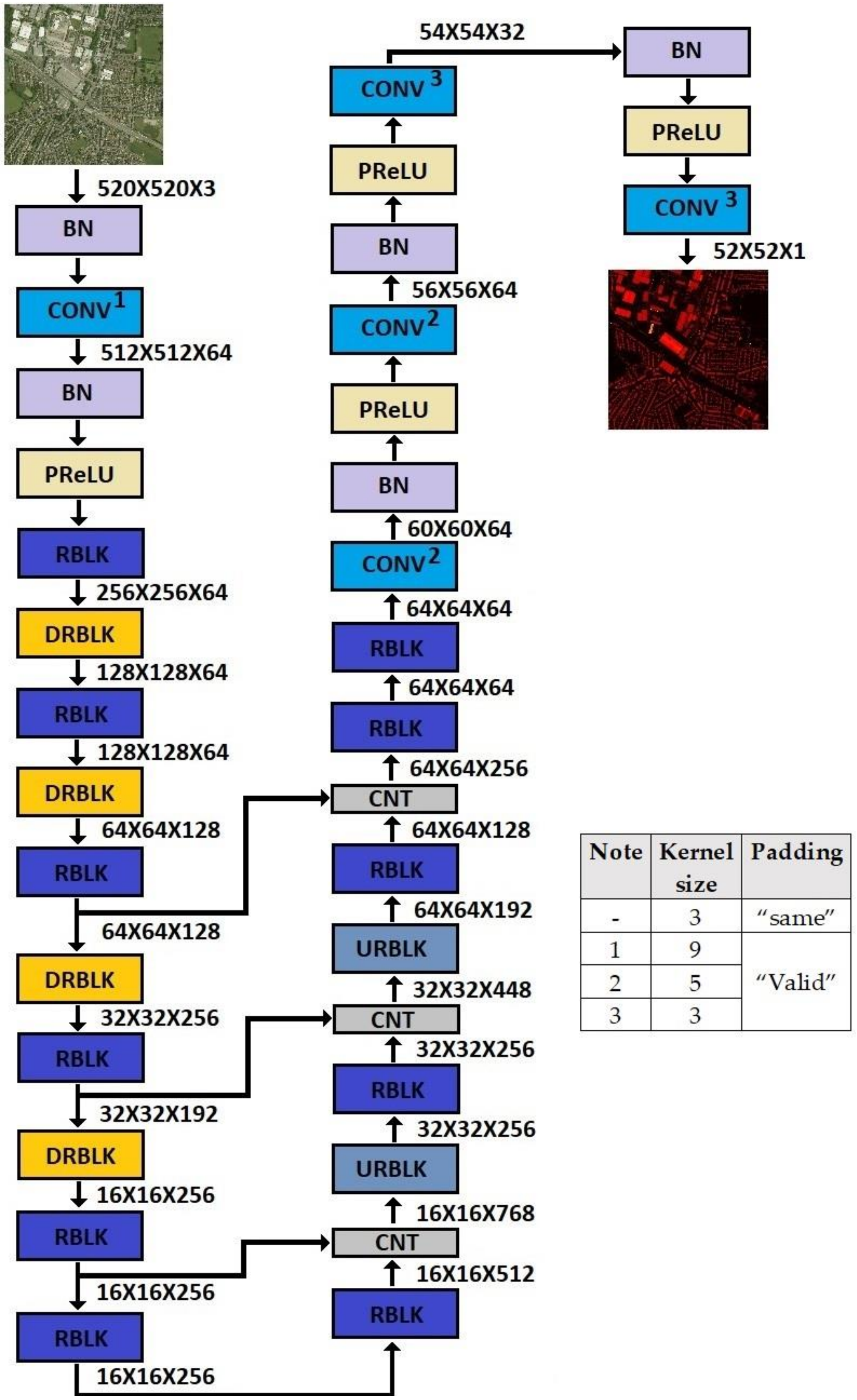


Figure 5. The architecture of the model trained with the Manchester dataset. All convolutional layers use kernel size 3 and "same" padding. *BN* represents a Batch Normalization layer and *CNT* a Concatenation layer.



Note	Kernel size	Padding
-	3	"same"
1	9	"Valid"
2	5	
3	3	

Figure 6. The architecture of the model trained with the DFC2018 dataset. Compared to the model trained with the Manchester dataset, the kernel sizes of certain convolutional layers are increased and their padding is changed from "same" to "valid" as indicated by the figure notes. Additionally, some convolutional layers are introduced at the end of the model. These modifications aim at applying a reduction factor of *10* between the input and the output of the model to match the resolution ratio between the aerial images and the nDSMs in the DFC2018 dataset. *BN* represents a Batch Normalization layer and *CNT* a Concatenation layer.

2.4 Training details

Simple augmentations are applied to the patches during training: rotations of 90, 180 and 270 degrees, small-value color shifting and contrast variations. Patches where the elevation maps contain incomplete or extreme elevation values ($> 100m$) are ignored in both datasets. Moreover, specifically for the Manchester area dataset, small elevation values ($< 1.5\ m$) were replaced with zeros to prevent the model from considering non-stationary objects and low vegetation. This pre-processing is important in part because of the time difference between the acquisition of the RGB images and the LiDAR point clouds, which results in inconsistencies between the images and the elevation maps due to the presence of mobile objects like cars in the viewing field of either of the two sensors (RGB or LiDAR). The time of acquisition inconsistency in the Manchester area dataset also introduces inconsistencies in vegetation height, and occasionally, in building heights (demolished buildings or newly constructed buildings). Consequently, regarding the Manchester area dataset, the model is focused on predicting the elevation of human-built structures like houses, factories, and public buildings. The DFC2018 dataset has better resolution, and no inconsistencies have been observed. This fact facilitates the prediction of vegetation height as well; thus a threshold filter is not applied to the ground truth nDSMs for the DFC2018 dataset.

