AGILE: GIScience Series, 3, 36, 2022. https://doi.org/10.5194/agile-giss-3-36-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.

Assessing the Influences of Band Selection and Pretrained Weights on Semantic-Segmentation-Based Refugee Dwelling Extraction from Satellite Imagery

Yunya Gao[1], Getachew Workineh Gella[1], and Nianhua Liu[2]

[1] Christian Doppler Laboratory for geospatial and EO-based humanitarian technologies (GEOHUM), Department of Geoinformatics – Z_GIS, Paris Lodron University of Salzburg, Salzburg, Austria
[2] Department of Geoinformatics – Z_GIS, Paris Lodron University of Salzburg, Salzburg, Austria

Correspondence: Yunya Gao (yunya.gao@plus.ac.at)

Abstract. This research assessed the influences of four band combinations and three types of pretrained weights on the performance of semantic segmentation in extracting refugee dwelling footprints of the Kule refugee camp in Ethiopia during a dry season and a wet season from very high spatial resolution imagery. We chose a classical network, U-Net with VGG16 as a backbone, for all segmentation experiments. The selected band combinations include 1) RGBN (Red, Green, Blue, and Near Infrared), 2) RGB, 3) RGN, and 4) RNB. The three types of pretrained weights are 1) randomly initialized weights, 2) pretrained weights from ImageNet, and 3) weights pretrained on data from the Bria refugee camp in the Central African Republic). The results turn out that three-band combinations outperform RGBN bands across all types of weights and seasons. Replacing the B or G band with the N band can improve the performance in extracting dwellings during the wet season but cannot bring improvement to the dry season in general. Pretrained weights from ImageNet achieve the best performance. Weights pretrained on data from the Bria refugee camp produced the lowest IoU and Recall values.

Keywords. Remote sensing, refugee dwellings, semantic segmentation, band selection, pretrained weights.

1 Introduction

1.1 Background

Sustainable Development Goals (SDGs) 2, 3, 6, and 7 emphasize the significance of distributing adequate living resources and health care services to refugees and their host countries based on the commitment "Leave No One Behind" (UNHCR, 2020). Population estimation of refugees in need is essential for logistics planning of the above resources during humanitarian operations (Çelik et al., 2012). However, it is usually difficult to collect such information in the field during conflicts. High-quality and updated footprints of refugee dwellings from satellite imagery could be beneficial for refugee population estimation (Checchi et al., 2013; Spröhnle et al., 2014), and thus, help achieve the related SDGs.

1.2 Related work

Deep learning approaches, especially Convolutional Neural Networks (CNN), have attracted researchers' attention for remote-sensing-based refugee-dwelling extraction in the last five years. Ghorbanzadeh et al. (2018) designed a shallow CNN model to extract refugee dwellings in the Minawao refugee camp. They trained the model from scratch based on four spectral bands (RGBN) of WorldView imagery. The results prove CNN has a high potential in this extraction task from Very High Spatial Resolution (VHSR) satellite imagery. Ghorbanzadeh et al. (2021) further combined the designed CNN with Object-Based Image Analysis (OBIA), which reveals the potential of combining CNN and expert knowledge for this task. Quinn et al. (2018) applied a Mask-RCNN model pretrained on the ImageNet dataset (Jia Deng et al., 2009) to extract dwellings in thirteen refugee camps. The model was trained with RGB bands of Google Earth imagery. Lu & Kwan (2020) compared the performance of two shallow CNN models, a deep fully CNN (FCN) model based on VGG16, and a Mask-RCNN model with ResNet-50 as a backbone in extracting refugee dwellings near Syria-Jordan border. Both the FCN model and Mask-RCNN

model were fine-tuned based on pretrained weights from ImageNet. The results turn out that the FCN model outperforms the other four models. Wickert et al. (2021) chose a Faster-RCNN model pretrained on the COCO dataset (Lin et al., 2014) to count dwelling numbers in nine refugee camps based on RGB bands of Google Earth imagery. Tiede et al. (2021) selected an untrained Mask-RCNN model to extract built-up structures in Sudan based on RGBN bands of Pléiades-1A satellite imagery. Gella et al. (2022) applied a Mask-RCNN model pretrained on the COCO dataset based on RGB bands of WorldView data.

