# ROOF TYPE CLASSIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORKS ON LOW RESOLUTION PHOTOGRAMMETRIC POINT CLOUDS FROM AERIAL IMAGERY

*Maria Axelsson*[⋆], *Ulf Söderman*[⋆], *Andreas Berg*[†], *Thomas Lithén*[†]

[⋆] Swedish Defence Research Agency (FOI), Linköping, Sweden
[†] Lantmäteriet, Gävle, Sweden

## ABSTRACT

Three-dimensional (3D) reconstruction of buildings is an active research area with applications in e.g. city planning, environmental simulations, and city navigation. Automatic 3D building reconstruction methods based on point clouds from laser scanning or methods based on high resolution dense photogrammetric point clouds are common in the literature. In applications where large land areas need to be covered regularly it is not practical to use laser scanning or acquire images with high resolution and large image overlaps. In these applications the reconstructed photogrammetric point cloud has low resolution with less building details. We present a method where the most common roof types are classified using a deep convolutional neutral network (CNN) pre-trained using RGB data in this challenging type of data. In addition, a method for roof height estimation for each roof type is presented to support automatic 3D building reconstruction using model building shapes. Results are shown for a low resolution dense photogrammetric point cloud generated using multi-view stereo reconstruction of standard overlapping aerial images from nationwide data collection. The method is intended to support automated generation of a nationwide 3D landscape model.

***Index Terms***— Building reconstruction, Deep learning, Convolutional neural network, Multi-view stereo, Aerial imagery

## 1. INTRODUCTION

The demand for three-dimensional (3D) models is growing and automatic 3D reconstruction of buildings is an active research topic in the research areas of computer vision, remote sensing, and photogrammetry. Today, multi-view stereo reconstruction of 3D geometry from two-dimensional (2D) images is well studied and used in large scale applications. Dense photogrammetric point clouds of large areas can be generated from highly overlapping aerial images and are common in city modeling, see e.g. [1, 2, 3]. These point clouds contain both height information and spectral information which makes them attractive for reconstructing buildings in 3D as they can be texture mapped using the spectral information.

Previous work on 3D building reconstruction include methods based on both relatively high resolution photogrammetric point clouds, see e.g. [3, 4, 5] and point clouds from laser scanning, see e.g. [6, 7, 8]. Approaches based on imagery from e.g. aerial images with small ground sampling distance or laser scanning, provide point clouds with high resolution and many resolved building details. However, in some applications it is not practical to acquire images with large image overlaps and small ground sampling distance or use laser scanning. One such example is when large areas needs to be covered regularly to keep the point cloud up to date, e.g. when generating a 3D map for an entire country on a regular basis. In this application the point cloud can be generated using aerial images with smaller image overlaps and imaged at larger distances. This gives a photogrammetric point cloud with low resolution, which is more challenging to use for 3D building reconstruction and 3D map generation than high resolution photogrammetric point clouds or a point cloud from laser scanning.

Building roof type classification is an important step in model-driven 3D building reconstruction. In this paper we present a method for automatic classification of the most common model building shapes, ridge roofs and flat roofs, using deep convolutional neural networks (CNN), as this approach to classification has shown strong results for hard problems in other application areas. In the literature on building classification there are some initial results using deep learning, see e.g. [9, 10]. An overview of our network architecture is shown in Fig. 1. The network is pre-trained using RGB images and we apply transfer learning to our data which is a point cloud sampled on a regular grid. We also propose a method for estimation of the highest roof height to enable reconstruction of the building from 2D building polygons to 3D models using model library roof shapes. The intended application is a large scale overview of buildings in a 3D landscape model e.g. a 3D

This work was funded by Lantmäteriet (the Swedish mapping, cadastral, and land registration authority). Maria Axelsson is also partly funded by the Swedish Innovation Board (VINNOVA).

