DWELLING EXTRACTION IN REFUGEE CAMPS USING CNN - FIRST EXPERIENCES AND LESSONS LEARNT

Omid Ghorbanzadeh*, Dirk Tiede, Zahra Dabiri, Martin Sudmanns & Stefan Lang

Department of Geoinformatics (Z_GIS), University of Salzburg, Schillerstrasse 30, 5020 Salzburg, Austria

KEY WORDS: (semi)-automated object-based image analysis (OBIA), convolutional neural network (CNN), camp dwellings extraction.

ABSTRACT:

There is a growing use of Earth observation (EO) data for support planning in humanitarian crisis response. Information about number and dynamics of displaced population in camps is essential to humanitarian organizations for decision-making and action planning. Dwelling extraction and categorisation is a challenging task, due to the problems in separating different dwellings under different conditions, with wide range of sizes, colour and complex spatial patterns. Nowadays, so-called deep learning techniques such as deep convolutional neural network (CNN) are used for understanding image content and object recognition. Although recent developments in the field of computer vision have introduced CNN networks as a practical tool also in the field of remote sensing, the training step of these techniques is rather time-consuming and samples for the training process are rarely transferable to other application fields. These techniques also have not been fully explored for mapping camps. Our study analyses the potential of a CNN network for dwelling extraction to be embedded as initial step in a comprehensive object-based image analysis (OBIA) workflow. The results were compared to a semi-automated, i.e. combined knowledge-/sample-based, OBIA classification. The Minawao refugee camp in Cameroon served as a case study due to its well-organised, clearly distinguishable dwelling structure. We use manually delineated objects as initial input for the training samples, while the CNN network is structured with two convolution layers and one max pooling.

1. INTRODUCTION

Up-to-date critical information products derived from very high resolution (VHR) Earth observation (EO) images have become one essential source of information in supporting humanitarian response, e.g. for the monitoring and management of refugee or internally displaced people (IDP) camps (Lang et al., 2015; Lang et al., 2017). The information derived from EO images includes amongst others the number and size of dwellings, dwelling type classification and derived population estimations (Spröhnle et al., 2014). Various achievements based on object-based image analysis (OBIA) workflows are documented in the literature, e.g. improving transferability of rule-sets (Tiede et al., 2013), challenges in operational mode (Füreder et al., 2014) or the integration of additional techniques like template-matching (Tiede et al., 2017). OBIA workflows rank high among the main strategies used in (semi-)automated camp analyses (Witmer, 2015; Lang et al., 2018). The accuracy and degree of automation of the dwelling extraction in refugee camps depends on various factors, such as image data quality, camp structure, weather conditions etc. (Tiede et al., 2013).

Recently, deep machine learning techniques and above all convolutional neural networks (CNN) have achieved higher accuracies in object detection compared to classical object detection methods. Conventional object detection methods are mainly based on the moving window techniques or fixed pixel arrangements by which the image is scanned in different scales. These object detection methods are mostly applied to distinct objects such as vehicles and airplanes (Zhang and Zhang, 2017; Deng et al., 2017). Currently there the community strives to use CNN networks based on labelled images for object detection (Dahmane et al., 2016). CNN networks are constructed by supervised machine learning, in which a training data set of labels is used to push learnable, i.e. adaptive, filters (feature extractors) to minimize a loss function (Yang, 2017). Recently, CNN networks have been used for various image analysis tasks in the remote sensing domain. A detailed review is provided by Zhu et al. (2017); examples include scene classification for high spatial resolution and aerial images (Hu et al., 2015; Othman et al., 2016; Han et al., 2017; Qayyum et al., 2017); remote-sensing image classification and object detection (Maggiori et al., 2017; Long et al., 2017; Radovic et al., 2017); so-called semantic segmentation (Long et al., 2015; Längkvist., 2016; Wang et al., 2017).

In this study, we trained a CNN for so-called semantic segmentation of dwellings. Labelled image patches of manually extracted objects were used as samples, obtained from an operational service for humanitarian mapping at the University of Salzburg, Department of Geoinformatics (Z_GIS). We used a World View 3 image captured in 2015 (4 bands; R-G-B-NIR, pansharpened spatial resolution of 0.5m) and split the study area into two different regions for training and testing (see Figure 1). We focused on three different target classes, namely *Tent I* (tunnel-shaped, bright tents), *Tent II* (rectangular shaped, bright tents) and *Larger Buildings* (supply infrastructure), as well as a class *Non-target Objects* comprising dark (i.e. traditional) dwellings, bare soil, vegetation, etc. The structured CNN was trained by objects of the target and non-target samples taken from the training region and implemented on the test region. The number of samples used for training and testing is presented in Table 1. Finally, the accuracy of the results was assessed against the manually delineated objects of the test region and compared to a (semi-automated OBIA approach.

https://doi.org/10.5194/isprs-archives-XLII-1-161-2018 | © Authors 2018. CC BY 4.0 License.