Based on findings from Lu & Kwan (2021), this research chose a semantic segmentation model for all experiments. Semantic segmentation algorithms can assign a label to each pixel in an image and produce a fine-grained delineation of target objects with embedded spatial information (Borba et al., 2021). We have test multiple semantic segmentation models during the preliminary stage. Eventually, we chose U-Net with VGG16 as a backbone for all segmentation experiments due to its effectiveness and efficiency. Besides, U-Net is one of the most popular architectures for detecting built-up structures from satellite imagery (Ansari et al., 2020; Jung et al., 2021; Li et al., 2019).

Most semantic segmentation models are adapted from deep CNN models pretrained on large image classification datasets such as ImageNet which consists of more than one million labelled images (Kemker et al., 2018). Using pretrained weights (or parameters) from large datasets is essential because most deep CNN models have millions of parameters. For example, VGG16 has around 138 million parameters (Simonyan & Zisserman, 2015). Limited annotated label data in remote sensing domains are usually incapable of computing proper settings for randomly initialized weights (Kemker et al., 2018). Therefore, choosing proper pretrained weights can play an important role in this extraction task. However, this topic has not been discussed.

Furthermore, for multispectral satellite imagery with more than three RGB bands, band selection is significant before feeding data to CNN models (Kemker et al., 2018). Dixit et al. (2021) compared the performance of a semantic segmentation model (Dilated-ResUnet) under three datasets, 1) RGB bands, 2) NRGB bands, and 3) NRG bands of Sentinel-2 imagery. They found the dataset merged by NRG bands outperforms the other two datasets. This finding inspires us to assess the influences of the band selection for refugee dwelling extraction tasks from VHSR satellite imagery.

1.3 Research problem

To the best of our knowledge, it is still unknown that band combination performs best for refugee dwelling extraction. Besides, it is unknown whether seasonal changes can influence the performance of various band combinations. This research aims to fill this gap by testing the performance of four band combinations (RGB, RGN, RNB and RGBN) in extracting dwellings in the Kule refugee camp in Ethiopia under a dry season and a wet season. Additionally, we tested the influences of pretrained weights by comparing randomly initialized weights (RIW), pretrained weights from ImageNet, and weights trained on data of the Bria refugee camp in the Central African Republic (CAR). The outcomes of this research may shed light on the selection of bands and weights for similar tasks in the future.

2 Methodology

2.1 Data preparation Data and Software Availability

The Kule refugee camp, located in the Gambella region, Ethiopia, was opened in 2014 in response to the major refugee influx from South Sudan and was fully occupied in 2016 (UNHCR, 2020a). Bria refugee camp is located in eastern CAR. The brutal attacks caused by religious conflicts displaced over 40000 people in 2017 (Médecins Sans Frontières, 2018). Fig. 1 presents examples of dwellings in the two camps. We can observe that the appearances of dwellings and background are different across two camps and two seasons.

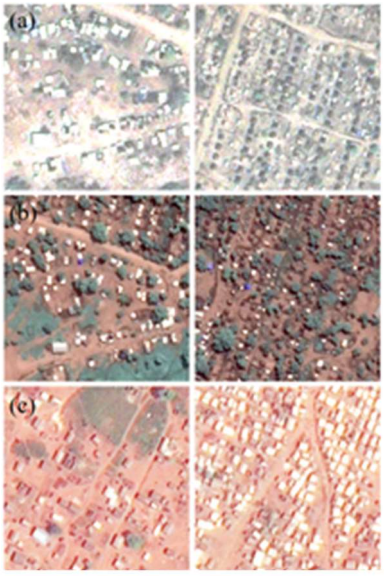

Figure 1. Examples of dwellings in the Kule refugee camp during the dry season (a) and the wet season (b), and in the Bria refugee camp (c).