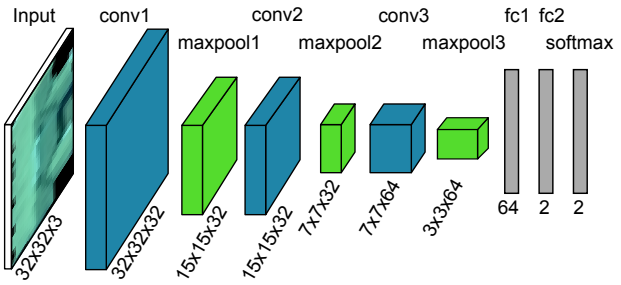


**Fig. 1**. Convolutional neural network for building classification. The input is a three band image of size $32 \times 32 \times 3$. The network consists of three iterations of convolutional layers and two fully connected layers and a softmax layer followed by a classifier.

map for an entire country. We show results using a relatively low resolution photogrammetric point cloud, a digital surface model, generated using multi-view stereo reconstruction from high altitude aerial imagery with relatively small image overlap. This type of point cloud often has missing points on the building roofs in areas where few images are used in the image matching and it is relatively noisy. An example overview from a small part of the point cloud that is used in the experiments is shown in Fig. 2.

## 2. ROOF TYPE CLASSIFICATION

In our framework we classify patches of buildings into the two most common roof types, ridge roofs and flat roofs. Ridge roofs include different types of ridge roofs such as gable, half-hip, hip, and mansard roofs. As only limited annotated data for building classification is available the building roof type classification is performed using transfer learning of a pre-trained CNN on RGB data.

### 2.1. Network architecture

The network architecture is illustrated in Fig. 1. It is a CNN where the input is a three band image of size $32 \times 32 \times 3$. The input size is well suited to the problem of building classification since many of the buildings fit well into this size without much interpolation. The network consists of three iterations of convolutional layers followed by ReLU and max poolings and two fully connected layers, where the first is followed by ReLU, and in the end a softmax layer followed by a classifier. The classifier outputs two classes, one for each roof type. The image input use zerocenter normalization of the data. We initialize the network using weights from a pre-trained network for object classification using CIFAR10 data [11] which is common RGB data. In our data the three spectral channels contain near infrared (NIR), red, and green. In addition to the spectral information it includes the height information from

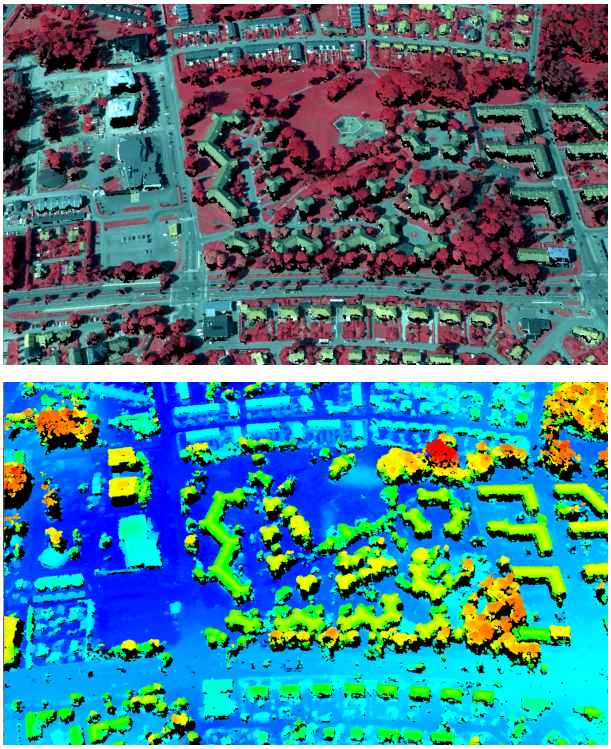

**Fig. 2**. (Top) The spectral information in the point cloud with near infrared, red, and green visualized as RGB. (Bottom) The height information in the point cloud.

the point cloud. Combinations of these four input channels in the training are evaluated in our experiments.