The models are trained with the Adam optimizer [56] and a learning rate of 1×10^{-4} which decreases by a factor of *10* each time the error plateaus for several iterations. Both datasets are randomly split into three sets each containing images of size 1000×1000 for the Manchester dataset and 5000×5000 for the DFC2018 dataset: a training set (70%), a validation set (15%), and a test set (15%). The validation set is used for hyper-parameter fitting and then merged with the training set for retraining the models. The test set is only used to report the models' performances. The models were trained for *5* different random dataset splits and the average of the performances on the test sets is reported. The pixel-wise Mean Absolute Error (MAE) is used between the ground-truth elevation maps and the predicted output as the loss function during training. Mean Squared Error (MSE) was also considered, but it was found that MAE performs slightly better probably because it does not penalize outliers as much as the MSE. The Root Mean Squared Error (RMSE) performance is also reported as an additional evaluation metric. All parameters are initialized with the He normal technique [57].

3. Results

The model presented herein achieves a MAE of 0.59 m and an RMSE of 1.4 m for the Manchester area dataset, as well as a MAE of 0.78 m and an RMSE of 1.63 m for the DFC2018 dataset. The lower error values on the first dataset most likely occur due to ignoring small nDSM values (<1.5 m), increasing the model accuracy in the prediction of buildings and human-made structures. The proposed architecture improves on the results of Carvallo et al. [40] and Liu et al. [17] by a significant margin (see Table 1), although a direct comparison cannot be accurate since all approaches use random data splits.

Table 1. Model's performance on the test set and comparison to other methods.

Method	MAE(m) ↓	RMSE(m) ↓
Manchester Area dataset [1]		
IMG2nDSM*	0.59	1.4
DFC2018 dataset [2]		
Carvallo et al. [40] (DSM)	1.47	3.05
Carvallo et al. [40] (DSM + semantic)	1.26	2.60
Liu et al. [17]	1.19	2.88
IMG2nDSM*	**0.78**	**1.63**

[1] 0.25 m/pixel RGB resolution, 1m/pixel LiDAR resolution, inconsistencies

[2] 0.05 m/pixel RGB resolution, 0.5m/pixel LiDAR resolution

*IMG2nDSM is the model presented in this work

3.1 Height prediction for the Manchester Area dataset

The estimated heights of areas in the Manchester Area test set are depicted in Figure 7. The estimations are shown in the form of heat maps for better visualization (i.e. providing a more precise display of the relative height values) and evaluation purposes. Since the model operates on patches of size 256 × 256 × 3, the RGB images are divided into several patches with overlapping regions of *16* pixels. Then, the model predicts a nDSM for each patch and, finally, the estimated maps are recombined to create the overall nDSM for the RGB image. During the recombination process, the outer *16* pixels of each predicted map are ignored to achieve a more natural blending and avoid artifacts.

Interestingly, the model avoids spiky estimations like the ones indicated with *note 1* in the images of Figure 7: ground truth LiDAR maps occasionally contain points of unnaturally high values compared to neighboring points that constitute false readings that occur for several reasons. These reasons relate mainly to the physical properties of the LiDAR sensor and the environmental conditions during data acquisition (see Section 1 for a discussion on this). Furthermore, some incorrect readings may have values that lie in the boundary of reasonable LiDAR values and are difficult to discriminate from incomplete readings with irrational values. Such spiky readings naturally occur in the training set too. Nevertheless, the model is not affected by such inconsistencies in the training set and its estimates corresponding to spiky measurements in the 'ground truth' data are closer to the actual ground truth (see Figure 7, *note 1*).

Moreover, the Manchester Area dataset contains several inconsistencies in regards to structures that are missing either from the RGB images or from the nDSM due to different acquisition times between the two data types. Such inconsistencies are shown in Figure 7 (indicated as *note 2*): In these cases, some structures present in the ground truth nDSM are missing from the RGB images; however, the model correctly predicts the corresponding regions containing the inconsistencies as undeveloped spaces. This behavior is, of course, desired and demonstrates effectively that the IMG2nDSM model presented herein is robust to false training instances. Furthermore, the results reveal some additional cases which indicate that the model is doing a good job estimating the height of buildings, surpassing the quality of the ground truth map. Notably with noisy data on specific structures with known forms. *Note 3* in Figure 7 demonstrates such a case where the ground truth map seems noisy, given the known form of the apex roof structure, while the estimation of the model is more detailed and smoother. This raises the point that, although the model performance is calculated against the LiDAR data as 'ground truth', it sometimes outperforms the LiDAR data and generates results closer to actual ground truth.