We chose satellite imagery from the Pléiades-1 sensor with a resolution pandsharpened to 0.5m in GeoTIFF format for both camps. Considering the area of refugee dwellings in the two camps mainly ranges from 8m^2 to 50m^2, the original resolution (2m) makes models incapable of detecting small dwellings. The Kule imagery of the dry season and the wet season was retrieved on 24 March 2017 and 22 June 2018 respectively. We use binary classes that are “built-up structures” and “background” in

label data. The label data were produced by OBIA and post-processed by manual correction (Lang et al., 2020). The testing label data were manually annotated and checked by two experts in ArcGIS 10.7 software. The polygon label data were converted to GeoTIFF format with the same resolution. We eventually created 8286 training patches and 921 validation patches in a shape of (128, 128) pixels (Gella et al., 2021) with an overlap of 32 pixels, 612 testing patches without any overlap. The data of the Bria camp were processed in the same way above. 6568 patches were produced to create initial weights for Kule cases.

2.2 Architecture and model set-up

U-Net was firstly developed for biomedical image segmentation (Ronneberger et al., 2015), which follows an encoder-decoder structure. The encoder path is designed to capture features of input images. The decoder path is the symmetric expansion of the encoder path, which could help enable precise localization. It requires no dense or fully connected layers, and thus, can render the learning process in an end-to-end fashion. VGG16 architecture wined ILSVRC in 2014 (Simonyan & Zisserman, 2015). We implemented the model based on the Segmentation Model Python library (Yakubovskiy, 2019). The brief structure of the model could be found in Fig. 2. Besides, we selected balanced cross-entropy loss as a loss function due to the high imbalance between the two classes (Zhou et al., 2017). The percent of pixels of built-up structures is only round 2%.

For other hyperparameters, the batch size is 32. Adam optimizer was chosen due to its fast speed in convergence (Bock et al., 2018). The model was trained by 200 epochs with $4x10^{-4}$ as an initial learning rate and $2x10^{-6}$ as a decay rate. We used NVIDIA RTX3090 GPU to train and test models in TensorFlow 2.7 environment.

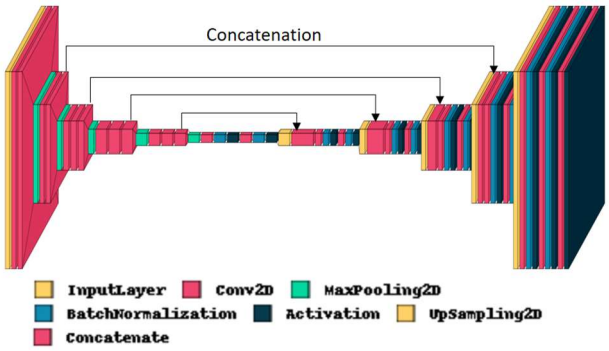

Figure 2. The structure of U-Net with VGG16 as a backbone.

2.3 Accuracy metrics

We evaluate the results with Precision, Recall, and Intersection over Union (IoU) of built-up structures (Van Beers et al., 2019). The calculation of the metrics could be found in Eq. (1) - (3) where TP, FP, and FN refer to the number of the True Positive, the False Positive, and the False Negative pixels for the semantic class.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (1)$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (2)$$

$$IoU = \frac{TP}{TP + \mathrm{FP} + \mathrm{FN}} \quad (3)$$

3 Results and Discussion

We present the Precision, Recall, IoU values of all experiments in Table 1. The highest and lowest IoU values were highlighted with red and black bold text separately for each season. “Bria” refers to weights from models trained on data of the Bria camp.

Table 1. The Precision, Recall, IoU values of all implemented experiments.

Band selection	Pretrained weights	Kule-Dry Season (24 March, 2017)			Kule-Wet Season (22 June, 2018)		
		Precision	Recall	IoU	Precision	Recall	IoU
RGBN	Random	0.8340	0.5819	0.5215	0.8539	0.5976	0.5422
	Bria	0.8380	0.5795	**0.5212**	0.8721	0.5679	0.5242
RGB	Random	0.8072	0.6031	0.5272	0.8056	0.6290	0.5461
	Bria	0.8426	0.5851	0.5275	0.8529	0.6042	0.5472
	ImageNet	0.8316	0.6022	0.5367	0.8368	0.6245	0.5567
RNB	Random	0.8333	0.5936	0.5306	0.8445	0.6376	0.5706
	Bria	0.8375	0.5840	0.5245	0.8723	0.5611	**0.5185**
	ImageNet	0.8275	0.5989	0.5325	0.8457	0.6378	0.5713
RGN	Random	0.8325	0.5921	0.5291	0.8445	0.6299	0.5645
	Bria	0.8410	0.5841	0.5261	0.8557	0.6126	0.5553
	ImageNet	0.8242	0.6080	**0.5382**	0.8374	0.6435	**0.5720**

