DEEP LEARNING BASED ROOF TYPE CLASSIFICATION USING VERY HIGH RESOLUTION AERIAL IMAGERY

M. Buyukdemircioglu[1], R. Can[1], S. Kocaman[1*]

Dept. of Geomatics Engineering, Hacettepe University, 06800 Beytepe Ankara, Turkey – (mbuyukdemircioglu, recepcan, sultankocaman)@hacettepe.edu.tr

Theme Session: Deep learning in Remote Sensing

KEY WORDS: Deep Learning, CNN, Roof type classification, 3D GIS, 3D City models

ABSTRACT:

Automatic detection, segmentation and reconstruction of buildings in urban areas from Earth Observation (EO) data are still challenging for many researchers. Roof is one of the most important element in a building model. The three-dimensional geographical information system (3D GIS) applications generally require the roof type and roof geometry for performing various analyses on the models, such as energy efficiency. The conventional segmentation and classification methods are often based on features like corners, edges and line segments. In parallel to the developments in computer hardware and artificial intelligence (AI) methods including deep learning (DL), image features can be extracted automatically. As a DL technique, convolutional neural networks (CNNs) can also be used for image classification tasks, but require large amount of high quality training data for obtaining accurate results. The main aim of this study was to generate a roof type dataset from very high-resolution (10 cm) orthophotos of Cesme, Turkey, and to classify the roof types using a shallow CNN architecture. The training dataset consists 10,000 roof images and their labels. Six roof type classes such as flat, hip, half-hip, gable, pyramid and complex roofs were used for the classification in the study area. The prediction performance of the shallow CNN model used here was compared with the results obtained from the fine-tuning of three well-known pre-trained networks, i.e. VGG-16, EfficientNetB4, ResNet-50. The results show that although our CNN has slightly lower performance expressed with the overall accuracy, it is still acceptable for many applications using sparse data.

1. INTRODUCTION

Buildings are the most important structural component of cities in many aspects. Building measurement and analysis have been used for many applications, such as urban planning, land management or climate change monitoring (Alidoost et al., 2019). 3D City models in LoD2 (Level of Detail 2) or higher levels include roof geometries that can be used in 3D GIS applications, such as solar potential estimation, quality evaluation and verification of existing data, roof reconstruction, and enhancing the LoD0/LoD1 data with the roof type attributes (Biljecki and Dehbi, 2019). A building roof type classification approach can be utilized for model-driven 3D building reconstruction, which also reduces the dependency for a digital surface model (DSM) (Partovi et al., 2017).

Deep learning (DL) and convolutional neural networks (CNNs) have contributed to the improvements in both photogrammetry and remote sensing tasks such as classification, 3D reconstruction change detection, object racking and extraction dramatically (Heipke and Rottensteiner, 2020). There exist different approaches in literature for roof type classification using the DL methods. Axellson et al. (2018) used low resolution photogrammetric point clouds from aerial imagery using deep CNNs (DCNNs) for roof type classification and roof height estimation. Partovi et al. (2017) utilized WorldView-2 pansharpened multispectral satellite image of Munich city, Germany, with 50 cm spatial resolution for roof type classification using VGG-Net model (Simonyan and Zisserman, 2015). Alidoost and Arefi (2016) have developed a model-based approach for automatic recognition of roof types using convolutional neural networks using LiDAR (Light Detection and Ranging) data and aerial images. Mohajeri et al. (2017) employed Support Vector Machine (SVM) classifier and LiDAR data for classifying six different roof types using a total of 10,085 roofs in Geneva, Switzerland; and obtained 66% overall accuracy. Qin et al. (2019) evaluated DCNN on the panchromatic and multispectral sensor (PMS) imagery of Gaofen-2 satellite in dense urban areas for image segmentation and obtained 94.67% accuracy. They also stated that DCNNs are promising for building mapping from very high resolution imagery in dense urban areas with different roof patterns.