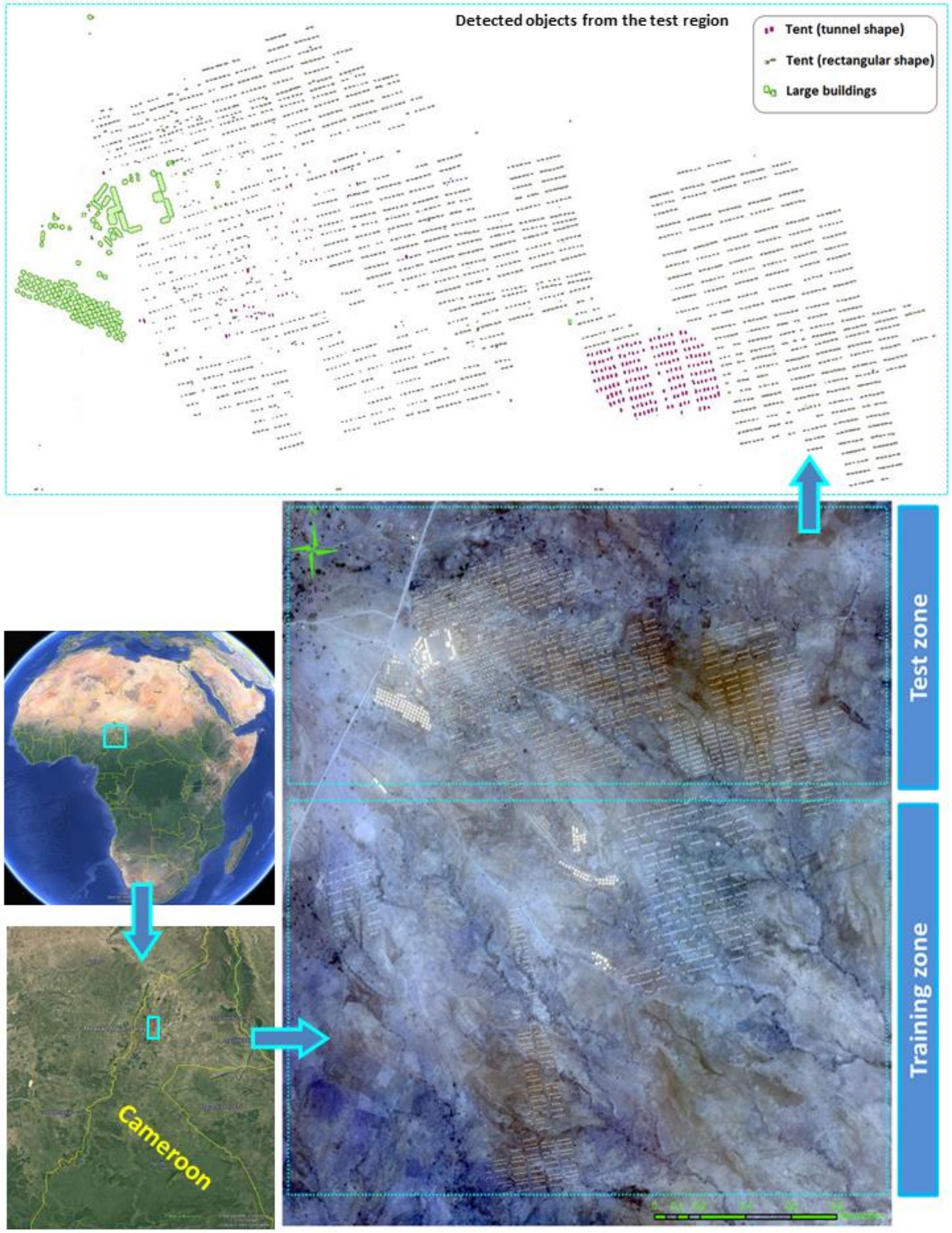

Figure 1. The case study area Minawao refugee camp situated in northern Cameroon (left), training and testing zones, and results of the CNN network for the testing area (right upper image).

