RURAL SETTLEMENTS SEGMENTATION BASED ON DEEP LEARNING U-NET USING REMOTE SENSING IMAGES

Zakaria AAMIR*[1], Mariem SEDDOUKI*[1], Oussama HIMMY[1], Mehdi MAANAN[1], Mohamed TAHIRI[2], Hassan RHINANE[1]

[1]Geosciences Laboratory, Faculty of Sciences-Ain Chock, Hassan II University, Casablanca, Morocco.
[2]Organic Synthesis Laboratory, Faculty of Sciences-Ain Chock, Hassan II University, Casablanca, Morocco.

Commission IV, WG7

KEY WORDS: Rural Settlements, Remote Sensing, Deep Learning, U-net, Image segmentation.

ABSTRACT:

Accurate and efficient extraction of rural settlements from high-resolution remote sensing imagery is of paramount importance for rural government management. Unplanned rural settlements are quite common. Understanding the spatial characteristic of these rural settlements is of great importance as it offers indispensable information for land management and decision-making. In this setting, the U-net architecture is proposed in this study for rural settlements differentiation by image segmentation on high-resolution satellite images of rural settlements in Zagora province, Draa-Tafilalet region, Morocco. To predict pixels in remote sensing images representing rural settlements in this province. Image segmentation is conducted using different encoders in the U-net architecture, and the results are compared. Experimental results demonstrate that the proposed method effectively mapped and discriminated rural settlements areas with an overall accuracy of 98%, achieving comparable and improved performance over other traditional rural extraction methods.

1. INTRODUCTION

Rural settlements, defined as settlements areas for rural residents to produce and live in, are directly related to population distribution and economic growth in rural areas. Our planet is becoming increasingly urbanized, and the global population is now concentrated more in urban than rural areas. Still, around 45% of the world’s population is still living in rural areas.

In Morocco, the total rural population from 2008 to 2018, over 13.53 million people were living in rural areas. In most existing land-use and land-cover maps, the distribution of rural settlements is not well characterized because they are incomplete and fragmented in organization and relatively stable over time. In this matter, the country adopted an inclusive and sustainable development of the rural world, which requires the study of expansion and reduction of rural settlements.

The acquisition of high-resolution remote sensing images has become more convenient, and easily accessible providing huge support generally for classification and change detection and precisely for the extraction of rural settlements. And with the advancement of technology of deep learning, target detection has become more useful and necessary in remote sensing image comprehension.

Neural networks are the basis of deep learning (DL) algorithms and are commonly used in many remote sensing applications, however, many machine learning (ML) methods have shown higher accuracy, such as Support Vector Machines (SVM) and ensemble classifiers, e.g., random forest (RF), for image classification and other tasks (e.g., change detection). This led to renewed interests in neural networks which made DL algorithms such as convolutional neural network (CNN), the regional recommendation convolutional neural network (R-CNN), and its new generations, achieve significant success in many images’ classification, object detection, image segmentation tasks.

In recent years, many deep learning methods have been used to identify rural buildings (Zhao Qingzhan, 2019). AlexNet and SVM were also employed (Li, Z., 2017). These studies were all based on object detection methods which were limited since it uses rectangular frames to locate objects and give irregular shapes as results of the real objects (Liu et al., 2019)

Each pixel should be assigned a class label in the rural settlement extraction task, which requires an algorithm to learn to make predictions for each pixel. U-net is a convolutional autoencoder widely used in the medical field and other industries; it performs high precision pixel-based segmentation on images (Ronneberger et al., 2015). However, the use of U-net to recognize rural settlements is still uncommon.

Given the above facts, this paper explores the potential of rural settlement segmentation based on satellite images using U-net algorithm and estimates the areas of rural settlements in a rapid and low-cost manner, and studies the change from the past to the present.

2. STUDY AREA AND DATASETS

2.1 Study Area

In this paper, 10 villages of Zagora province were selected as a study area, Zagora province is located in the south-eastern of Morocco as a part of the Draa-Tafilalet region (Figure 1) covering 3.55 % of the total of the country with a surface of 21 041 km^2. The climate is dominated by low rainfall, stormy character, and large fluctuation in daily and yearly temperatures.