### 2.2. Input preprocessing

Our approach requires that the point cloud is on a regular grid in the ground coordinate system. This enables analysis of the data as an image with one or more spectral bands and corresponding height information in each pixel instead of a 3D point cloud. The annotated training and test data used for classification are mapped to fit the input layer of the network architecture. We base the preprocessing of the data on the 2D polygon associated with each building. A patch for each building is cropped from the point cloud using the building polygon and the background is set to zero. The 2D building polygon, which can cover a whole building or part of a building, outlines the projection of the building on the ground. The building patches are also rotated to align the main axis with the image coordinates using the lengths of the segments in the building polygon. Depending on the building shape the main axis can be aligned both horizontally or vertically in the image. After rotation the building patches are resampled to $32 \times 32 \times 3$ pixels to fit the input layer. Examples of training patches with the bands NIR, red, and green are shown in Fig. 3. Before training the patches are also augmented using

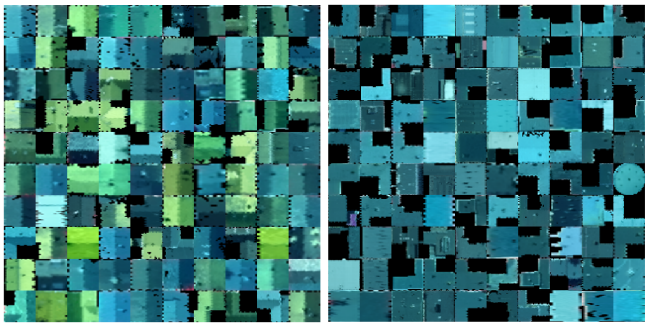

**Fig. 3**. Examples of training data patches for ridge roofs and flat roofs using the spectral bands in near infrared, red, and green as the three input layers.

rotation and flipping to create more training data using the annotated data. This also makes the two main directions equal and removes any differences in the alignment after rotating the patches.

## 3. BUILDING HEIGHT ESTIMATION

In addition to the roof type, the building height is also needed to approximate each building using a model roof shape in a 3D map. The building height, defined as the highest roof excluding for example chimneys and antennas, is estimated for each building depending on the result from the roof type classification.

### 3.1. Height estimation for ridge roofs

The building height for ridge roofs is estimated by identifying points on the highest ridge and estimating the height from these using the values in the point cloud. First, the height image and an intensity image, which is the average of the spectral information for all bands in each pixel, are smoothed using normalized convolution [12] and the image gradients are calculated for both images. Candidate ridge points are extracted using the two gradient images. The height gradient should be small, below a threshold, and the image gradient should be large, above a threshold. Ridges are assumed to be roof parts where the surface normal point upwards and there is some structure in the intensity image, either from different illumination on the two surfaces around the ridge or that the ridge appear as a line structure in the image. A certainty measurement for each candidate point is calculated by analyzing several local image profiles with center in the candidate point. An example ridge roof and profiles though a candidate point are shown in Fig. 4. The candidate point should be the global maximum in the profile. Lines are fitted jointly to image points around the maximum. The fit of the line should be good in a least square error sense and there should be a slope from the candidate point on both sides. If at least one image profile fulfill these criteria the point is denoted validated

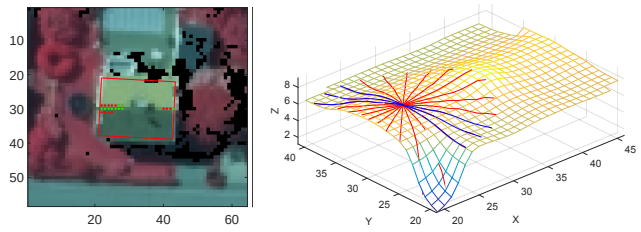


**Fig. 4**. Example ridge roof. (Left) Validated candidate points in green in the spectral data. (Right) Example of image profiles used to validate a ridge point in the height data.

ridge point. The height is estimated using the height values at the positions of the validated ridge points. The final height estimate use the 80th percentile of the height values to reject outliers.

### 3.2. Height estimation for flat roofs

The building height for flat roofs is estimated using the histogram of all heights inside the building polygon. Large flat roofs have histograms with only one large peak and sectioned flat roofs have histograms with multiple peaks. If there is only one large histogram peak where all values in a small interval around the maximum represent a large part of the building polygon, e.g. more than 80 percent, the building height is calculated as the 80th percentile of these values.