Firstly, we could observe ImageNet performs the best followed by RIW and then "Bria" in general. ImageNet models achieve the highest IoU values, and the most balanced Precision and Recall values for all three-band combinations. Whereas "Bria" models produce the lowest IoU values in all combinations except RGB bands. Additionally, they produce the highest Precision and the lowest Recall values in all combinations. These results demonstrate that "Bria" models missed the most TP pixels even though they extracted fewer FP pixels than other pretrained weights. However, the results of all experiments expose the imbalance issue in this extraction task. The issue is probably caused by the high imbalance of two semantic classes as mentioned before. It is a critical issue in deep learning domains (Johnson & Khoshgoftaar, 2019). More techniques should be applied to solve the issue to achieve better performance. Besides, it is worth noting that RIW performs better than "Bria", which indicates fine-tuning pretrained weights from other refugee camps can be harmful. The explanation is beyond the scope of this research but is worthy of attention for future research, especially in domain adaptation.

Secondly, we can find three-band combinations outperform four-band combinations in general. It shows feeding four-band data directly to a semantic segmentation model is not recommended under the given conditions of this research. Besides, using the N band to replace the G or B band can improve by around 0.02 IoU values compared to conventional RGB bands for the wet season but cannot influence the performance for the dry season. N band is significant in identifying vegetation (Huang et al., 2021) and probably makes it more important for the wet season when the surrounding environment is covered by more vegetation. Therefore, RGN or RNB bands are highly recommended to replace RGB bands when extracting refugee dwellings in areas covered by a lot of vegetation. This finding is consistent with the outcomes of (Dixit et al., 2021) which prove NRG bands outperform RGB and NRGB bands in terms of F1-score of the class building based on Sentinel-2 imagery. These findings indicate the enhancement of input images based on band combinations can influence on the performance of semantic segmentation models.

Fig. 3 presents predicted labels of a subset of testing data for every band selection, every type of pretrained weights during the dry season and the wet season. Overall, we can observe that many FP and FN pixels occur around the boundary of built-up structures. This type of errors is hard to be avoided. It has been found that the label data annotated by different experts can have slightly differences in the boundary of target objects. Additionally, all models are incapable of detecting built-up structures occluded by trees (seen the example during the dry season in Fig. 3).