Bittner et al. (2019) experimented Conditional Generative Adversarial Network (cGAN) using very dense DSM with 50 cm resolution generated from Worldview-1 satellite imagery for roof type classification in their study. Castagno and Atkins (2018) proposed a method for labelling and classifying roof types with an automatic approach using different type of supervised machine learning and DL methods by fusing LiDAR and satellite imagery. Assouline et al. (2017) used LiDAR data and the random forest method for classifying building rooftops for large-scale solar photovoltaic deployment and obtained an average accuracy of 67%. Another prominent dataset used for similar tasks was provided by (Rottensteiner et al., 2012) within the ISPRS building reconstruction and urban classification benchmark. The dataset consisted of high resolution (8 cm) aerial imagery and airborne laser scanning data (6 points/m^2) for

* Corresponding author

building, tree detection and 3D building reconstruction. The benchmark results were presented by (Rottensteiner et al., 2014) and the common problems of the state-of-the-art methods were discussed.

In this study, we present a dataset with 10 cm resolution for urban roof type classification that consists of 10,000 roof images categorized in six commonly used roof types, which can also be used for segmentation or building reconstruction. A shallow CNN architecture was evaluated with the dataset and compared with the results achieved by three state-of-the-art pre-trained CNN models, which were fine-tuned with the study data. We achieved overall accuracy values between 80%-86% the shallow CNN and one pre-trained models.

2. DATA PREPARATION

In this study, orthophotos with 10 cm spatial resolution, which were produced from a total of 4468 aerial images taken with 80% forward overlap and 60% lateral overlap using UltraCam Falcon large-format digital camera, were utilized to prepare the training data. The building footprint vectors were manually delineated from the stereo models during a mapping project by operators (photogrammetry professionals) (Buyukdemircioglu et al., 2018). Since the building footprints do not contain attribute information and often differ from roof boundaries; a visual comparison by overlaying the orthophotos and the building footprints was performed to manually adjust the roof border vectors, and to include the roof type attribute information for each vector. The roofs that are completely or partially covered by trees or shadows were also visually checked, and were not included in the dataset. A roof library with six different commonly types (flat, hip, halfhip, gable, pyramid, complex) was populated for classification. A view of a roof polygon before and after the manual editing (adjustment) is given in Figure 1.

Figure 1. Building roof polygon before (left) and after (right) manual adjustment.

When creating dataset, a balance between the different roof type classes was sought as much as possible. Yet, due to the different numbers of instances for each roof type in the study area, classes such as halfhip and pyramid have lower number of samples than the others. The class sample distribution of the dataset is presented in Table 1.

Roof Type	Training (72%)	Validation (18%)	Test (10%)	Total
Complex	1620	405	225	2250
Flat	1260	315	175	1750
Gable	1260	315	175	1750
Hip	1260	315	175	1750
Pyramid	1080	270	150	1500
Halfhip	720	180	100	1000

Table 1. The distribution of the class samples used in the study.

In the next step, the roof images were clipped automatically from the orthophotos by using the vector data. Since there are a total of 927 orthophotos in the study, a roof may be located at the border of an image, and thus covered by multiple orthophotos. Therefore, an orthophotos mosaic was produced prior to clipping with the same resolution (10 cm). The roof images were then clipped and classified based on their roof type attributes from the mosaic using the FME software (Safe Software, 2021) and stored in individual folders for each class. The pixel values outside the roof polygons were stored as "NoData". Sample roofs from all classes are given in Figure 2.