2. METHODS: WORKFLOW

2.1 Deep convolutional neural network

A deep CNN is typically structured by multiple convolutional layers. Moreover, based on the user’s intended goal, other layers may be used, e.g., normalization layer, pooling layers, and fully connected layers (Cozzolino et al., 2017). A convolutional layer as the core of the CNN consists of different learnable filters. Pooling layers used for size reduction by the maximum or average value or other measurements. As pooling layers are a crucial part of biological visual systems, they are common in the CNN applications of the computer vision (Yang, 2017).

The window size of our input samples was set to 16×16 pixels by cross validating a variety of window sizes, including 12×12, 16×16, 18×18 and 32×32. As we fed the CNN network with the four-layer image, the sample patch had 16×16×4 units. We worked in Trimble’s eCognition software environment with the CNN implementation based on Google TensorFlow library. We generated the samples extracted from a layer containing all manually delineated objects of the training area. The number of our feature maps was 40, thus 16×16×4×40 different weights were trained during the first hidden layer. As a result, 40 feature maps within 12×12×1 units were obtained after convolution with a kernel size of 5. There is also a max pooling in the first hidden layer, which reduced the units to 6×6×1 in the same number of feature maps. The results forwarded to the second hidden layer as input data. Consequently, convolution with a kernel size of 3 led to 12 feature maps within 4×4×1 units. It should be noted that the kernel sizes and the number of feature maps were selected by us with attention of the camp situation, e.g. the quite small ratio of dwelling size vs. pixel size (see Figure 3).

In each training step, gradients for each weight is assessed, i.e. estimated using backpropagation. During this process, a statistical gradient descent function is used to optimize the weights. We choose a very small value for learning rate of 0.0001 because of the simplicity of our samples. Training steps and batch size were 5000 and 50 respectively. Batch size is the number of samples used as input data at each training step.

2.2 (semi-) automated object-based dwelling extraction

For comparison of the results we conducted a semi-automated, i.e. combined knowledge-/sample-based OBIA dwelling extraction for the same dwelling types was conducted. The approach combines OBIA elements with supervised classification techniques in a user-friendly interface for fast parameter selection (see Tiede et al., 2013). The following steps were performed. (1) Image segmentation of the area of interest and initial target class detection (bright dwellings) based on relative contrast difference of the initial segments compared to their surroundings. Brightness contrast in the blue band has been selected for the initial detection of bright dwellings types. (2) Then segments classified as initial bright dwellings were merged to image objects describing single dwellings (if dwellings are densely attached to each other, they are merged into larger objects containing more than one dwelling). (3) Third, a stratified supervised classification was performed on the target dwellings only, which allows the usage of only a few samples per dwelling class (here: ~ 10 samples per class were selected). A support vector machine (SVM) classifier has been has been used considering also spatial features (form and size) next to spectral information per object (mean and standard deviation of the 4 spectral bands); (4) Finally, after the differentiation of the initial dwelling types into the three dwelling classes, knowledge-based post processing is conducted automatically, to select only dwellings of at least 10 m² in size and remove outliers, which are not within the camp extent (based on dwellings density estimations).

In this workflow, the number of free parameters to be user defined included (i) segmentation parameters, (ii) a relative threshold for initial dwellings type detection and (iii) the selection of (few) training samples for the SVM classifier. The last step is significantly reduced, due to the stratified approach of initial target class detection and differentiation of classes only within the initial range of target objects. The result of this approach is represented in figure 2.

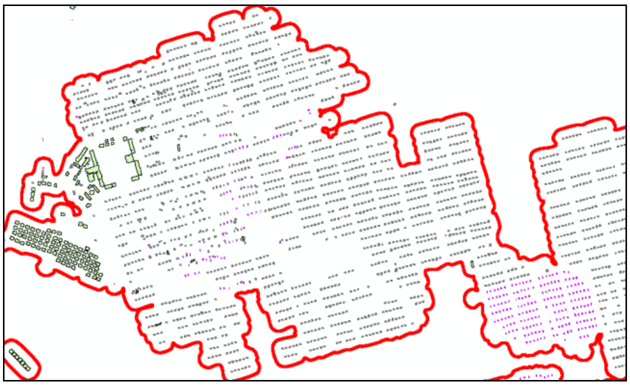

Figure 2. Subset of the results obtained from the OBIA approach, including the automatically derived camp extent (based on dwelling density estimations).