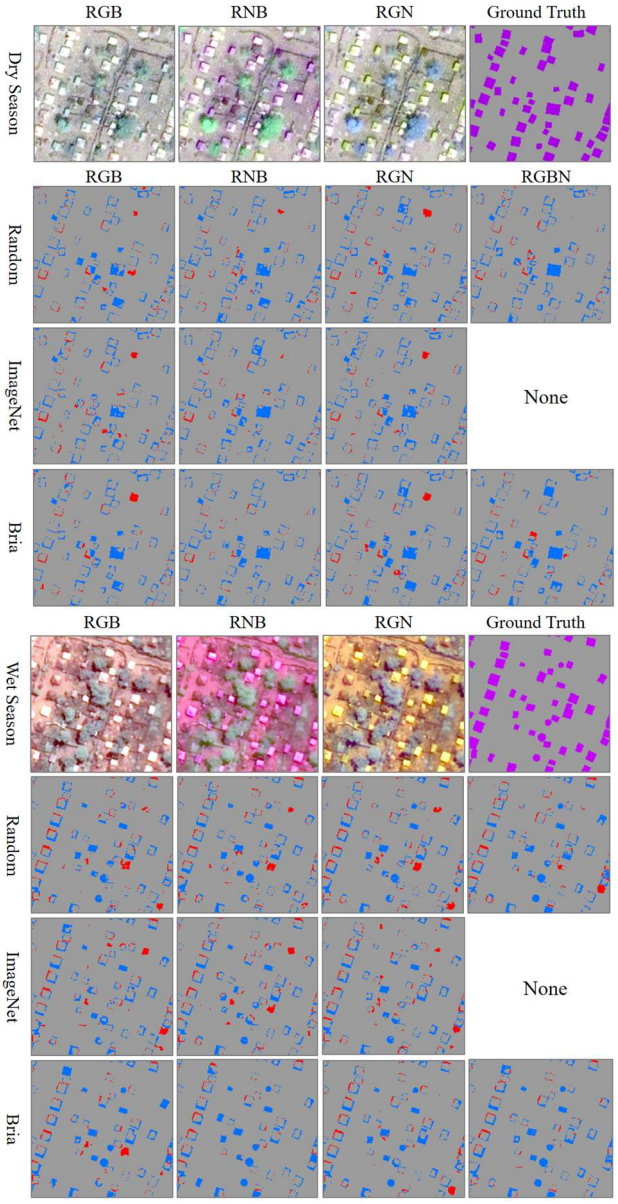

Figure 3. Predicted labels of a subset of testing data for every band combination, every type of pretrained weights during the dry season and the wet season. Blue: False Negative pixels; Red: False Positive pixels.

4 Conclusions

This research compared the performance of four band combinations (RGBN, RGB, RGN, RNB) and three types of pretrained weights (RIW, "Bria", "ImageNet") in extracting refugee dwellings in the Kule refugee camp during the dry season and the wet season. The results illustrate that ImageNet outperforms RIW and "Bria" in terms of IoU and Recall values. On the contrary, "Bria" weights produce the lowest IoU and Recall values. Overall, three-band combinations achieve better results than four-band combinations. Using the N band to replace B or G band is recommended for extraction tasks during the wet season. This finding may be caused by the

significance of N band in identifying vegetation, which probably makes it more important for the wet season when the surrounding environment is covered by more vegetation.

Data and Software Availability

The VHSR satellite imagery and label data are not available restricted by licenses and sensitivity of refugees.

Acknowledgement

This work was supported by Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, the Christian Doppler Research Association (CDG), and Médecins Sans Frontières (MSF) Austria.

References

Ansari, R. A., Malhotra, R., & Buddhiraju, K. M. (2020). Identifying informal settlements using contourlet assisted deep learning. *Sensors (Switzerland)*, *20*(9), 1–15. https://doi.org/10.3390/s20092733

Bock, S., Goppold, J., & Weiß, M. (2018). *An improvement of the convergence proof of the ADAM-Optimizer*. 1–5.

Borba, P., de Carvalho Diniz, F., da Silva, N. C., & de Souza Bias, E. (2021). Building Footprint Extraction Using Deep Learning Semantic Segmentation Techniques: Experiments and Results. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 4708–4711.

Çelik, M., Ergun, Ö., Johnson, B., Keskinocak, P., Lorca, Á., Pekgün, P., & Swann, J. (2012). Humanitarian logistics. In *New directions in informatics, optimization, logistics, and production* (pp. 18–49). INFORMS.

Checchi, F., Stewart, B. T., Palmer, J. J., & Grundy, C. (2013). Validity and feasibility of a satellite imagery-based method for rapid estimation of displaced populations. *International Journal of Health Geographics*, *12*. https://doi.org/10.1186/1476-072X-12-4

Dixit, M., Chaurasia, K., & Kumar Mishra, V. (2021). Dilated-ResUnet: A novel deep learning architecture for building extraction from medium resolution multi-spectral satellite imagery. *Expert Systems with Applications*, *184*(June), 115530. https://doi.org/10.1016/j.eswa.2021.115530

Gella, G. W., Wendt, L., Lang, S., & Braun, A. (2021). Testing Transferability of Deep- Learning-Based Dwelling Extraction in Refugee Camps Methodology 2 . 1 The test sites. *GI_Forum*, *9*(1), 220–227. https://doi.org/10.1553/giscience2021