3. METHODOLOGY

The DL based methods, especially the CNNs, exhibit great prediction performances on image classification tasks. To train a CNN model from the scratch means finding optimal values for the large quantities of parameters. The number of parameters depends on the model design, and the modeling is only possible if there is sufficient amount of training data. The sufficiency of the data for model training in a CNN architecture is problem specific. In this study, we developed a CNN architecture for the roof type classification task. In order to compare the prediction performance of our CNN with the state-of-the-art architectures, we modified the pre-trained EfficientNet (Tan and Le, 2019), ResNet (He et al., 2016) and VGG-16 (Simonyan and Zisserman, 2015) models and fine-tuned them for the roof type classification task using the study dataset. For the training step, the generated roof images were split into train, validation and test data (72%, 18% and 10% respectively) to be used as input for the CNN models. Python 3.8 with the open source DL library Tensorflow 2.4 environment was used to train the deep CNN using the generated roof images for the six classes. Since the DL methods in general require large amount of data for obtaining accurate results, a data augmentation technique was also employed in the study while training the CNN to increase the accuracy and to prevent from overfitting. We also used data augmentation with horizontal flip, vertical flip, zoom (0.1) and rotation (0.1) while training the CNN models. As the last step, we compared the classification results and the accuracy of the different CNN models for performance assessment using precision, recall, F1-score, and accuracy values. The details of our shallow CNN and the pre-trained models are explained in the following sub-sections.

3.1 Roof Type Classification using Shallow CNN Model

The shallow CNN architecture implemented here includes 312,550 trainable parameters. The batch size parameter was chosen as 64. There are five convolutional blocks in the architecture (Figure 3). A 3x3 kernel filter size was chosen for the convolutional layers (Conv2D). The default pool size (2x2) was used in the pooling layers (MaxPooling2D, GlobalAveragePooling2D). The batch normalization (Ioffe and Szegedy, 2015) with the momentum of 0.01 was used to prevent from overfitting. We used global average pooling layer instead of flatten layer to reduce the parameter size. Adam optimizer (Kingma and Ba, 2015) with a learning rate of 0.0003 was used with the categorical cross entropy loss. The model was trained for 150 epochs. A more detailed view of the model is given in Figure 3.