The building height for flat roofs with more than one distinct roof height is found using all histogram peaks. For each histogram peak a binary image is analyzed where all points in a small interval around the maximum are set to foreground. Connected foreground components in this image larger than a threshold, e.g. 30 pixels, are identified using 8-connected labeling. For sectioned flat roofs one height is estimated for each of the connected components by selecting heights using the connected component as a binary mask. All estimates use the 80th percentile of the height values to reject outliers. The final building height is selected as the highest estimate.

## 4. EVALUATION

The evaluation of the building roof type classification using CNN and the method for building height estimation is performed using a relatively low resolution photogrammetric point cloud with nationwide coverage.

### 4.1. Data

The aerial images used in this work cover approximately $6.6 \times 3.7$ km on the ground with a ground sampling distance of about 0.25 m. The images are overlapping in both the flight direction with 60 % and in cross direction with 25 %. A dense photogrammetric point cloud is generated from the

**Table 1**. Classification results in terms of accuracy for different input configurations. Average over ten trained networks.

| Input layers (RGB) | Ridge roof | Flat roof | Total |
|---|---|---|---|
| Height, red, green | 97.48% | 90.80% | 96.65% |
| NIR, red, green | 97.37% | 90.80% | 96.55% |
| Height, height, height | 96.35% | 81.19% | 94.45% |

images using multi-view stereo with Semi-Global Matching and fusion of depth where redundant depth estimates from overlapping stereo models are merged [13]. Due to the image overlap the number of available images for an area varies from two to six. Stereo models are calculated from each image pair and depths are fused. The resulting 2.5D point cloud, which is called Digital Surface Model from Aerial Photos[1], is sampled on a regular grid of $0.5 \times 0.5$ m. Each point that is matched contains both spectral and height information. The data has nationwide coverage over Sweden.

In addition to the point cloud, 2D building polygons of the building footprint are used to crop out the relevant point cloud area for each building or building part. Also a Digital Terrain Model (the National elevation model) with resolution $1 \times 1$ m is used to recover the building height over the local terrain.

### 4.2. Experiments and results

The proposed classification method using CNNs has been evaluated using buildings with manually marked roof heights and roof types from two classes, houses with ridge roofs and houses with flat or very low-slope roofs. The training set contains 1200 ridge roofs and 400 flat roofs and the test set contains 403 ridge roofs and 197 flat roofs. Multiple copies of the flat roof data was added to the training set to remove unwanted bias towards the ridge roof class. For evaluation of the height estimation a smaller number of houses were also annotated with the maximal building height not including antennas and chimneys. In this data set there are 76 buildings with ridge roofs and 24 buildings with flat roofs.

The network was trained using stochastic gradient descent with momentum using the two classes. In our experiments we have evaluated different combinations of the three spectral bands and the height information as the three input layers. The best result was obtained by combining height, red, and green as the three input layers, but almost the same result was obtained using only the spectral information from the data using NIR, red and green, see Table 1. For reference the result using only the height information in all three bands is also shown.

In the evaluation of the building height estimation, the estimated height is compared to the marked ground truth value using the absolute difference. The maximal height value inside the building polygon is also compared to the ground truth in the same way. The mean and standard deviation of the absolute differences are 0.106 and 0.103 for the absolute difference between the estimated value and the ground truth $d_{est}$ and 0.631 and 0.634 for the absolute difference between the maximal value and the ground truth $d_{max}$. The error histograms for the absolute differences are shown in Fig. 5. This shows that the proposed method for estimating building height using the ridge height for ridge roofs and the highest flat area for flat roofs gives a better height estimate of the highest roof height than the maximal height from the point cloud. For both roof types there are often smaller protruding structures on the main roof that gives higher maximal values in the point cloud.