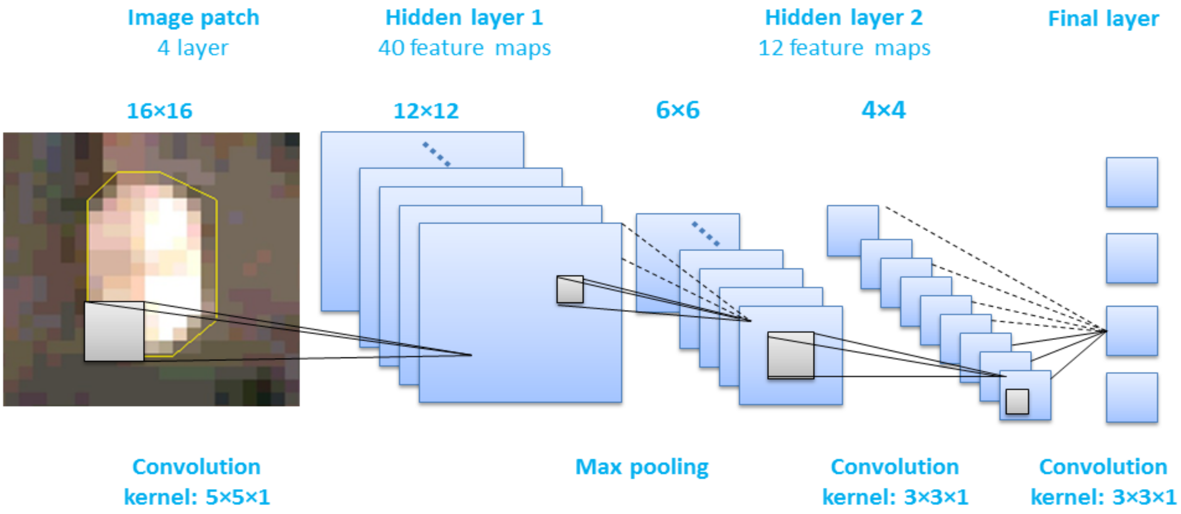

Figure 3. The structure of our CNN for mapping of dwellings in a refugee camp

3. RESULTS AND DISCUSSION

We used a threshold of 85% for extraction the objects from the resulted heat map of the CNN model (Figure 1). For the accuracy assessment, three different metrics were used: precision (P) was used to find how many detected objects were true. Recall (R) was used to find how many actual objects were detected. F1 measure was used to determine the balance between mentioned metrics (see Figure 4).

Precision = True Positives ⁄ ((True Positives + False Positives))

Recall = True Positives ⁄ ((True Positives + False Negatives))

F1 measure = 2 × (precision × recall) ⁄ ((precision + recall))

The accuracy values (P, R, and F1, see Table 1 & 2) of both approaches show more than 85% for the extraction of all three types of dwellings except of the P metric of the class *Tent I* by (semi)-automated OBIA which reaches 76%. For this class, although the F1 measure shows the same result of accuracy, there is a big difference between P and R metrics in the results of our two different methodologies. In the case of using CNN network for the *Tent I*, P and R metrics were almost the same. For the OBIA approach, the metric of R almost 20% more than the P metric (less detection of *Tent I* objects, but with a high confidence, i.e. less false negatives).
For the *large buildings*, the CNN network revealed a P measure of 100% which means this method could successfully detect all the objects of this type (both methods treated attached large dwellings as one large dwelling). However, the lesser value of R metric illustrates that the CNN network indicated more falsely classified objects as *large buildings*, whereas the OBIA method revealed a balanced result on a high accuracy level (98% / 94%).

Among three types of dwellings, the *Tent I* type (tunnel shape) was most difficult to be detected, while the other classes show very high accuracy values for both approaches. This might be due to the smaller amount of training samples (compared to the number of training samples for *Tent II*), or more variabilities in spatial context and more complex spatial structure of the dwellings (tunnel shape) in comparison to the large buildings or the rectangular dwelling type.

