Testing Transferability of Deep-Learning-Based Dwelling Extraction in Refugee Camps

Getachew Workineh Gella[1], Lorenz Wendt[1], Stefan Lang[1], Andreas Braun[1,2], Dirk Tiede[1], Barbara Hofer[1], Yunya Gao[1], Barbara Riedler[1], Ahmad Alobaidi[3] and Gina Marciela Schwendemann[3]

[1]University of Salzburg, Austria
[2]University of Tübingen, Germany
[3]Spatial Services, Salzburg, Austria

Abstract

For effective humanitarian response in refugee camps, reliable information concerning dwelling type, extent, surrounding infrastructure, and respective population size is essential. As refugee camps are inherently dynamic in nature, continuous updating and frequent monitoring is time and resource-demanding, so that automatic information extraction strategies are very useful. In this ongoing research, we used labelled data and high-resolution Worldview imagery and first trained a Convolutional Neural Network-based U-net model architecture. We first trained and tested the model from scratch for Al Hol camp in Syria. We then tested the transferability of the model by testing its performance in an image of a refugee camp situated in Cameroon. We were using patch size 32, at the Syrian test site, a Mean Area Intersection Over Union (MIoU) of 0.78 and F-1 score of 0.96, while in the transfer site, MIoU of 0.69 and an F-1 score of 0.98 were achieved. Furthermore, the effect of patch size and the combination of samples from test and transfer sites are investigated.

Keywords: deep learning, dwelling extraction, refugee camp, transferability, U-net

1   Introduction

Humanitarian aid organizations, human rights groups, and concerned parties working in emergency response need accurate and reliable information related to the camp extent, dwelling type, number, and structure of dwellings of camps of refugees and internally displaced persons (IDP). Especially in the absence of exact population numbers, estimating the number of people in need of relief from these proxies is a viable option. Earth observation is routinely used to this end when information collection on the ground would be too time-consuming or dangerous (e.g., Lang and Füreder, 2015; Bjorgo, 2000; Lang et al., 2020).

Despite recent advances in computer vision and particularly deep learning for information extraction from satellite images (Ma et al., 2019; Li et al., 2020), manual digitization or at least extensive clean-up of automatically extracted dwelling features is still required. Quinn et al. (2018), who demonstrated the potential of deep learning for automatic dwelling counting,

noted problems of transferability of models emanated from differences in sensors and inherent characteristics of dwelling structures. To overcome these challenges, Ghorbanzadeh et al. (2018) investigated combining object-oriented and deep learning approaches for dwelling extraction to improve the transferability of classifiers from one satellite scene to another.

The large number of dwellings we extracted from satellite images of refugee and IDP camps in an operational setting for a major international humanitarian organization over the past years now enables us to experiment on the optimal combination of sample-based machine learning / deep learning techniques, prior-knowledge based machine teaching methods and an optimized selection of samples. As part of ongoing work, this study focuses on the spatial transferability of deep learning-based dwelling extraction. More specifically, it has the following contributions: firstly, it tests the U-net architecture (Ronneberger, Fischer, & Brox, 2015) for dwelling extraction in two refugee camps; secondly, it investigates the impact of patch size on model capability for dwelling extraction; thirdly, it explores the transferability of the model trained at one site to another refugee camp situated in a different geographic region; finally, it compares the performance of U-net segmentation network trained on local samples, samples from a different camp setting, and network using a combined set of samples.
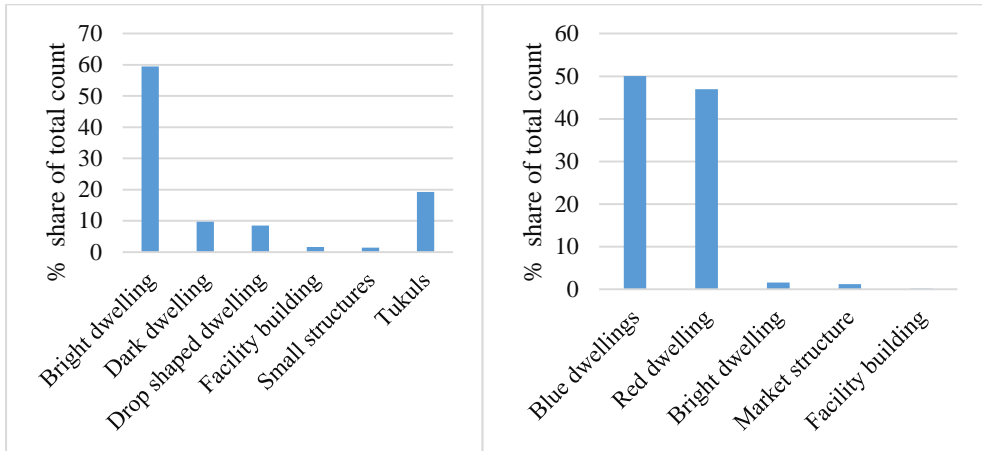
2   Methodology