[1]In Swedish: "Ytmodell från flygbilder", see: www.lantmateriet.se

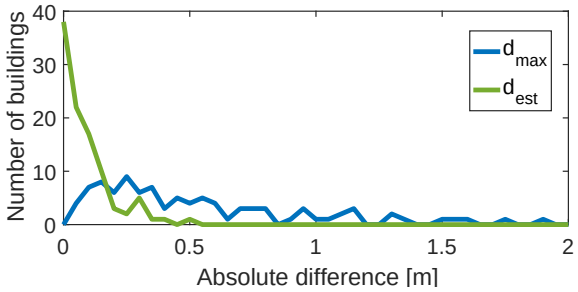


**Fig. 5**. Error histograms for the absolute difference between the true and estimated building height, $d_{est}$ and the true and maximal building height, $d_{max}$.

## 5. DISCUSSION AND CONCLUSION

We have presented a method using deep CNNs for roof type classification and a method for roof height estimation to support 3D building reconstruction from a low resolution photogrammetric point cloud generated using multi-view stereo reconstruction of standard overlapping aerial images from nationwide data collection. This type of point cloud is very challenging compared to point clouds from laser scanning or from high resolution aerial imagery. We show using annotated data that building roof types can be identified with 96.65% accuracy overall and the highest roof height excluding e.g. chimneys and antennas can be estimated with high accuracy for each roof type. The estimated roof height represent the building height better than using the maximum of the point cloud inside the building polygon directly.

In future work it is interesting to investigate possibilities to divide larger building polygons in subparts automatically and analyze them separately using the proposed method. Also, depending on the availability of large training sets it is also interesting to compare transfer learning to training using only the spectral and height information from the same type of data.

## 6. REFERENCES


[1] Norbert Haala, "Benchmark on Image Matching Final report," *EuroSDR Official Publication No 64* , pp. 115 – 144, 2014.

[2] Norbert Haala and Martin Kada, "An update on automatic 3D building reconstruction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 6, pp. 570 – 580, 2010.

[3] A. P. McClune, J. P. Mills, P. E. Miller, and D. A. Holland, "Automatic 3D building reconstruction from a dense image matching dataset," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B3, pp. 641–648, 2016.

[4] S. Malihi, M. J. Valadan Zoej, M. Hahn, M. Mokhtarzade, and H. Arefi, "3D building reconstruction using dense photogrammetric point cloud," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B3, pp. 71–74, 2016.

[5] B. Xiong, S. Oude Elberink, and G. Vosselman, "Building modeling from noisy photogrammetric point clouds," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3, pp. 197–204, 2014.

[6] Andre Henn, Gerhard Groger, Viktor Stroh, and Lutz Plumer, "Model driven reconstruction of roofs from sparse LIDAR point clouds," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 76, pp. 17 – 29, 2013.

[7] Elisabeth Orthuber and Janja Avbelj, "3D building reconstruction from lidar point clouds by adaptive dual contouring," in *PIA15+HRIGI15 - Joint ISPRS conference*, March 2015, vol. II-3 of *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 157–164.

[8] A. Sampath and J. Shan, "Segmentation and reconstruction of polyhedral building roofs from aerial lidar point clouds," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1554–1567, March 2010.

[9] F. Alidoost and H Arefi, "Knowledge based 3D building model recognition using convolutional neural networks from lidar and aerial imageries," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLI-B3, pp. 833–840, 2016.

[10] T. Partovi, F. Fraundorfer, S. Azimi, D. Marmanis, and P. Reinartz, "Roof type selection based on patch-based classification using deep learning for high resolution satellite imagery," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLII-1/W1, pp. 653–657, 2017.

[11] G Hinton A Krizhevsky, "Learning multiple layers of features from tiny images," *Technical report, University of Toronto*, 2009.

[12] H. Knutsson and C.-F. Westin, "Normalized and differential convolution: Methods for interpolation and filtering of incomplete and uncertain data," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1993, pp. 515–523.

[13] M. Rothermel, K. Wenzel, D. Fritsch, and N Haala, "SURE: Photogrammetric surface reconstruction from imagery," in *Proceedings LC3D Workshop*, December 2012.