According to the 2014 National Census, 307 306 inhabitants populate this province with a density of 15 habitat/km^2. Zagora is a region with a rural character, which makes it an ideal study area to examine our proposed method, the material for the construction of houses mainly relied on clay help the study to distinguish between the rural and urban areas and study the extension and changes of rural settlements in this region.

In this region, rural settlements are more concentrated on the sides of Draa valley as the main source of water for inhabitant’s

life based on agriculture, we can distinguish two categories of settlements;

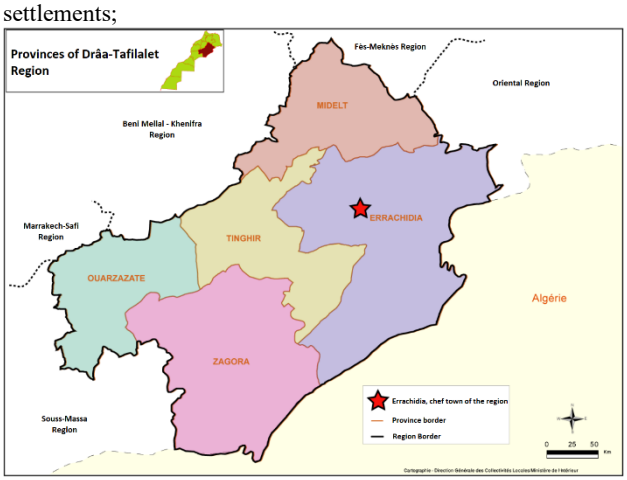

Figure 1. Provinces of Draa-Tafilalet Region

1. Low-density settlements, which are disorderly, distributed and have different orientations; this type of settlement is more noticed close to Drâa valley obscured by the surrounding palm trees and vegetation.

2. High-density settlements are newly built residential areas where such settlements have higher density in the area and have identical spacing and same surface, this type is more noticed adjacent to the newly built transportation roads.

Figure 2. Satellite image of the study area in 2020 and example of (a) low-density rural settlement and (b) high-density settlement

2.2 Datasets

The high-resolution remote sensing imagery used in this work was collected from SASPlanet maps, primarily Bing maps. The spatial resolution of every pixel in the remote sensing imagery is 90cm. SASPlanet is a program designed for viewing and downloading high-resolution satellite imagery and conventional maps submitted by such services as Google Maps, DigitalGlobe, Kosmosnimki, Yandex.Maps, Yahoo! Maps, VirtualEarth, Gurtam, OpenStreetMap, eAtlas, Genshtab maps, iPhone maps, Navitel maps, Bings Maps (Bird's Eye) etc.

The acquired dataset covers rural regions in Zagora province, which a similar signature covered near the Draa valley by the vegetation, which makes rural settlements a little more challenging, and less in areas far from the valley. Most deep learning-based segmentation approaches achieve high accuracy using this kind of imagery; however, a problem of generalization capability occurs, predicting new rural settlements requires using the same type as the training datasets.

3. METHODOLOGY

The general flowchart (Figure 3) of our proposed method, starts with preparing RGB remote sensing images and splitting the dataset into a training set and test set. Second, U-net model was used to perform rural settlements segmentation. Finally, an accuracy assessment was conducted on the test set.