2.1 The test sites

The study is based on the two refugee camps in Al Hol, Syria, and Minawao, Cameroon. The two camps are characterized by a large fraction of standard-issue shelters as used by UNHCR and other organizations but are located in a different part of the world. Therefore, these sites allow testing the generalization capability of the trained model from one geographic setting to the other. Al Hol consisted of approximately 19,396 dwellings at the time of the investigation. It has a total area of 289 hectares with dwelling density ranging from below 15 dwellings per hectare at the outskirts of the camp towards 108 dwellings per hectare in the inner parts. Almost 97% of the dwellings are standard-issue tents (UNHCR, 2016). It has experienced a high population influx (Neil, 2020; REACH, 2020). Minawao has a camp area of 623 hectares with a total of 16,601 dwellings with a dwelling density of 1-2 dwellings per hectare towards 180 dwellings per hectare. The camps have different dwellings, which include small structures (~2-5 m²) towards large facility structures (~102 m²).

2.2 Data and sample generation

We used a WorldView-2 image with a spatial resolution of 0.5 meters for Al Hol acquired on 27 April 2020, and a WorldView-3 image with a resolution of 0.3 meters for Minawao acquired on 3 June 2016. The Worldview-3 image is resampled to Worldview-2 resolution. The labelled vector data used in this study were generated as part of ongoing operational humanitarian service by combining object-based image analysis (OBIA) and subsequent manual digitization with proper post-processing operations. For labelled data, a qualitative check is made for completely missing polygons and the presence of systematic positional shifts. Then these

vector files were converted to binary raster tiles (dwelling and non-dwelling), irrespective of dwelling type (Figure 1).

Figure 1: Dwelling types in Minawao (right) and Al Hol (left)

To train, validate and test the model, the test site in Syria is partitioned into training and test area. During this partitioning, some part of the area surrounding the camp is also included in training and validation samples to reduce model confusion on unseen features from the surrounding environment during the prediction phase both onsite and at the transferability test site. As patch size influences the variation of Fully Convolutional Network (FCN) model performance (Hamwood, Alonso-Caneiro, Read, Vincent, & Collins, 2018), from a training area, three sample sets with a mutually exclusive patch size of 32 by 32, 64 by 64, and 128 by 128 pixels were generated. To train the model, these sample sets are randomly partitioned into training and validation samples with a ratio of 0.8 and 0.2. Testing of the trained model is done with unseen samples taken from camp parts partitioned to test areas that are not included in the training and validation samples and also in the transferability test site. To see the impacts of training on mixed samples from test and transfer sites, we have also generated some samples from Cameroon and mixed them with samples from Al Hol Syria.

2.3 The model and training process

We used the U-net model architecture (Ronneberger et al., 2015). The model is a family of Convolutional Neural Network (CNN), which are FCN architectures (Long, Shelhamer, & Darrell, 2015) and was reported robust in many problems that need semantic segmentation, like medical (Ibtehaz & Rahman, 2020) and aerial (Ivanovsky, Khryashchev, Pavlov, & Ostrovskaya, 2019) image segmentation. The model mainly works with an encoder-decoder architecture where the contracting encoder extracts abstract features from an image while the expanding decoder block reconstructs segmented features (Ibtehaz & Rahman, 2020). Both encoder and decoder blocks are constructed from a stack of convolution, pooling, and activation layers with skip connections between the decoder and encoder blocks at some stage.