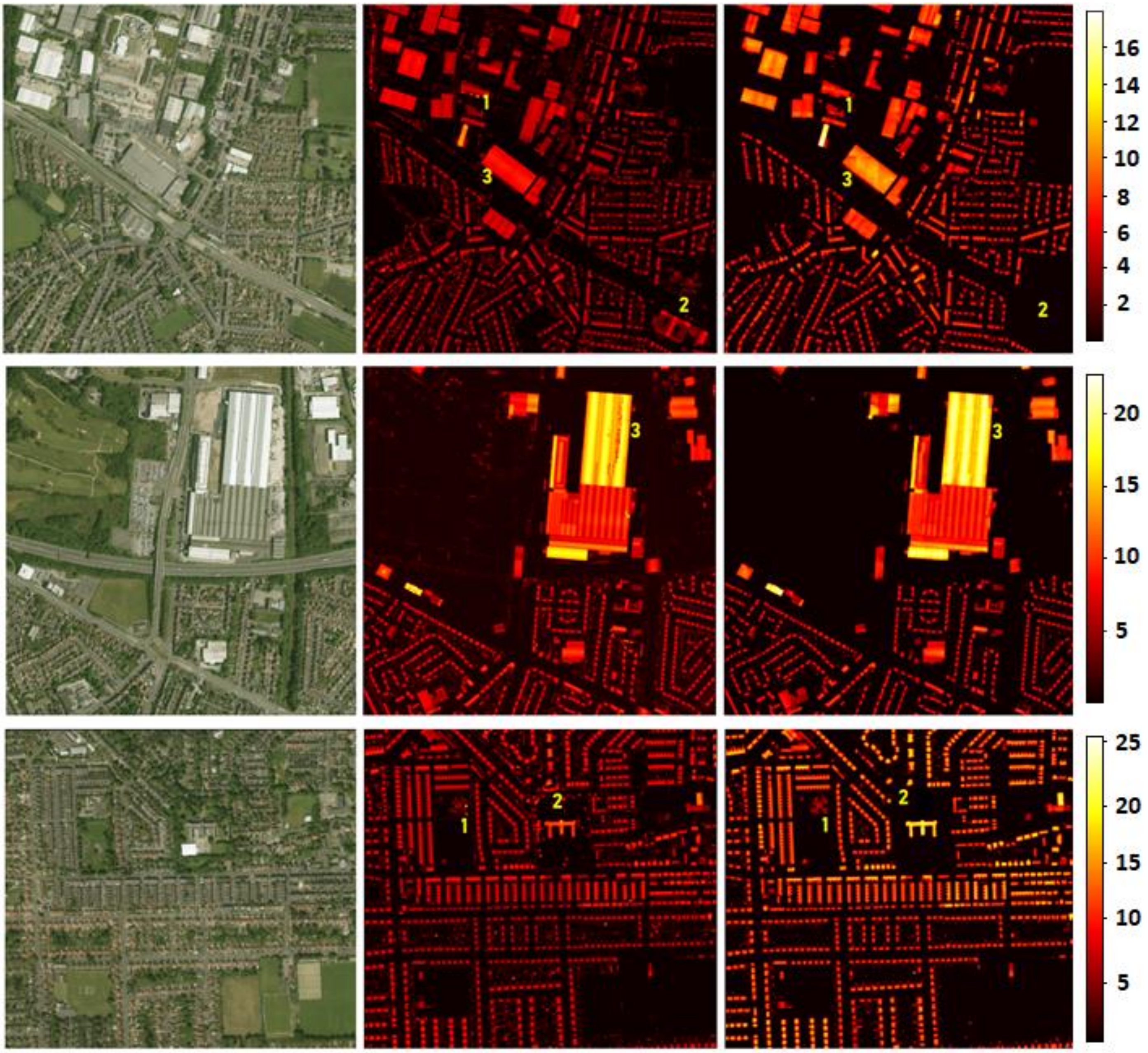


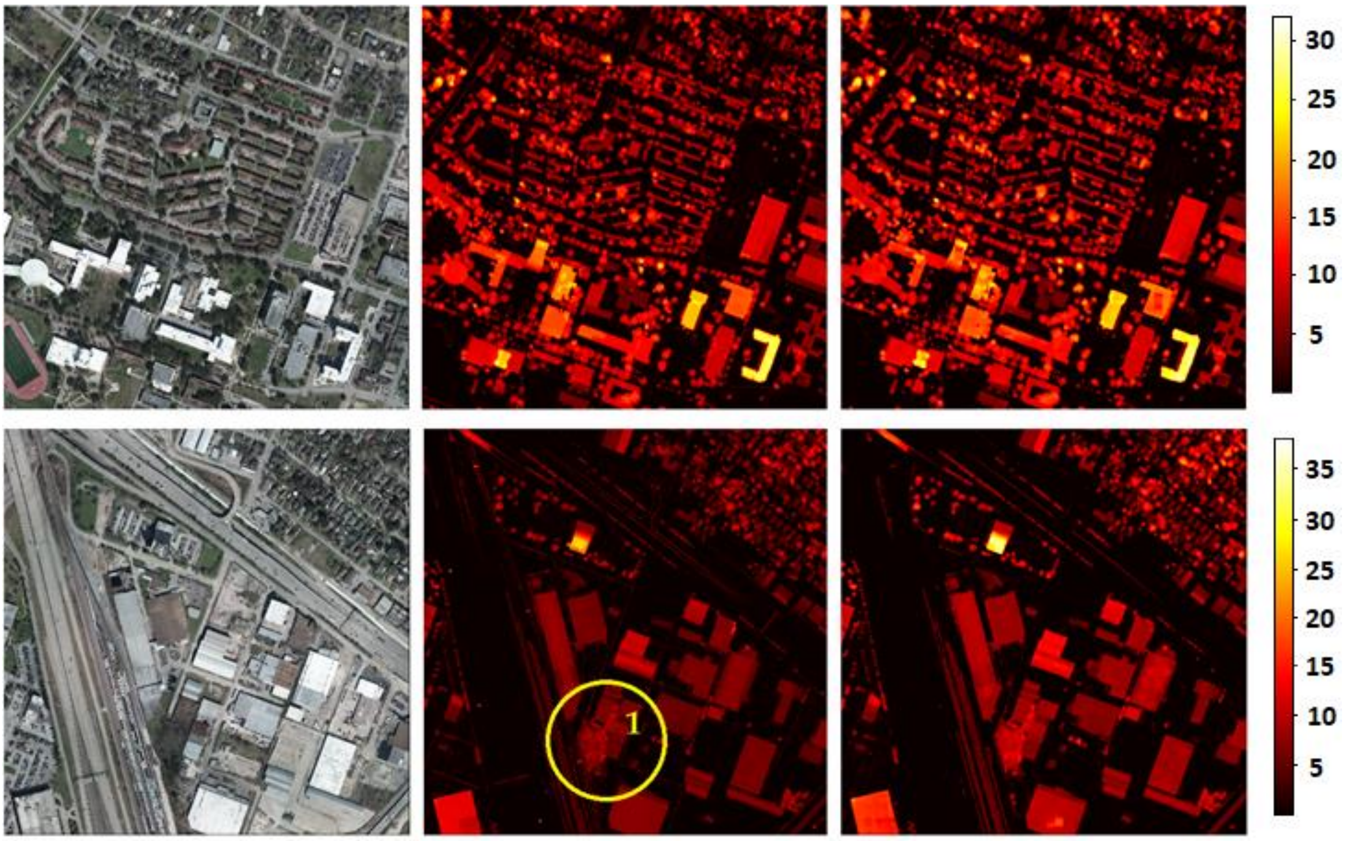
Figure 7. Left: RGB images of an area in the test set of the Manchester area dataset. Middle: The ground truth nDSMs. Right: The elevation heat maps as predicted by the model. *Note 1* shows cases of spurious points in the ground truth that the model correctly avoids estimating. *Note 2* shows occasional inconsistencies in the dataset due to different acquisition time of the RGB images and the LiDAR measurements. Although these inconsistencies are also evident in the training set, the model is robust to such problematic training instances. *Note 3* shows cases where the model produces better quality maps than the ground truth in terms of surface smoothness and level of detail as the LiDAR data contains noisy values.