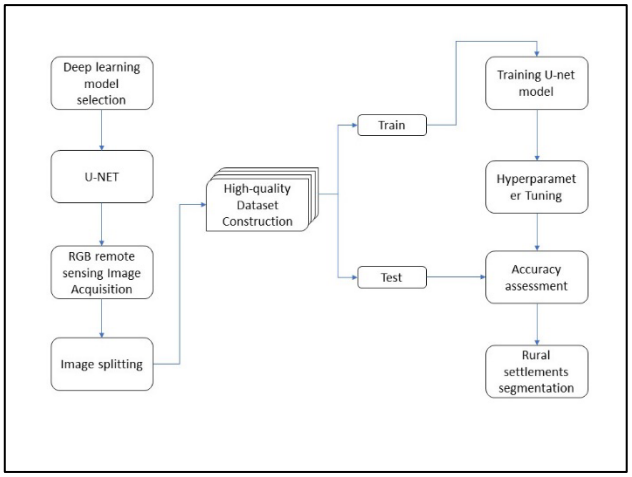

Figure 3. The flowchart in this study

3.1 U-net architecture and parameter settings

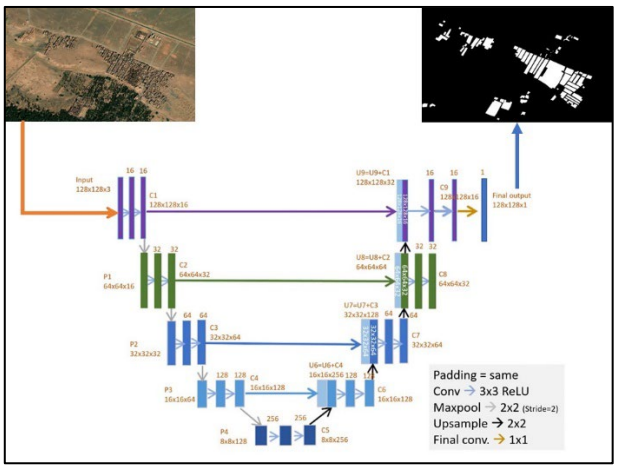

Figure 4. Illustration of U-net architecture for semantic segmentation (modified from Ronneberger et al.)

In this study, the process of rural settlement segmentation was developed based on U-net architecture following the equivalent diagram developed by (Ronneberger et al., 2015). The U-net architecture is synonymous with an encoder-decoder architecture, it is a deep learning framework based on FCNs and

is applied mainly for the segmentation of small datasets of medical images; it comprises two parts:

1. The contraction path like Encoder, following the typical architecture of convolutional networks. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a Rectified Linear Unit (ReLU) and a 2x2 max-pooling operation using Stride 2 for down sampling. At each down sampling step, we double the number of feature channels. The purpose of this contracting path is to capture the context of the input image to be able to do segmentation.

2. Like the expansion path of the decoder, where each step consists of up sampling the feature map, followed by a 2x2 convolution that bisects the number of feature channels, with the appropriate pruned feature map in the shrinking path. connections, and two 3x3 convolutions, each followed by a ReLU. The purpose of this expanded path is to provide accurate localization along with contextual information from the contracted path.

3.2 Network Model Training and accuracy assessment

Preparing the dataset is essential before the training task, it starts labelling the images to highlight data features, properties, characteristics, or classifications that can be analysed for patterns that help predict the target, in this case, rural settlements.

The model version of the U-net used in this paper requires input images with the size of 128 x 128 pixels, which means it was necessary to split dataset images into tiles of the size 128 x 128 pixels. After splitting the dataset, we clean it to remove unwanted images from the dataset to make it ready for analysis. For now the best (most reliable) way to clean image datasets is manually.

The operations of splitting and labelling input images were done in the ArcGIS 10.3 environment, for labelling it is based on digitizing rural settlements as shapefiles and then converting shapefiles into binary raster masks with the same dimension as the original image (black for non-rural settlements and white for rural settlements). The splitting task relies on the split raster tool for both original RGB images and masks having the same names in the folder output which is necessary for the training part.

Here, a total of 2600* RGB input images with their mask were prepared, and randomly divided into training samples of 70% and test samples of 30% for model training. The network was trained for 50 epochs using binary cross-entropy as a loss function using Keras v2.8.0 with TensorFlow 2.5.0 using Python programming language (v3.9.7).

Model training was performed taking into consideration hyperparameter tuning and optimization, the hyperparameters were estimated with consideration of the number of iterations equal to 50, batch size equal to 16, and learning rate equal to 0.01.