For the current study, decoder and encoder blocks used Rectified Linear Unit (ReLU) activation (Nair & Hinton, 2010; Zeiler et al., 2013). Weight updating and feature learning are done using a categorical cross-entropy loss and stochastic gradient descent (SGD) optimizer (Zaheer & Shaziya, 2019) with a learning rate of 0.01. To prevent model overfitting (Ying, 2019) and reduce unnecessary computational time, an early stopping strategy with a patience of 10 epochs taking validation cross-entropy loss as a target monitoring metric has been implemented. Final dwelling structure presence probabilities are predicted by using a softmax activation, which is further converted to hard binary classes of dwelling features. To see the pixel-wise overall model performance, the F-1 score is used. Given that pixel-based metrics yield relatively inflated values in segmentation tasks, especially in unbalanced samples, an object-based metric, Mean Intersection over Union (MIoU) (Atiqur & Yang, 2001), was used. This metric evaluates the spatial (geometric) congruency of predicted and reference objects where a perfect match gives MIoU of 1 while complete disjoint MIoU of 0.
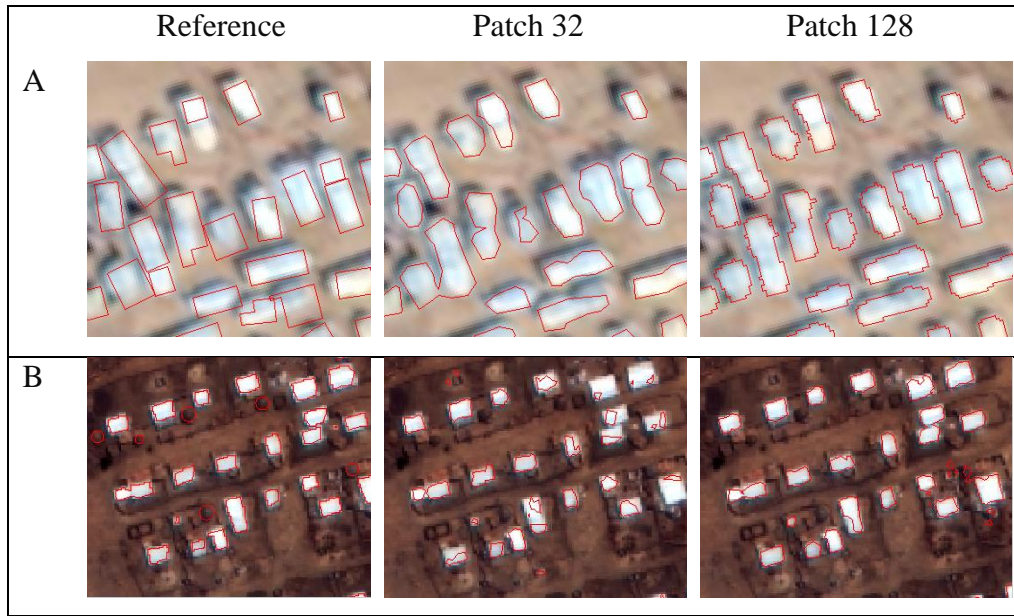
3   Results

Table 1 shows dwelling extraction accuracy metrics for a model in Syria trained on Syria, Cameroon trained on Syria, and Syria and Cameroon trained on samples from both sites combined for different patch sizes. The accuracy is calculated over all dwellings, irrespective of their type (Figure 1). As model prediction is made on an image that includes areas outside of the camp, false positives outside the camp area are masked out before the calculation of evaluation metrics.

Obtained results show variation as per utilized patch size for model training. In terms of MIoU metrics, 78.2% areal fit is achieved when trained with patch sizes of 128 pixels by 128 in Syria and 79.5% in Cameroon. The combination of samples from the transferability test site to train the model has yielded MIoU values almost similar to outputs from the model trained on samples from a single site. The model transferability metric is also varying as per input patch size, where a model trained with a patch size of 128 performed better. This also holds the same for a model trained with samples mixed from model training and transferability site (Cameroon).

Table 1: Accuracy metrics for experiments

Test site	Syria			Cameroon		
Patch size	32	64	128	32	64	128
F-1 score	0.964	0.961	0.963	0.985	0.983	0.989
Mean IoU	0.781	0.776	0.782	0.691	0.758	0.795
	Combined samples			Combined samples		
F-1 score	0.963	0.961	0.963	0.986	0.983	0.988
Mean IoU	0.778	0.777	0.779	0.726	0.763	0.770