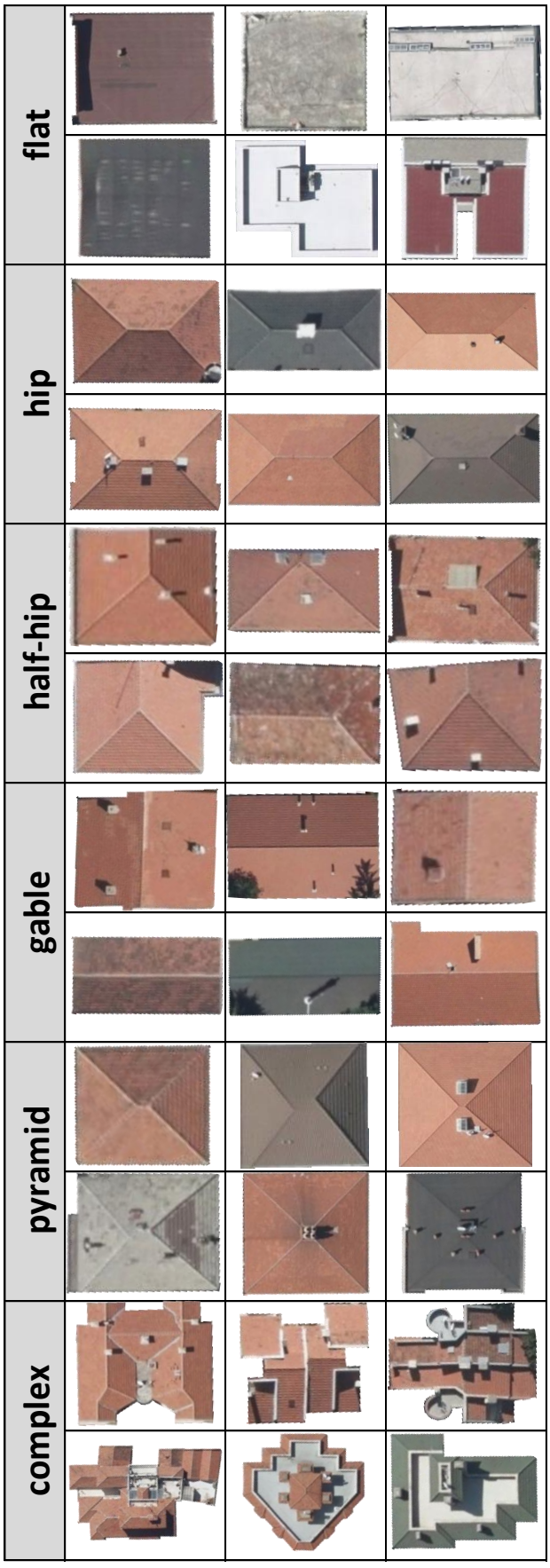

Figure 2. Roof type class samples in the study area (Cesme, Turkey).

Layer (type)	Output Shape	Param #
conv2d_280 (Conv2D)	(None, 140, 140, 32)	864
batch_normalization_233 (Bat	(None, 140, 140, 32)	128
activation_216 (Activation)	(None, 140, 140, 32)	0
max_pooling2d_60 (MaxPooling	(None, 70, 70, 32)	0
conv2d_281 (Conv2D)	(None, 70, 70, 64)	18496
conv2d_282 (Conv2D)	(None, 70, 70, 64)	36864
batch_normalization_234 (Bat	(None, 70, 70, 64)	256
activation_217 (Activation)	(None, 70, 70, 64)	0
max_pooling2d_61 (MaxPooling	(None, 35, 35, 64)	0
conv2d_283 (Conv2D)	(None, 35, 35, 128)	73856
conv2d_284 (Conv2D)	(None, 35, 35, 128)	147456
batch_normalization_235 (Bat	(None, 35, 35, 128)	512
activation_218 (Activation)	(None, 35, 35, 128)	0
global_average_pooling2d_34	(None, 128)	0
dense_76 (Dense)	(None, 256)	33024
dense_77 (Dense)	(None, 6)	1542

Total params: 312,998
Trainable params: 312,550
Non-trainable params: 448

Figure 3. An overview of the generated CNN model

3.2 Roof Type Classification using Transfer Learning

Fine-tuning is a method, which freezes the base model of a previously trained network with massive amount of data (e.g. millions of images), and trains only the selected top layers of the network with the study data. The approach helps to overcome weak performance problems of the CNNs trained with smaller datasets and to ensure that the model is more relevant for the study area. In this study, we modified and fine-tuned three deep CNNs, such as VGG-16, EfficientNetB4 and ResNet-50, which were pre-trained on ImageNet (Deng et al., 2009) for image classification tasks. The fully connected layers of the three networks were replaced with the new fully connected layer block designed and implemented in this study. The batch size of the models was chosen as 64, and the categorical cross entropy was used as loss function. The Adam method was used as optimizer. After the replacement, the modified networks were configured in order to train the new fully connected layer block for 10 epochs. In this configuration, the layers of the base networks, i.e. EfficientNet, ResNet and VGG-16, were configured as non-trainable; and only the fully connected layer block of each network was configured as trainable. After 10 epochs, the last layers of the base networks were configured as trainable and the whole networks was re-trained for 50 epochs.

4. RESULTS AND DISCUSSIONS

The performance evaluation metrics used the study were F1-Score, precision, recall and accuracy. A total of 1,000 roof images were employed as test data, which were randomly selected from the 10,000 samples. The classification results are presented and discussed in the following subsections in detail.

4.1 Performance of Shallow CNN Model

The shallow CNN model implemented here achieved 80% accuracy as the overall performance from the test set as shown in Table 2. In the table, the precision, recall, F1-score and the

weighted and macro averages of the accuracy values obtained from all classes are also presented. In the support column, the numbers of test samples in each class are given. According to the results, the flat roof type, which has relatively uniform geometric and spectral properties, has the highest precision and F1-score values. The precision of the complex roof type class was in the second place. This type also has the highest number of samples in the dataset. The least F1-score was obtained from the half-hip roof type, which also has the smallest number of test samples.