3.2 Height prediction for the DFC2018 dataset 427

The estimated nDSMs of consecutive areas of the DFC2018 test set are illustrated in 428 Figure 8. As in the case of the Manchester Area test set, the RGB images are divided into 429 overlapping patches and the model predicts the nDSM for each of the patches. The only 430 difference is that the size of the patches for this dataset is $520 \times 520 \times 3$ pixels. The 431 estimated nDSMs are amalgamated, as described in section 3.1, to create the nDSM of the 432 entire area. 433

434

435



436

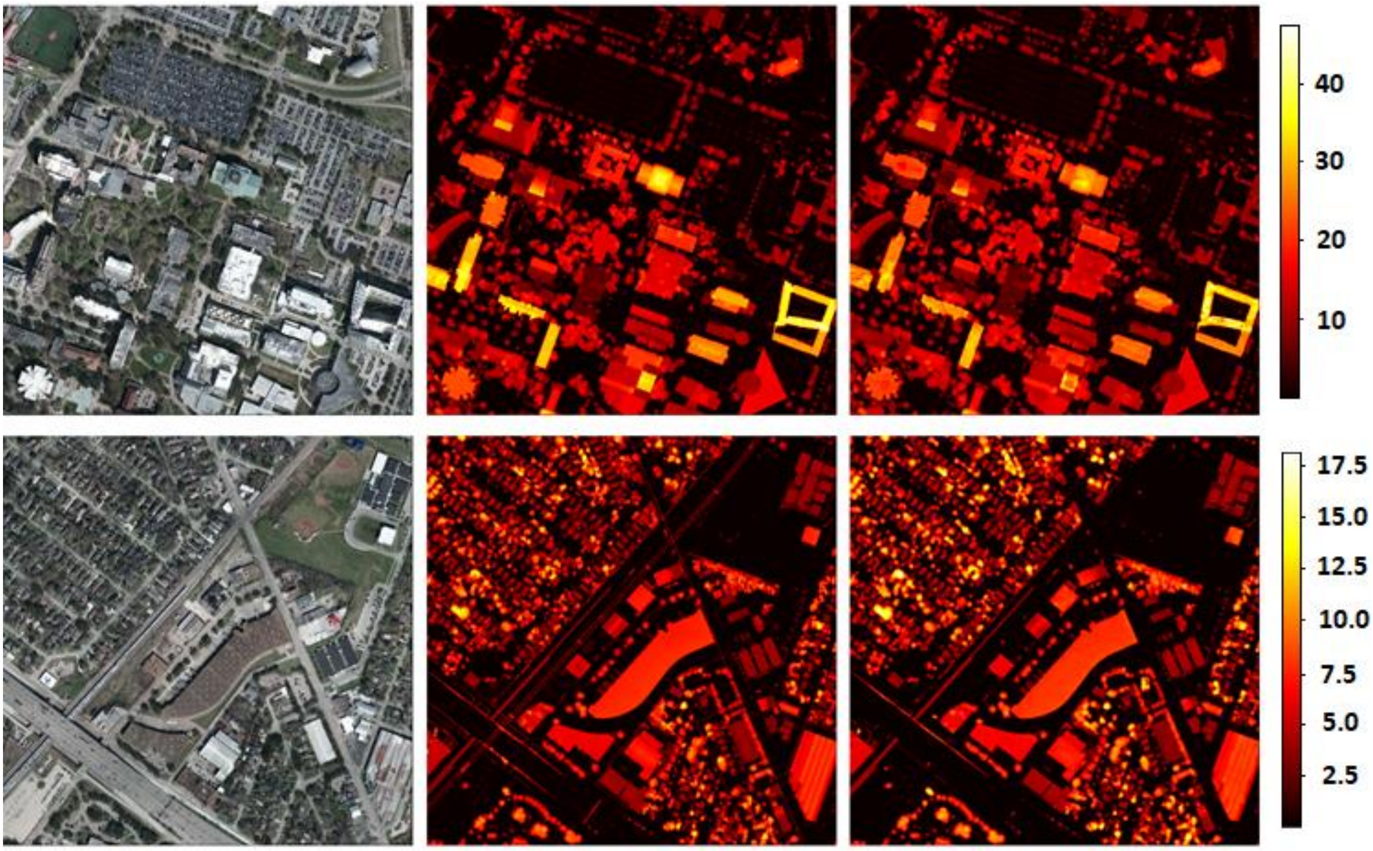


Figure 8. Left: RGB images from the DFC2018 test set. Middle: Ground truth nDSMs. Right: Model's height estimations. *Note 1* indicates an area that contains a group of trees and is magnified in Figure 9 to demonstrate how the model treats vegetation in the RGB images.

The predicted nDSMs look very similar to the ground truth. The higher resolution of the RGB images and the consistency between the RGB and the LiDAR measurements in terms of data acquisition time have a positive impact on the model's performance. For this dataset, the model can estimate vegetation height accurately. Regarding vegetation, the model is consistently overestimating the area covered by foliage, as it fills the space between the foliage. *Note 1* in the second row of Figure 8 (located at the ground truth nDSM) shows the height measurements for a group of trees. Figure 9 shows the magnification of that area, the magnified ground truth map and the model's height estimation; and demonstrates the tendency of the model to overestimate the volume of foliage. It is thought that this behavior contributes to the higher MAE that the model scores on the DFC2018 dataset compared to the better performance on the Manchester area dataset. As described above, the latter dataset has lower resolution and more inconsistencies, but the model training ignores vegetation and low standing objects to its favor. However, this behavior of the model with vegetation height estimation could be beneficial under some circumstances, such as projects that focus on tree counting, monitoring tree growth or tree coverage in an area [12].