Figures 1 and 2 show a visual comparison of classification and reference data. In the first test site, except for lack of crispness at the edge of dwellings, it has extracted dwellings with good completeness. In the transferability test site (Cameroon), relatively, there are some dwellings flagged by a model as false negatives. Combining samples from both sites during model training has not added any accuracy improvements except for patch size 32 where combining samples resulted in a 3.5% improvement of MIoU.
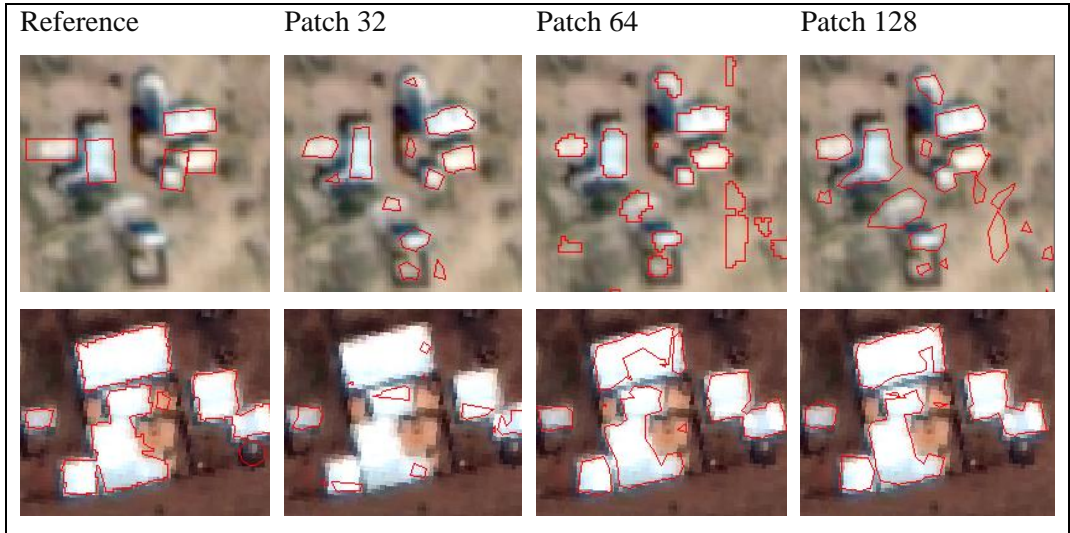

Figure 2: Dwelling extraction results in A) from Syrian test site and B) from Cameroon transferability test site

4   Discussion

Though the model fails to detect small and circular buildings made from natural materials, called tukuls (UNHCR, 2016) in the transferability test (Figure 1B), overall, the obtained results were in good agreement with segmentation results reported for building extraction (Rastogi, Bodani, & Sharma, 2020) and tent detection in refugee camps (Kahraman, Ates, & Kucur Ergunay, 2013). Current results were slightly better than segmentation results reported using the CNN with OBIA approach (Ghorbanzadeh et al., 2018), which achieved F-1 scores of 85.2%, 96.3%, and 93.3% for tunnel-shaped, rectangular, and large buildings, respectively. Variations in findings could also be attributed to differences in model architecture, patch size and details included in dwelling types. Increasing patch size has resulted in a reduction of false negatives within the dwelling blocks with respective trade-offs, including some reflective features like roads and bare land as false positives, especially in areas outside of the dwelling camp. Contrary to this, when the patch size is reduced, the model fails to properly segment larger and linearly attached dwelling structures (area >280 km2). This is especially prevalent in

the transferability test site. Using a different segmentation model, Ergunay et al. (2013) have also reported the same impact of window size on the proper segmentation of tents. It should be noted that though the dwellings structures are extracted with good performance, the segmentation is not exact at the edges. The samples for this study had been produced in a time-constraint operational setting, where outlining the dwellings precisely was not essential. At some blocks, even segmentation outputs have more precise outlines than reference polygons (Figure 1A). We assume that somewhat unsharp outlines in the samples resulted in a less-than-perfect segmentation, which also affects evaluation metrics.

Figure 3: Confusion of background features with dwelling structures in Syria (first row) and patch wise variation of larger structure extraction (lower row)

5 Conclusion

Though the deep learning-based U-net model showed good overall performance, not all types of all dwelling structures were similarly well extracted. Structures that have good contrast with the background (bright and drop-shaped dwellings) are well extracted, while those with poor contrast (tukuls and small dark structures) were not well extracted. The model's capability to extract features varies with the patch sizes used for model training, especially to segregate background features with resembling reflectance characteristics. The model trained with a larger patch size can extract features in the transferability test site. In this study, except with patch size 32, the combination of samples from both sites has not changed the model performance. To get a more robust model that can universally extract dwelling structures, we plan further research with different pre-processing strategies that can enhance low contrast dwelling from its background, test segmentation models, and further detailed mapping of types of dwelling structures in the camp.