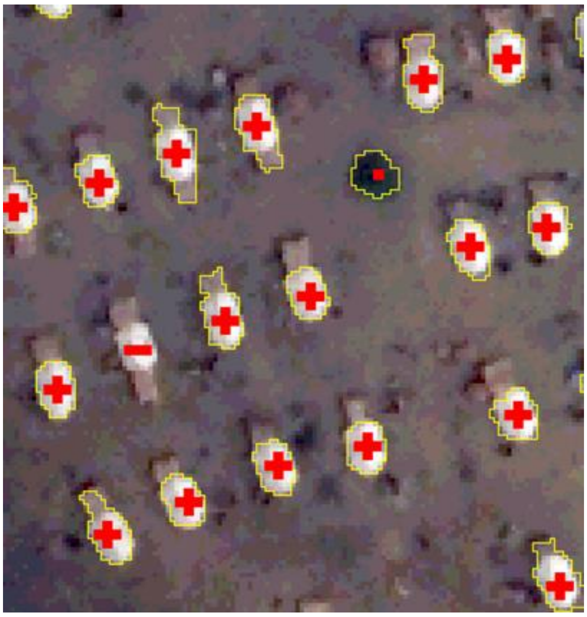

Figure 4. Subset of the results obtained from the model. The symbols (+), (-) and (.) are true positive (TP), false positive (FP) and false negative (FN) respectively.

Table 1. Accuracy results of CNN approach

Object	train	test	TP	FP	FN	P (%)	R (%)	F1 (%)
Tent I (tunnel shape)	675	297	219	41	37	84.2	85.5	85.2
Tent II (rectangular shape)	1644	2639	2458	129	52	95.0	97.9	96.3
Large buildings	535	121	106	0	15	100	87.6	93.3

Table 2. Accuracy results of object-based approach

Object	test	TP	FP	FN	P (%)	R (%)	F1 (%)
Tent I (tunnel shape)	297	221	67	9	76.7	96.0	85.2
Tent II (rectangular shape)	2639	2541	28	70	98.9	97.3	97.8
Large buildings	121	112	2	7	98.2	94.1	96.0

4. CONCLUSIONS

In this paper, we evaluated the potential of CNNs, as an alternative learning strategy for or an integral part in OBIA workflows. We focused on the issue of improving the detection and extraction of dwelling types in refugee camps based on VHR data. The results were compared with an established (semi)-automated OBIA approach.

Both approaches showed quite high accuracy values for the extraction of the selected three different dwelling types. The two approaches differ by the number of training samples and the number of free parameters to be specified for transferability to other time stamps and/or areas. While the CNN approach needs a multiple of samples in the initial training phase, the transferability – once a proper CNN is trained – is expected to be high, at least to similar sites (Penatti et al. 2015; Yosinski et al. 2014). However, transferability of the CNNs to areas covered by different sensors or atmospheric conditions or more complex camp structures also highly depends on many un-biased samples for training and supervised learning (LeCun, Bengio, and Hinton 2015). This is difficult to achieve in the case of refugee camps (sample scarce situation). Another problem we faced using the CNN approach, was the quite small object size under consideration compared to the image resolution. The best suited window size of the training samples was selected as 16×16 pixels, which covers for small objects (e.g., trees and small dwellings) more than one object in a single window. On the other hand, if smaller window sizes are selected, no sufficient object context is taken into consideration for the convolution and pooling operations. Maybe other approaches like scene detection rather than object detection for the smaller objects could be a solution. The semi-automated OBIA shows also a very good performance on the single test site. The approach is quite fast to implement, since only a few parameters need to be defined (see section 2.2), but adaptation is needed for every new site.

Further research will focus on the scalability of the two approaches regarding:

- Other time stamps or different sensors of the same refugee camp
- Improving the CNN by integrating training samples from different refugee camps and testing the transferability to other camps
- Comparison of the performance of both approaches if scaled to larger or different sites with respect to accuracy and speed, manual intervention etc.

It is then envisaged to integrate the CNN probability layer as input for a subsequent object-based analysis, to increase the accuracy and decrease the number of free parameters of existing, knowledge-based rule-sets in this time-critical application domain.

ACKNOWLEDGEMENTS

This work has been co-funded by the Austrian Research Promotion Agency (FFG) under the Austrian Space Application Programme (ASAP) within the project EO4HumEn+ (contract no: 854041) and partly funded by the Austrian Science Fund (FWF) through the GIScience Doctoral College (DK W 1237-N23).

REFERENCES

Cozzolino, D., Di Martino, G., Poggi, G., & Verdoliva, L., 2017. A fully convolutional neural network for low-complexity single-stage ship detection in Sentinel-1 SAR images. In Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International (pp. 886-889). IEEE.

Dahmane, M., Foucher, S., Beaulieu, M., Riendeau, F., Bouroubi, Y., & Benoit, M., 2016. Object detection in pleiades images using deep features. In Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International (pp. 1552-1555). IEEE.

Deng, Z., Sun, H., Zhou, S., Zhao, J., Lei, L., & Zou, H., 2017. Fast multiclass object detection in optical remote sensing images using region based convolutional neural networks. In Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International (pp. 858-861). IEEE.