The ability of the U-net algorithm to segment and extract rural settlements were evaluated by the following quantitative indicators: overall accuracy, precision, recall, and F1 score. These indicators are presented as calculated true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs). For a class l, TP is the number of correctly classified pixels as l. FP is the number of pixels that are misclassified as l. Finally, FN represents pixels that belong to l but are associated by the model with some other classes.

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

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

$$F1 = \frac{2.Precision.Recall}{Precision+Recall} \quad (3)$$

$$Overall\ Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision is a measure of how many of the positive predictions made are correct (true positives), Recall is a measure of how many of the positive cases the classifier correctly predicted, and overall the positive cases in the data.

Precision and Recall are common indicators used to evaluate classification performance (Fawcett, 2006). However, neither precision nor recall is necessarily useful alone, since we rather generally are interested in the overall picture. This is why Accuracy is always good to check as one option. F1-score is another.

F1-score combines precision and recall; it is generally described as the harmonic mean of the two. Harmonic mean is just another way to calculate an "average" of values, generally described as more suitable for ratios (such as precision and recall) than the traditional arithmetic mean, it works also for cases where datasets are imbalanced as it requires both precision and recall having a reasonable value.(Moon et al., 2020).

Intersection-over-union (IoU) is another way to evaluate the performance of the developed approach, it represents the proximity of the predicted object to the ground truth. In Equation (6), A and B are two different data samples (Papadomanolaki et al., 2019).

The calculation formula of IoU is as follows:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (5)$$

$$IoU(A,B) = \frac{A \cap B}{A \cup B} \quad (6)$$

4. RESULTS

The proposed method achieved the OA of 98 % with a dice score (F1score) of 0.90. The validation results of the U-net algorithm-based homestead identification are presented in Table 1.

Indicators	Precision	Recall	F1	Overall accuracy	IoU
U-net	0.94	0.88	0.90	0.98	0.90

Table 1. Validation of the U-net algorithm identification results.

The indicators summarized in Table 1, indicate very good performance of rural settlements segmentation.

Figure 5 shows the results of U-net identification, including 5 training sites. The predicted images in the training sites and test sites showed a good match with the label image.

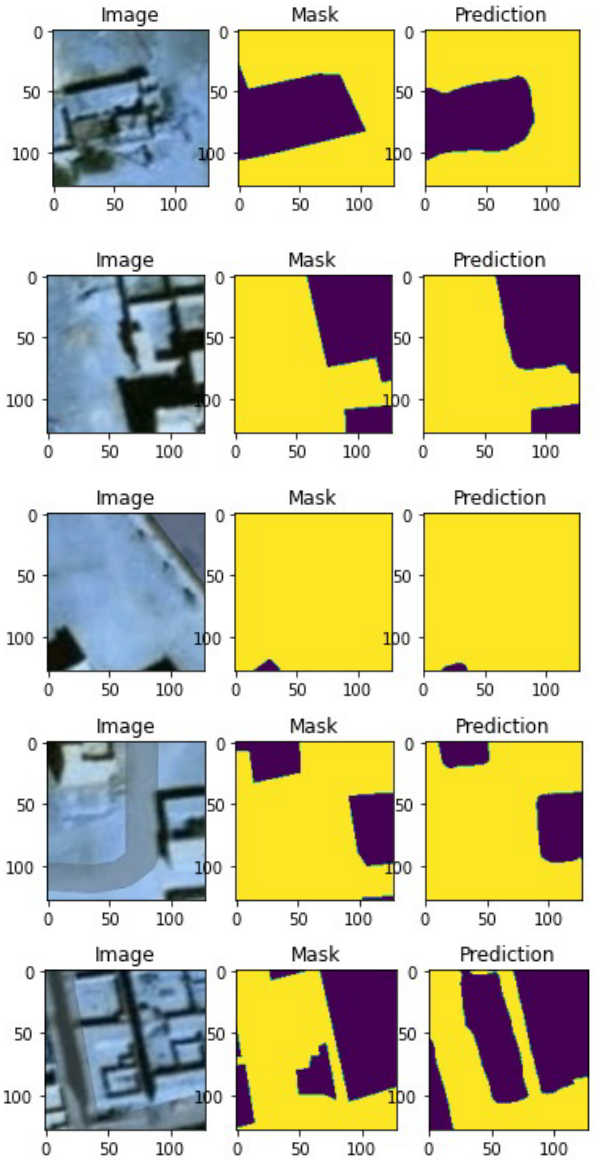

Figure 5. Comparison of remote sensing original RGB images, ground truth, and U-net algorithm recognition results in indicative regions.