Roof Type	Precision	Recall	F1-score	Support
Complex	87%	80%	83%	225
Flat	89%	84%	86%	175
Gable	76%	86%	81%	175
Halfhip	74%	62%	67%	100
Hip	78%	75%	76%	175
Pyramid	73%	86%	79%	150
Accuracy			**80%**	1000
Macro avg.	79%	79%	79%	1000
Weighted avg.	80%	80%	80%	1000

Table 2. Classification performance results of the shallow CNN model implemented in the study.

4.2 Performance of Fine-tuned VGG-16

The VGG-16 (Simonyan and Zisserman, 2015) model achieved 92.7% accuracy with top-5 score in ImageNet Benchmark, which is a dataset of 1000 different classes with more than 14 million images. The VGG-16 model architecture is given in Figure 4. The classification results obtained from the fine-tuned VGG-16 model in this study are presented in Table 3. The overall classification accuracy obtained from the model was 86%. Here, the pyramidal roof types exhibited the highest precision and F1-score values. The half-hip roof type again yielded to lower prediction performance in comparison to the other classes.

Figure 4. VGG-16 model architecture (Simonyan and Zisserman, 2015).

Roof Type	Precision	Recall	F1-score	Support
Complex	90%	80%	84%	225
Flat	81%	93%	86%	175
Gable	90%	84%	87%	175
Halfhip	82%	80%	81%	100
Hip	83%	92%	87%	175
Pyramid	94%	89%	91%	150
Accuracy			**86%**	1000
Macro avg.	86%	86%	86%	1000
Weighted avg.	87%	86%	86%	1000

Table 3. Classification performance results of the fine-tuned VGG-16 architecture.

4.3 Performance of Fine-tuned EfficientNetB4

EfficientNet (Tan and Le, 2019) is a model developed by Google for scaling up CNNs by increasing number of layers for classification tasks. Most of the CNN models arbitrarily scales network dimensions, where EfficientNet uniformly scales each dimension with a fixed set of scaling coefficients. The model balances network depth, resolution and width for better performance. The EfficientNet model architecture is given in Figure 5. The classification results obtained from the fine-tuned EfficientNetB4 model are provided in Table 4. This model yielded to an overall accuracy of 83%, whereas the class with highest F1-score was pyramidal as in VGG-16. The precision values obtained from the halfhip and the complex roof types were equally high (85%), and followed by the pyramid type (84%).

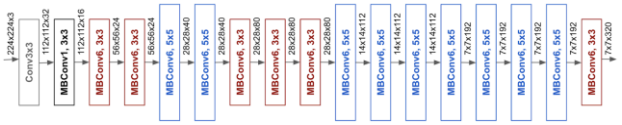

Figure 5. EfficientNet model architecture (Tan and Le, 2019).

Roof Type	Precision	Recall	F1-score	Support
Complex	85%	77%	81%	225
Flat	82%	83%	83%	175
Gable	79%	88%	83%	175
Halfhip	85%	81%	83%	100
Hip	82%	81%	81%	175
Pyramid	84%	89%	87%	150
Accuracy			**83%**	1000
Macro avg.	83%	83%	83%	1000
Weighted avg.	83%	83%	83%	1000

Table 4. Classification performance results of the fine-tuned EfficientNetB4 architecture.

4.4 Performance of Fine-tuned ResNet-50

The ResNet-50 (He et al., 2016) is a CNN model, which is commonly used in many DL and computer vision studies. The model was the winner of the ImageNet challenge in 2015. The ResNet allows users to train very DCNNs with hundreds or thousands of layers with high performance. The ResNet-50 model architecture is provided in Figure 6. The classification performance results of the fine-tuned ResNet-50 model in this study is provided in Table 5. The results show that an overall accuracy of 85% was obtained from the model, which is slightly inferior to the VGG-16 results. Similar to the VGG-16, the pyramidal roof types exhibited the highest precision and F1-score values. On the contrary, the halfhip roof type also exhibit high precision and F1-score and placed as second.