Han, X., Zhong, Y., Zhao, B., & Zhang, L., 2017. Scene classification based on a hierarchical convolutional sparse auto-encoder for high spatial resolution imagery. International Journal of Remote Sensing, 38(2), 514-536.

Hu, F., Xia, G. S., Hu, J., & Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing, 7(11), 14680-14707.

Lang, S., Füreder, P. and Rogenhofer, E., 2018. Earth Observation for Humanitarian Operations. In Yearbook on Space Policy 2016 (pp. 217-229). Springer, Cham.

Lang, S., Schoepfer, E., Zeil, P. and Riedler, B., 2017 Earth observation for humanitarian assistance. GI Forum - Journal for Geographic Information Science, 1/2017, pp. 157-165.
2017

Lang, S., Füreder, P., Kranz, O., Card, B., Roberts, S. and Papp, A., 2015. Humanitarian emergencies: causes, traits and impacts as observed by remote sensing. in Thenkabail, P., (ed.) Remote Sensing Handbook,New York: Taylor and Francis. pp. 483-512.

Längkvist, M., Kiselev, A., Alirezaie, M., & Loutfi, A, 2016. Classification and segmentation of satellite orthoimagery using convolutional neural networks. Remote Sensing, 8(4), 329.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep learning." nature 521 (7553):436.

Long, Y., Gong, Y., Xiao, Z., & Liu, Q., 2017. Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks. IEEE Transactions on Geoscience and Remote Sensing, 55(5), 2486-2498.

Long, J., Shelhamer, E., & Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3431-3440).

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. IEEE Transactions on Geoscience and Remote Sensing, 55(2), 645-657.

Othman, E., Bazi, Y., Alajlan, N., Alhichri, H., & Melgani, F., 2016. Using convolutional features and a sparse autoencoder for land-use scene classification. International Journal of Remote Sensing, 37(10), 2149-2167.

Penatti, Otávio AB, Keiller Nogueira, and Jefersson A dos Santos. 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition workshops.

Qayyum, A., Malik, A. S., Saad, N. M., Iqbal, M., Faris Abdullah, M., Rasheed, W., ... & Bin Jafaar, M. Y., 2017. Scene classification for aerial images based on CNN using sparse coding technique. International Journal of Remote Sensing, 38(8-10), 2662-2685.

Radovic, M., Adarkwa, O., & Wang, Q., 2017. Object Recognition in Aerial Images Using Convolutional Neural Networks. Journal of Imaging, 3(2), 21.

Spröhnle, K., Tiede, D., Schoepfer, E., Füreder, P., Svanberg, A., & Rost, T., 2014. Earth observation-based dwelling detection approaches in a highly complex refugee camp environment A comparative study. Remote Sensing, 6(10), 9277-9297.

Tiede, D., Füreder, P., Lang, S., Hölbling, D., Zeil, P., 2013. Automated Analysis of Satellite Imagery to provide Information Products for Humanitarian Relief Operations in Refugee Camps – from Scientific Development towards Operational Services. PFG Photogramm. - Fernerkundung - Geoinf. 2013, 185–195. doi:10.1127/1432-8364/2013/0169.

Tiede, D., Krafft, P., Füreder, P., & Lang, S., 2017. Stratified Template Matching to Support Refugee Camp Analysis in OBIA Workflows. Remote Sensing, 9(4), 326.

Wang, H., Wang, Y., Zhang, Q., Xiang, S., & Pan, C., 2017. Gated Convolutional Neural Network for Semantic Segmentation in High-Resolution Images. Remote Sensing, 9(5), 446.

Witmer, F.D.W., 2015. Remote sensing of violent conflict: eyes from above. Int. J. Remote Sens. 36, 2326–2352. doi:10.1080/01431161.2015.1035412

Yang, H. L., Lunga, D., & Yuan, J., 2017. Toward country scale building detection with convolutional neural network using aerial images. In Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International (pp. 870-873). IEEE.

Yosinski, Jason, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks? Paper presented at the Advances in neural information processing systems.

Zhang, L., & Zhang, Y., 2017. Airport detection and aircraft recognition based on two-layer saliency model in high spatial resolution remote-sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(4), 1511-1524.

Zhu, Xiao X., Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. 2017. “Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources.” IEEE Geoscience and Remote Sensing Magazine 5 (4): 8–36. doi:10.1109/MGRS.2017.2762307.