The separation between the rural settlements and non-rural settlements is obvious, and the edges of the U-net recognition seem to be somewhat regular.

A visual explanation indicates the proposed method can effectively distinguish rural residential areas from other man-made.

Another thing that demonstrates the strength of the model and proves that it is well learned is that despite the fact that the mask erred in some parts of the house (Figure 5). However, the algorithm very well predicted that there was a building there.

The accuracy of a model is usually determined after the model parameters and is calculated in the form of a percentage. It is the measure of how accurate the model's prediction is compared to the true data.

In our case, the used model has given great results (Figure 6), indeed we get almost 98.35% in terms of accuracy during training and 98.87% during model testing which is considered as good results for this kind of problems.

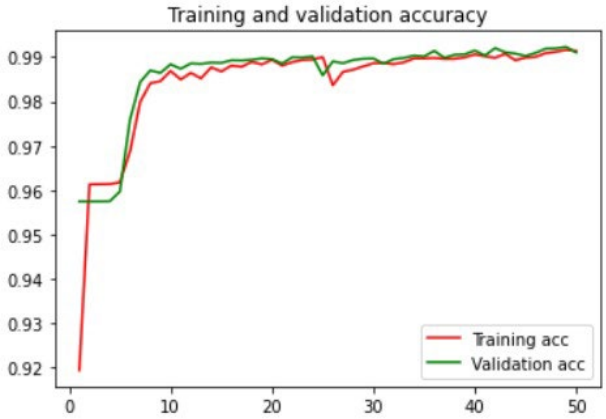

Figure 6. Accuracy of U-NET

On the other hand, the loss function is used to optimize our trained model. The loss is calculated on training and validation and its interpretation is based on how well the model is doing in these two operations. It is the sum of errors made for each example in training or validation sets. Loss value implies how poorly or well a model behaves after each iteration of optimization. We got 0.0248 in terms of error and 0.0279 after training which are considered as low error values for the used Binary Cross-Entropy function to measure the error of the model (Figure 7).

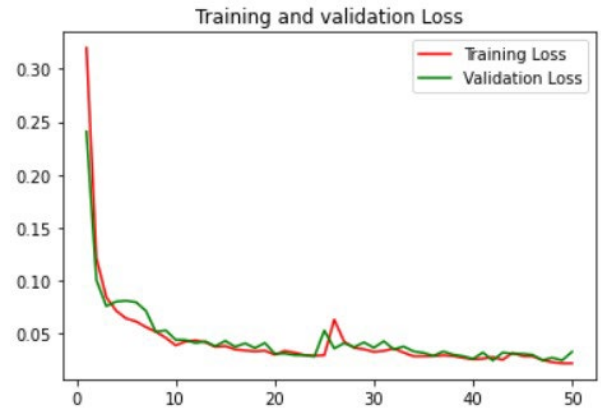

Figure 7. Loss of U-NET

5. DISCUSSION

This paper, like many other deep learning projects for feature extraction in remote sensing, provides great potential for rural settlements extraction from RGB composite remote sensing images, the experiment achieved reasonable and encouraging results.

Despite the fact that soil and rural settlements have nearly the same RGB signature, the proposed method seemed promising in terms of potential for extracting rural settlements from RGB

images and yields good results indicated by indicators shown in Table 1. Moreover, this does not deny the fact that the use of multispectral remote sensing images will bring a huge improvement to this model because by including more bands, convolution filtering can improve the extraction of building texture, shape, and edge features. This view supports the suggestion that multispectral images are superior to three-channel images for deep neural network training, despite the rare availability of high-resolution multispectral images in Morocco.