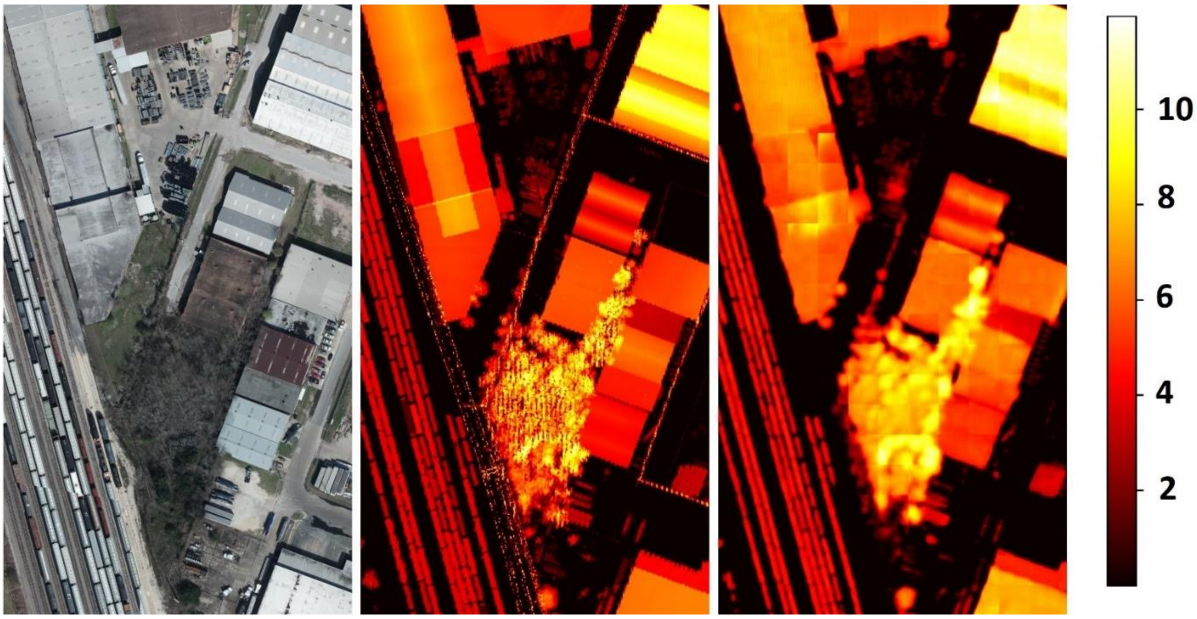


Figure 9. Magnification of the noted region (*Note 1*) in Figure 8. Left: The magnified RGB image. Middle: The Ground truth nDSM. Right: Model output. The model consistently overestimates foliage volume by filling the spaces between foliage with similar values to neighboring estimations.

3.3 Model Analysis

The very good results of the model, as shown in Table 1, result from its carefully designed architecture which was selected after many experiments and trials with various alternate options. The initial form of the model was a basic model having the U-NET scheme proposed in [50] with typical residual blocks (Figure 4.a) only, max-pooling (down-sampling) layers and nearest-neighbor interpolation (up-sampling) layers. Then, the basic model was improved upon by replacing individual architectural features with ones that improved performance. The modifications that affected performance the most are listed according to their contribution (higher contribution first):

- Use of the up-sampling residual block (URBLK) as shown in Figure 4.c instead of nearest-neighbor interpolation.
- Use of the down-sampling residual block (DRBLK) with strided convolutions as shown in Figure 4.b instead of max-pooling.
- Modification of the basic U-NET scheme so that the first two concatenation layers are applied before the up-sampling steps and not after them as originally proposed in [50].
- Use of "same" instead of "valid" padding in the U-NET scheme.
- Replace the ReLU activation functions with PReLUs.

The first three modifications (use of URBLKs, DRBLKs and the change in the concatenation layers positions) enabled the model to surpass the performance of other state-of-the-art works, while the remaining modifications (the use of "same" padding and PReLUs) further increased the performance gap in favor of the proposed model. Overall, the proposed model relies on a task-specific architecture for achieving good results in predicting the nDSM of a scene from an aerial image.

3.4 Limitations

Despite the overall promising results of the proposed model, there are still some cases where the model does not perform correctly. Buildings are well represented in both datasets, and thus, the model can predict their height with little error. The same applies to vegetation in the DFC2018 dataset. However, for objects that are rarely seen in the data (e.g., objects that are tall and thin simultaneously, such as light poles and telecommunication towers), the model sometimes fails to estimate their height correctly. In cases of very scarce objects, the model treats them as if they do not exist. Rarely seen tall objects that are not bulky or whose structure has empty interior spaces are tough for the model to assess. Examples of such failed cases are shown in Figure 10. The leading cause of the problem is the under-representation of these structures in the dataset. It can be mitigated by introducing more images containing these objects during training.

Although the model performs well, it is acknowledged that it has many parameters. However, predicting the nDSM of an individual patch is quite fast, especially when the model runs on a Graphical Processing Unit (GPU). Inferring the nDSM of a large area requires the splitting of the RGB image into several patches. Using a GPU, the estimation of the nDSMs of all patches is performed in parallel by processing a batch (or batches) of patches, taking advantage of the hardware and its parallel computing capabilities.