Acknowledgments

References

Atiqur, R., & Yang, W. (2001). Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation. *International Symposium on Visual Computing ISVC 2016: Advances in Visual Computing*, 234–244.

Bjorgo, E. (2000). Using very high spatial resolution multispectral satellite sensor imagery to monitor refugee camps. *International Journal of Remote Sensing*, *21*(3), 611–616. https://doi.org/10.1080/014311600210786

Ghorbanzadeh, O., Tiede, D., Dabiri, Z., Sudmanns, M., & Lang, S. (2018). Dwelling extraction in refugee camps using CNN - First experiences and lessons learned. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(1), 161–166. https://doi.org/10.5194/isprs-archives-XLII-1-161-2018

Hamwood, J., Alonso-Caneiro, D., Read, S. A., Vincent, S. J., & Collins, M. J. (2018). Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers. *Biomedical Optics Express*, *9*(7), 3049. https://doi.org/10.1364/boe.9.003049

Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, *121*, 74–87. https://doi.org/10.1016/j.neunet.2019.08.025

Ivanovsky, L., Khryashchev, V., Pavlov, V., & Ostrovskaya, A. (2019). Building detection on aerial images using U-NET neural networks. *Conference of Open Innovation Association, FRUCT*, *2019-April*, 116–122. https://doi.org/10.23919/FRUCT.2019.8711930

Kahraman, F., Ates, H. F., & Kucur Ergunay, S. S. (2013). Automated Detection Of Refugee / IDP TENTS FROM Satellite Imagery Using Two-Level Graph Cut Segmentation. *CaGIS/ASPRS Fall Conference*, (September 2015). San Antonio, Texas.

Lang, S., Füreder, P., Riedler, B., Wendt, L., Braun, A., Tiede, D., … Hochschild, V. (2020). Earth observation tools and services to increase the effectiveness of humanitarian assistance. *European Journal of Remote Sensing*, *53*(sup2), 67–85. https://doi.org/10.1080/22797254.2019.1684208

Li, X., Liu, B., Zheng, G., Ren, Y., Zhang, S., Liu, Y., … Wang, F. (2020). Deep-learning-based information mining from ocean remote-sensing imagery. *National Science Review*, *7*(10), 1584–1605. https://doi.org/10.1093/nsr/nwaa047

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 431–440. https://doi.org/10.1109/CVPR.2015.7298965

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 152, pp. 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, 807–814.

Neil, S. (2020). The Al Hol camp in Northeast Syria: health and humanitarian challenges. *BMJ Global Health*, *5*(e002491). https://doi.org/doi:10.1136/bmjgh-2020-00249

Quinn, J. A., Nyhan, M. M., Navarro, C., Coluccia, D., Bromley, L., & Luengo-Oroz, M. (2018). Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128). https://doi.org/10.1098/rsta.2017.0363

Rastogi, K., Bodani, P., & Sharma, S. A. (2020). Automatic building footprint extraction from very high-resolution imagery using deep learning techniques. *Geocarto International*. https://doi.org/10.1080/10106049.2020.1778100

REACH. (2020). *Camp Profile: Al HolAl-Hasakeh governorate, Syria October 2020*. Retrieved from https://reliefweb.int/sites/reliefweb.int/files/resources/Al-Hol_camp-profile_October-2020_FINAL.pdf

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9351, pp. 234–241). https://doi.org/10.1007/978-3-319-24574-4_28

UNHCR. (2016). *SHELTER DESIGN CATALOGUE* (p. 68). p. 68. Retrieved from https://cms.emergency.unhcr.org/documents/11982/57181/Shelter+Design+Catalogue+January+2016/a891fdb2-4ef9-42d9-bf0f-c12002b3652e

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*(2). https://doi.org/10.1088/1742-6596/1168/2/022022

Zaheer, R., & Shaziya, H. (2019). A Study of the Optimization Algorithms in Deep Learning. *Proceedings of the 3rd International Conference on Inventive Systems and Control, ICISC 2019*, (Icisc), 536–539. https://doi.org/10.1109/ICISC44355.2019.9036442

Zeiler, M. D., Ranzato, M., Monga, R., Mao, M., Yang, K., Le, Q. V., … Hinton, G. E. (2013). On rectified linear units for speech processing. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 3517–3521. https://doi.org/10.1109/ICASSP.2013.6638312