When the overall results were evaluated, it can be said that the pre-trained model exhibit higher prediction performances due to the very high number of training data used in the model building phase. On the other hand, our shallow CNN model also exhibited high accuracy even though the training data was sparse. The model can be tuned further for increasing the prediction performance.

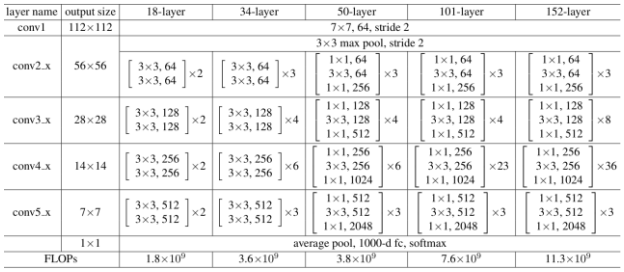

Figure 6. ResNet-50 model architecture (He et al., 2016).

Roof Type	Precision	Recall	F1-score	Support
Complex	86%	79%	82%	225
Flat	84%	85%	85%	175
Gable	82%	88%	85%	175
Halfhip	89%	85%	87%	100
Hip	81%	86%	83%	175
Pyramid	90%	88%	89%	150
Accuracy			**85%**	1000
Macro avg.	85%	85%	85%	1000
Weighted avg.	85%	85%	85%	1000

Table 5. Classification report of fine-tuned ResNet-50

5. CONCLUSIONS AND FUTURE WORK

In this study, a roof type dataset compiled from very high resolution aerial imagery was generated for automatic roof type classification tasks. We defined six different roof types including complex, flat, gable, hip, halfhip and pyramid in the dataset. A shallow CNN model was also implemented in the study and its prediction performance was investigated by comparing with three different pre-trained CNN models, i.e. VGG-16, EfficientNetB4, and ResNet-50. The pre-trained models were fine-tuned here prior to the comparison. The highest roof type classification accuracy was obtained from the fine-tuned VGG-16 model (86% overall accuracy). The shallow CNN implemented here yielded to 80% overall accuracy. The fine-tuned models help to overcome performance problems of smaller datasets. The ranking of the class accuracy values obtained from the four models were not uniform, such as the halfhip type yielded the lowest accuracy with the VGG-16 model and one of the highest with the ResNet-50 model.

The classification results show that initial results of the method were promising, but can be improved further by using more data. Although the accuracy of the used models was comparable, the pre-trained models achieved slightly higher performance than the implemented shallow CNN. The quality, the resolution and the total number of the generated roof patches for training are very important for obtaining satisfying performances. As planned work, we will focus on improving the results of developed shallow CNN, improve the training dataset with images from different data sources and study areas, and expand the classification with new roof type classes. Since the success rate of transfer learning results is also satisfactory, we plan to fine-tune some other popular pre-trained CNNs with the same dataset and evaluate their performances as well.

ACKNOWLEDGEMENTS

This paper is part of the Ph.D. thesis research of Mehmet Buyukdemircioglu.

REFERENCES

Alidoost, F., Arefi, H., 2016. Knowledge based 3D building model recognition using convolutional neural networks from LiDAR and aerial imageries: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Prague, Czech Republic*, Vol. XLI-B3, part. XXIII, pp. 833-839, https://doi.org/10.5194/isprsarchives-XLI-B3-833-2016

Alidoost, F., Arefi, H. and Tombari, F., 2019. 2D image-to-3D model: knowledge-based 3D building reconstruction (3DBR) using single aerial images and convolutional neural networks (CNNs). *Remote Sensing*, 11(19), p.2219. https://doi.org/10.3390/rs11192219