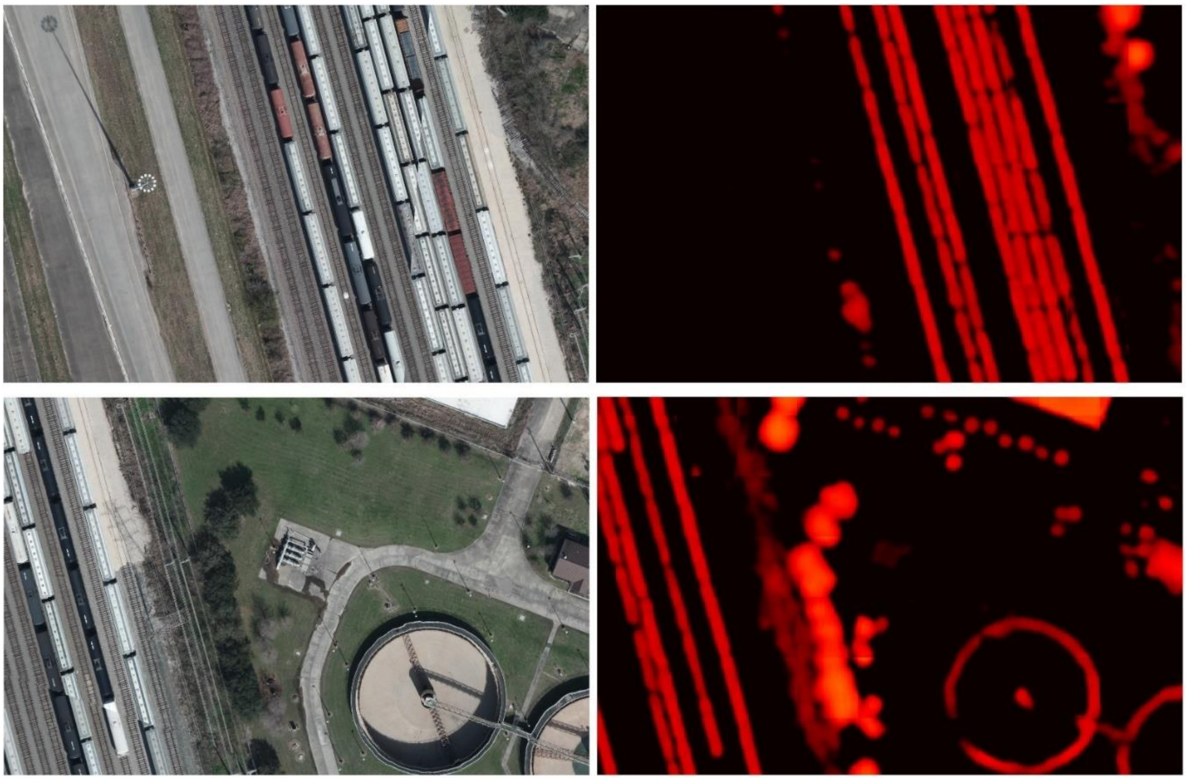


Figure 10. Sample failed cases where the model misses the presence of an object completely. The cases are magnified regions from the second RGB image (second row) of Figure 8. The top left image shows a very high pole standing on a highway (on the left of the train wagons) with a height of *30* meters (according to its LiDAR measurement). Despite the pole's long shadow, the model does not detect it. The left bottom magnified region contains a tall electric energy transmission tower (close and on the right of the train wagons) that is also not detected by the model.

4. Discussion

Obtaining the height of objects in aerial photography with hardware equipment can be costly, time-consuming, and require human expertise and sophisticated instruments. Furthermore, the acquisition techniques of such data are demanding and require specialized operators. On the other hand, inferring this data solely from aerial RGB images is easier, faster, and especially helpful if the availability of image pairs is limited for a certain terrain modeling task. Height estimation from aerial imagery is difficult due to its ill-posed nature, yet DL techniques offer a promising perspective towards providing adequate solutions to the task.

The authors propose a model, named IMG2nDSM, with a task-focused DL architecture that tackles the problem with very good results, which are better than state-of-the-art to date. The model has been tested on two different datasets: one with $0.25\ m$ by $0.25\ m$ image resolution, $1\ m$ LiDAR resolution, and different acquisition times (thus, it has spatial inconsistencies) and one with $0.05\ m$ by $0.05\ m$ image resolution and $0.5\ m$ LiDAR resolution. The first dataset (capturing lower resolution images) covers the Trafford area in Manchester, UK, while the second dataset is part of the 2018 IEEE GRSS Data Fusion Contest. The first dataset is used to estimate building heights only, while the second dataset is used to estimate both buildings and vegetation heights. Despite the inconsistencies encountered in the first dataset, the effectiveness of the model indicates its high robustness and ability to build domain knowledge without resorting to dataset memorization. This indication is also suggested by the fact that data curation or special prepossessing, besides data augmentation, was not employed.

The authors aspire to the idea that the possibility of deriving high-precision digital elevation models from RGB images without expensive equipment and high costs, will accelerate global efforts in various application domains that require geometric analysis of areas and scenes. Such domains include urban planning and digital twins for smart cities [11], tree growth monitoring and forest mapping [12], modeling ecological and hydrological dynamics [58], detecting farmland infrastructures [59], etc. Such low-cost estimation of building heights will allow policy-makers to understand the potential revenue of rooftop photovoltaics based on yearly access to sunshine [60] and law enforcement to verify whether urban/or rural infrastructures comply with local land registry legislation.

Finally, it is noted that the model experiences some cases of poor performance with tall-thin and generally under-represented objects. This issue can be solved by including more examples of such objects in the training images, which is an aspect of future work.

5. Conclusions

A DL model, IMG2nDSM, is proposed for inferring the heights of objects in single aerial RGB images. The model is trained with aerial images and their corresponding nDSMs acquired from LiDAR point clouds, but only the RGB images are required during inference. The model was tested on two datasets and its performance is significantly better than other state-of-the-art methods. Results prove that the model builds good domain knowledge and sometimes produces results that are better compared to the LiDAR data when assessing the ground truth scenario. The model's behavior regarding vegetation height estimation is also analyzed and some failed cases are reported.

Future research directions and model improvements include the reduction of failed cases for under-represented structures in the aerial imagery such as rarely seen specialpurpose structures with electronic devices, telecommunication towers and energy transmission towers. The value of the proposed methodology stems from its convenient and easy application and the fact that it only requires the RGB images during inference. Achieving the height estimation task from single RGB images without requiring LiDAR or any other information greatly reduces the cost; required effort and time; and the difficulties emerging from using complex data acquisition techniques or complex analytical computations.