There are still some issues to discuss regarding the efficiency of network model training. The potential and importance of deep learning methods for object extraction stem from their fast and complete automation, sample repeatability, and adaptability. The performance of the U-net depends on a large amount of training data. A model that uses significantly fewer training samples and produces accurate results is highly desirable (Ball et al., 2017).

It is still difficult to say how many samples are enough to train a robust network model. In addition, there is potential for further improvements in U-net design, data expansion, finer image resolution, etc.

6. CONCLUSION

Rural settlements segmentation using RGB remotely sensed images remains a challenging task, due to the spatial scale variation of RGB signature of rural settlements, although it seems to work perfectly fine within Zagora province with the possibility to be extended to a larger area, especially in neighbor regions having same rural settlements structure as Zagora.

The study succeeded to show the capability of deep learning for rural settlements mapping in which segmentation was tested.

This paper presented an effective rural settlements extraction method based on a U-net from RGB remote sensing images. U-net was used to perform segmentation with an overall accuracy of 0.98.

Although our experiment does not involve up-to-date remote sensing images, it still provides a potential alternative to time-consuming surveys related to rural areas in Zagora province.

In future works, further improvements could be made by integrating UAV imagery that excels in real-time image acquisition, pixel-based identification, and even 3D modelling recognition.

REFERENCES

Ball, J.E., Anderson, D.T., Sr, C.S.C., 2017. A comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. J. Appl. Remote Sens. 11, 042609. https://doi.org/10.1117/1.JRS.11.042609

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett., ROC Analysis in Pattern Recognition 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Liu, Z., Cao, Y., Wang, Y., Wang, W., 2019. Computer vision-based concrete crack detection using U-net fully convolutional networks. Autom. Constr. 104, 129–139. https://doi.org/10.1016/j.autcon.2019.04.005

Moon, W.K., Lee, Y.-W., Ke, H.-H., Lee, S.H., Huang, C.-S., Chang, R.-F., 2020. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. Comput. Methods Programs Biomed. 190, 105361. https://doi.org/10.1016/j.cmpb.2020.105361

Papadomanolaki, M., Vakalopoulou, M., Karantzalos, K., 2019. A Novel Object-Based Deep Learning Framework for Semantic Segmentation of Very High-Resolution Remote Sensing Data: Comparison with Convolutional and Fully Convolutional Networks. Remote Sens. 11, 684. https://doi.org/10.3390/rs11060684

Pan, Z., Xu, J., Guo, Y., Hu, Y. and Wang, G., 2020. Deep learning segmentation and classification for urban village using a worldview satellite image based on U-Net. Remote Sensing, 12(10), p.1574. https://www.mdpi.com/2072-4292/12/10/1574/htm#B24-remotesensing-12-01574 (accessed 5.7.22).

Ji, H., Li, X., Wei, X., Liu, W., Zhang, L. and Wang, L., 2020. Mapping 10-m resolution rural settlements using multi-source remote sensing datasets with the Google Earth Engine platform. Remote Sensing, 12(17), p.2832. https://www.mdpi.com/2072-4292/12/17/2832/htm (accessed 5.7.22).

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Xu, R., 2021. Mapping Rural Settlements from Landsat and Sentinel Time Series by Integrating Pixel- and Object-Based Methods. Land 10, 244. https://doi.org/10.3390/land10030244

Zhang, X., 2020. Village-Level Homestead and Building Floor Area Estimates Based on UAV Imagery and U-Net Algorithm. ISPRS Int. J. Geo-Inf. 9, 403. https://doi.org/10.3390/ijgi9060403

Zhao Qingzhan, R.Y., 2019. Target Detection of Rural Buildings in UAV Remote Sensing ImagesBased on Convolutional Neural Network.

Zhang, L., Zhang, L. and Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and remote sensing magazine, 4(2), pp.22-40. https://ieeexplore.ieee.org/abstract/document/7486259 (accessed 5.7.22).

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F. and Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), pp.8-36. https://ieeexplore.ieee.org/abstract/document/8113128/ (accessed 5.7.22).