Gella, G. W., Wendt, L., Lang, S., Tiede, D., Hofer, B., Gao, Y., & Braun, A. (2022). Mapping of Dwellings in IDP/Refugee Settlements from Very High-Resolution Satellite Imagery Using a Mask Region-Based Convolutional Neural Network. *Remote Sensing*, *14*(3). https://doi.org/10.3390/rs14030689

Ghorbanzadeh, O., Tiede, D., Dabiri, Z., Sudmanns, M., & Lang, S. (2018). Dwelling extraction in refugee camps using CNN - First experiences and lessons learnt. *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

Ghorbanzadeh, O., Tiede, D., Wendt, L., Sudmanns, M., & Lang, S. (2021). Transferable instance segmentation of dwellings in a refugee camp - integrating CNN and OBIA. *European Journal of Remote Sensing*, *54*(sup1), 127–140. https://doi.org/10.1080/22797254.2020.1759456

Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, *32*(1), 1–6. https://doi.org/10.1007/s11676-020-01155-1

Jia Deng, Wei Dong, Socher, R., Li-Jia Li, Kai Li, & Li Fei-Fei. (2009). *ImageNet: A large-scale hierarchical image database*. 248–255. https://doi.org/10.1109/cvprw.2009.5206848

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0192-5

Jung, H., Choi, H. S., & Kang, M. (2021). Boundary Enhancement Semantic Segmentation for Building Extraction From Remote Sensed Image. *IEEE Transactions on Geoscience and Remote Sensing*, 1–12. https://doi.org/10.1109/TGRS.2021.3108781

Kemker, R., Salvaggio, C., & Kanan, C. (2018). Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *145*(June 2017), 60–77. https://doi.org/10.1016/j.isprsjprs.2018.04.014

Lang, S., Füreder, P., Riedler, B., Wendt, L., Braun, A., Tiede, D., Schoepfer, E., Zeil, P., Spröhnle, K., & Kulessa, K. (2020). Earth observation tools and services to increase the effectiveness of humanitarian assistance. *European Journal of Remote Sensing*, *53*(sup2), 67–85.

Li, W., He, C., Fang, J., Zheng, J., Fu, H., & Yu, L.

(2019). Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. *Remote Sensing*, *11*(4). https://doi.org/10.3390/rs11040403

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *European Conference on Computer Vision*, 740–755.

Lu, Y., & Kwan, C. (2020). Deep Learning for Effective Refugee Tent. *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, *18*(8), 16–20.

Médecins Sans Frontières. (2018). *Renewed violence threatens people and healthcare in Bria*. https://www.msf.org/central-african-republic-renewed-violence-threatens-people-and-healthcare-bria

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

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–14.

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. https://doi.org/10.3390/rs6109277

Tiede, D., Schwendemann, G., Alobaidi, A., Wendt, L., & Lang, S. (2021). Mask R-CNN- based building extraction from VHR satellite data in operational humanitarian action: An example related to Covid- - 19 response in. *Transactions in GIS*, 1–15. https://doi.org/10.1111/tgis.12766

UNHCR. (2020a). *Kule refugee camp* (Issue May).

UNHCR. (2020b). *The Sustainable Development Goals and the Global Compact on Refugees*. https://www.unhcr.org/5efcb5004.pdf

Van Beers, F., Lindström, A., Okafor, E., & Wiering, M. A. (2019). Deep neural networks with intersection over union loss for binary image segmentation. *ICPRAM 2019 - Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, *Icpram*, 438–445. https://doi.org/10.5220/0007347504380445

Wickert, L., Bogen, M., & Richter, M. (2021). Lessons Learned on Conducting Dwelling Detection on VHR Satellite Imagery for the Management of Humanitarian Operations. *Sensors & Transducers*, *249*(2), 45–53.

Yakubovskiy, P. (2019). Segmentation Models. *GitHub Repository*.

Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., & Liang, J. (2017). EAST: An efficient and accurate scene text detector. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-January*, 2642–2651. https://doi.org/10.1109/CVPR.2017.283