Author Contributions: Conceptualization, S.K. and A.K; methodology, S.K.; software, S.K.; validation, S.K.; formal analysis, S.K.; investigation, S.K.; resources, S.K., I.C; data curation, S.K., I.C.; writing—original draft preparation, S.K.; writing—review and editing, A.K., I.C.; visualization, S.K.; supervision, A.K.; project administration, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739578 and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

Acknowledgments: The authors thank Chirag Prabhakar Padubidri for his consulting on handling the DFC2018 dataset.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wellmann, T.; Lausch, A.; Andersson, E.; Knapp, S.; Cortinovis, C.; Jache, J.; Scheuer, S.; Kremer, P.; Mascarenhas, A.; Kraemer, R.; et al. Remote Sensing in Urban Planning: Contributions towards Ecologically Sound Policies? *Landsc. Urban Plan.* **2020**, *204*, 103921, doi:https://doi.org/10.1016/j.landurbplan.2020.103921.
2. Bechtel, B. Recent Advances in Thermal Remote Sensing for Urban Planning and Management. In Proceedings of the Joint Urban Remote Sensing Event, JURSE 2015, Lausanne, Switzerland, March 30 - April 1, 2015; IEEE, 2015; pp. 1–4.
3. Zhu, Z.; Zhou, Y.; Seto, K.C.; Stokes, E.C.; Deng, C.; Pickett, S.T.A.; Taubenböck, H. Understanding an Urbanizing Planet: Strategic Directions for Remote Sensing. *Remote Sens. Environ.* **2019**, *228*, 164–182, doi:https://doi.org/10.1016/j.rse.2019.04.020.
4. Lesiv, M.; Schepaschenko, D.; Moltchanova, E.; Bun, R.; Dürauer, M.; Prishchepov, A. V; Schierhorn, F.; Estel, S.; Kuemmerle, T.; Alcántara, C.; et al. Spatial Distribution of Arable and Abandoned Land across Former Soviet Union Countries. *Sci. Data* **2018**, *5*, 1–12, doi:10.1038/sdata.2018.56.
5. Ma, L.; Li, M.; Blaschke, T.; Ma, X.; Tiede, D.; Cheng, L.; Chen, Z.; Chen, D. Object-Based Change Detection in Urban Areas: The Effects of Segmentation Strategy, Scale, and Feature Space on Unsupervised Methods. *Remote Sens.* **2016**, *8*, doi:10.3390/rs8090761.
6. Muro, J.; Canty, M.; Conradsen, K.; Hüttich, C.; Nielsen, A.A.; Skriver, H.; Remy, F.; Strauch, A.; Thonfeld, F.; Menz, G. Short-Term Change Detection in Wetlands Using Sentinel-1 Time Series. *Remote Sens.* **2016**, *8*, doi:10.3390/rs8100795.
7. Lyu, H.; Lu, H.; Mou, L. Learning a Transferable Change Rule from a Recurrent Neural Network for Land Cover Change Detection. *Remote Sens.* **2016**, *8*, doi:10.3390/rs8060506.
8. Kaku, K. Satellite Remote Sensing for Disaster Management Support: A Holistic and Staged Approach Based on Case Studies in Sentinel Asia. *Int. J. Disaster Risk Reduct.* **2019**, *33*, 417–432, doi:https://doi.org/10.1016/j.ijdrr.2018.09.015.
9. Wing, M.G.; Burnett, J.; Sessions, J.; Brungardt, J.; Cordell, V.; Dobler, D.; Wilson, D. Eyes in the Sky: Remote Sensing Technology Development Using Small Unmanned Aircraft Systems. *J. For.* **2013**, *111*, 341–347, doi:10.5849/jof.12-117.
10. Mulac, B.L. Remote Sensing Applications of Unmanned Aircraft: Challenges to Flight in United States Airspace. *Geocarto Int.* **2011**, *26*, 71–83, doi:10.1080/10106049.2010.537786.
11. Xue, F.; Lu, W.; Chen, Z.; Webster, C.J. From LiDAR Point Cloud towards Digital Twin City: Clustering City Objects Based on Gestalt Principles. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 418–431, doi:https://doi.org/10.1016/j.isprsjprs.2020.07.020.
12. Michałowska, M.; Rapiński, J. A Review of Tree Species Classification Based on Airborne LiDAR Data and Applied Classifiers. *Remote Sens.* **2021**, *13*, doi:10.3390/rs13030353.
13. Schönberger, J.L.; Frahm, J.-M. Structure-from-Motion Revisited. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016; IEEE Computer Society, 2016; pp. 4104–4113.