Assouline, D., Mohajeri, N. and Scartezzini, J.L., 2017, October. Building rooftop classification using random forests for large-scale PV deployment. Proc. SPIE 10428, Earth resources and environmental remote Sensing/GIS Applications VIII (Vol. 10428, p. 1042806). International Society for Optics and Photonics. https://doi.org/10.1117/12.2277692

Axelsson, M., Soderman, U., Berg, A. and Lithen, T., 2018, April. Roof Type Classification Using Deep Convolutional Neural Networks on Low Resolution Photogrammetric Point Clouds from Aerial Imagery. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP)(pp.1293-1297). IEEE, https://doi.org/10.1109/ICASSP.2018.8461740

Biljecki, F. and Dehbi, Y., 2019. Raise the Roof: Towards Generating Lod2 Models Without Aerial Surveys Using Machine Learning, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-4/W8, 27–34, https://doi.org/10.5194/isprs-annals-IV-4-W8-27-2019

Bittner, K., Körner, M., Fraundorfer, F. and Reinartz, P., 2019. Multi-task cGAN for simultaneous spaceborne DSM refinement and roof-type classification. *Remote Sensing*, 11(11), p.1262. https://doi.org/10.3390/rs11111262

Buyukdemircioglu M, Kocaman S, Isikdag U., 2018. Semi-Automatic 3D City Model Generation from Large-Format Aerial Images. *ISPRS International Journal of Geo-Information*. 2018; 7(9):339. https://doi.org/10.3390/ijgi7090339

Castagno, J. and Atkins, E., 2018. Roof shape classification from LiDAR and satellite image data fusion using supervised learning. *Sensors*, 18(11), p.3960. https://doi.org/10.3390/s18113960

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In 2009 *IEEE conference on computer vision and pattern recognition* (pp. 248-255). https://doi.org/10.1109/CVPR.2009.5206848

Heipke, C. and Rottensteiner, F., 2020. Deep learning for geometric and semantic tasks in photogrammetry and remote sensing. *Geo-spatial Information Science*, 23(1), pp.10-19. https://doi.org/10.1080/10095020.2020.1718003

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

Ioffe, S. and Szegedy, C., 2015, June. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *In International conference on machine learning* (pp. 448-456). https://arxiv.org/abs/1502.03167v3

Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint*, https://arxiv.org/abs/1412.6980v9

Mohajeri, N., Assouline, D., Guiboud, B., Bill, A., Gudmundsson, A. and Scartezzini, J.L., 2018. A city-scale roof shape classification using machine learning for solar energy applications. *Renewable Energy*, 121, pp.81-93. https://doi.org/10.1016/j.renene.2017.12.096

Partovi, T., Fraundorfer, F., Azimi, S., Marmanis, D., & Reinartz, P., 2017. Roof Type Selection based on patch-based classsification using deep learning for high Resolution Satellite Imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives*, 42(W1), 653-657, https://doi.org/10.5194/isprs-archives-XLII-1-W1-653-2017

Qin, Y., Wu, Y., Li, B., Gao, S., Liu, M. and Zhan, Y., 2019. Semantic segmentation of building roof in dense urban environment with deep convolutional neural network: A case study using GF2 VHR imagery in China. *Sensors*, 19(5), p.1164. https://doi.org/10.3390/s19051164

Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., and Breitkopf, U., 2012. The ISPRS Benchmark on Urban Object Classification And 3D Building Reconstruction, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., I-3, 293–298*, https://doi.org/10.5194/isprsannals-I-3-293-2012

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J.D., Breitkopf, U. and Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS journal of photogrammetry and remote sensing*, 93, pp.256-271. http://dx.doi.org/10.1016/j.isprsjprs.2013.10.004

Safe Software, FME, 2021. https://www.safe.com/ (accessed on 24 April 2021).

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, *arXiv preprint*, https://arxiv.org/abs/1409.1556v6

Tan, M. and Le, Q., 2019, May. Efficientnet: Rethinking model scaling for convolutional neural networks. *In International Conference on Machine Learning*, (pp. 6105-6114). https://arxiv.org/abs/1905.11946