14. Bosch, M.; Foster, K.; Christie, G.A.; Wang, S.; Hager, G.D.; Brown, M.Z. Semantic Stereo for Incidental Satellite Images. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, WACV 2019, Waikoloa Village, HI, USA, January 7-11, 2019; IEEE, 2019; pp. 1524–1532.
15. Voumard, J.; Derron, M.-H.; Jaboyedoff, M.; Bornemann, P.; Malet, J.-P. Pros and Cons of Structure for Motion Embarked on a Vehicle to Survey Slopes along Transportation Lines Using 3D Georeferenced and Coloured Point Clouds. *Remote Sens.* **2018**, *10*, doi:10.3390/rs10111732.
16. Liu, X. Airborne LiDAR for DEM Generation: Some Critical Issues. *Prog. Phys. Geogr. Earth Environ.* **2008**, *32*, 31–49.
17. Liu, C.-J.; Krylov, V.A.; Kane, P.; Kavanagh, G.; Dahyot, R. IM2ELEVATION: Building Height Estimation from Single-View Aerial Imagery. *Remote Sens.* **2020**, *12*, doi:10.3390/rs12172719.
18. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R.B. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397, doi:10.1109/TPAMI.2018.2844175.
19. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *CoRR* **2017**, *abs/1706.0*.
20. Güler, R.A.; Neverova, N.; Kokkinos, I. DensePose: Dense Human Pose Estimation in the Wild. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018; IEEE Computer Society, 2018; pp. 7297–7306.
21. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016; IEEE Computer Society, 2016; pp. 2818–2826.
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR); 2016; pp. 770–778.
23. Krizhevsky, A.; Sutskever, I.; Geoffrey E., H. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS); Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc., 2012; pp. 1097–1105.
24. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In Proceedings of the Artificial Neural Networks and Machine Learning - ICANN 2018 - 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III; Kurková, V., Manolopoulos, Y., Hammer, B., Iliadis, L.S., Maglogiannis, I., Eds.; Springer, 2018; Vol. 11141, pp. 270–279.
25. Eigen, D.; Puhrsch, C.; Fergus, R. Depth Map Prediction from a Single Image Using a Multi-Scale Deep Network. In Proceedings of the Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; 2014; pp. 2366–2374.
26. Laina, I.; Rupprecht, C.; Belagiannis, V.; Tombari, F.; Navab, N. Deeper Depth Prediction with Fully Convolutional Residual Networks. *CoRR* **2016**, *abs/1606.0*.
27. Alhashim, I.; Wonka, P. High Quality Monocular Depth Estimation via Transfer Learning. *CoRR* **2018**, *abs/1812.1*.
28. Huang, G.; Liu, Z.; Weinberger, K.Q. Densely Connected Convolutional Networks. *CoRR* **2016**, *abs/1608.0*.
29. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. Imagenet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE conference on computer vision and pattern recognition; 2009; pp. 248–255.
30. Bhat, S.F.; Alhashim, I.; Wonka, P. AdaBins: Depth Estimation Using Adaptive Bins. *CoRR* **2020**, *abs/2011.1*.
31. Geiger, A.; Lenz, P.; Urtasun, R. Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR); 2012.
32. Nathan Silberman Derek Hoiem, P.K.; Fergus, R. Indoor Segmentation and Support Inference from RGBD Images. In Proceedings of the ECCV; 2012.
33. Mahjourian, R.; Wicke, M.; Angelova, A. Unsupervised Learning of Depth and Ego-Motion From Monocular Video Using 3D Geometric Constraints. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018; IEEE Computer Society, 2018; pp. 5667–5675.
34. PNVR, K.; Zhou, H.; Jacobs, D. SharinGAN: Combining Synthetic and Real Data for Unsupervised Geometry Estimation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020; IEEE, 2020; pp. 13971–13980.
35. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M. Generative Adversarial Networks. *CoRR* **2014**, *abs/1406.2*.
36. Yu, D.; Ji, S.; Liu, J.; Wei, S. Automatic 3D Building Reconstruction from Multi-View Aerial Images with Deep Learning. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 155–170, doi:https://doi.org/10.1016/j.isprsjprs.2020.11.011.
37. Mou, L.; Zhu, X.X. IM2HEIGHT: Height Estimation from Single Monocular Imagery via a Fully Residual Convolutional-Deconvolutional Network. *CoRR* **2018**, *abs/1802.1*.
38. Amirkolaee, H.A.; Arefi, H. Height Estimation from Single Aerial Images Using a Deep Convolutional Encoder-Decoder Network. *ISPRS J. Photogramm. Remote Sens.* **2019**, *149*, 50–66, doi:https://doi.org/10.1016/j.isprsjprs.2019.01.013.
39. Srivastava, S.; Volpi, M.; Tuia, D. Joint Height Estimation and Semantic Labeling of Monocular Aerial Images with CNNS.

In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2017, Fort Worth, TX, USA, July 23-28, 2017; IEEE, 2017; pp. 5173–5176.

40. Carvalho, M.; Saux, B. Le; Trouvé-Peloux, P.; Champagnat, F.; Almansa, A. Multitask Learning of Height and Semantics From Aerial Images. *IEEE Geosci. Remote. Sens. Lett.* **2020**, *17*, 1391–1395, doi:10.1109/LGRS.2019.2947783.
41. Ghamisi, P.; Yokoya, N. IMG2DSM: Height Simulation From Single Imagery Using Conditional Generative Adversarial Net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 794–798, doi:10.1109/LGRS.2018.2806945.
42. Panagiotou, E.; Chochlakis, G.; Grammatikopoulos, L.; Charou, E. Generating Elevation Surface from a Single RGB Remotely Sensed Image Using Deep Learning. *Remote Sens.* **2020**, *12*, doi:10.3390/rs12122002.
43. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; 2016; ISBN 9780874216561.
44. Nielsen, M. Neural Networks and Deep Learning Available online: http://neuralnetworksanddeeplearning.com/ (accessed on 24 March 2021).
45. Digimap Available online: https://digimap.edina.ac.uk/ (accessed on 25 March 2021).
46. Edina Available online: https://edina.ac.uk/ (accessed on 25 March 2021).
47. Defra (Department for Environment, Food and Rural Affairs) Spatial Data Available online: https://environment.data.gov.uk/DefraDataDownload/ (accessed on 25 March 2021).
48. 2018 IEEE GRSS Data Fusion Contest Available online: http://dase.grss-ieee.org/index.php (accessed on 24 March 2021).
49. IEEE France GRSS Chapter Available online: https://site.ieee.org/france-grss/2018/01/16/data-fusion-contest-2018-contest-open/ (accessed on 24 March 2021).
50. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference of Medical Image Computing and Computer-Assisted Intervention 18 (MICCAI); 2015; pp. 234–241.
51. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI); Ourselin, S., Joskowicz, L., Sabuncu, M.R., Ünal, G.B., Wells, W., Eds.; 2016; Vol. 9901, pp. 424–432.
52. Iglovikov, V.; Shvets, A. TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. *CoRR* **2018**, *abs/1801.0*.
53. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA,; 2015; pp. 3431–3440.
54. Shi, W.; Caballero, J.; Huszar, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016; IEEE Computer Society, 2016; pp. 1874–1883.
55. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167* **2015**, 1–11, doi:10.1007/s13398-014-0173-7.2.
56. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *ICLR* **2015**, 1–15, doi:http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503.
57. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the International Conference on Computer Vision (ICCV); 2015; pp. 1026–1034.
58. Jones, K.L.; Poole, G.C.; O'Daniel, S.J.; Mertes, L.A.K.; Stanford, J.A. Surface Hydrology of Low-Relief Landscapes: Assessing Surface Water Flow Impedance Using LIDAR-Derived Digital Elevation Models. *Remote Sens. Environ.* **2008**, *112*, 4148–4158, doi:https://doi.org/10.1016/j.rse.2008.01.024.
59. Sofia, G.; Bailly, J.; Chehata, N.; Tarolli, P.; Levavasseur, F. Comparison of Pleiades and LiDAR Digital Elevation Models for Terraces Detection in Farmlands. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1567–1576, doi:10.1109/JSTARS.2016.2516900.
60. Palmer, D.; Koumpli, E.; Cole, I.; Gottschalg, R.; Betts, T. A GIS-Based Method for Identification of Wide Area Rooftop Suitability for Minimum Size PV Systems Using LiDAR Data and Photogrammetry. *Energies* **2018**, *11*, 1–22.