GEOMATICS AND ENVIRONMENTAL ENGINEERING • Volume 18 • Number 3 • 2024

https://doi.org/10.7494/geom.2024.18.3.25

Miguel Chicchon[1], Eva Savina Malinverni[2], Marsia Sanità[3], Roberto Pierdicca[4], Francesca Colosi[5], Francisco James León Trujillo[6]

Building Semantic Segmentation Using UNet Convolutional Network on SpaceNet Public Data Sets for Monitoring Surrounding Area of Chan Chan (Peru)

Abstract: The amount of damage to cultural heritage sites is increasing rapidly every year. This is due to inadequate heritage management and uncontrolled urban growth as well as unpredictable seismic and atmospheric events that manifest themselves in a continuously deteriorating ecosystem. Thus, applications of artificial intelligence (AI) in remote-sensing (RS) techniques (machine-learning and deep-learning algorithms) for monitoring archaeological sites have increased in recent years. This research involves the surrounding area of the archaeological site of Chan Chan in Peru in particular. An approach that is based on the use of AI algorithms for building footprint segmentation and change-detection analysis by means of RS images is proposed. It involves a UNet convolutional network based on an EfficientNet B0 to B7 encoder. The network was trained on two public data sets from SpaceNet that were based on WV2 and WV3 satellite images: SpaceNet V1 (Rio), and SpaceNet V2 (Shanghai). In the pre-processing phase, the images from the two data sets have been equalized in order to improve their quality and avoid overfitting. The building segmentation has been performed on HRV images of the study area that were downloaded from Google Earth Pro. The value that was achieved in the IoU metric was around 70% in both experiments. The purpose of this proposed methodology is to assist scientists in drafting monitoring and conservation protocols based on already-recorded data in order to prevent future disasters and hazards.

Keywords: building detection, neural network, segmentation, HRV images, SpaceNet data set

Received: November 30, 2023; accepted: February 29, 2024

[1] Pontificia Universidad Católica del Perú, Escuela de Posgrado, San Miguel, Lima, Perú, email: mchicchon@pucp.edu.pe, https://orcid.org/0000-0002-2228-8557

[2] Università Politecnica delle Marche, Dipartimento di Ingegneria Civile, Edile e Architettura, Ancona, Italy, email: e.s.malinverni@staff.univpm.it, https://orcid.org/0000-0001-6582-2943

[3] Università Politecnica delle Marche, Dipartimento di Ingegneria Civile, Edile e Architettura, Ancona, Italy, email: m.sanita@pm.univpm.it (corresponding author), https://orcid.org/0009-0002-9828-2626

[4] Università Politecnica delle Marche, Dipartimento di Ingegneria Civile, Edile e Architettura, Ancona, Italy, email: r.pierdicca@staff.univpm.it, https://orcid.org/0000-0002-9160-834X

[5] Consiglio Nazionale delle Ricerche (CNR), Istituto di Scienze del Patrimonio Culturale, Rome, Italy, email: francesca.colosi@cnr.it, https://orcid.org/0000-0002-1384-2560

[6] Universidad de Lima, Instituto de Investigación Científica, Carrera de Ingeniería Civil, Lima, Perú, email: francisco.leon2903@gmail.com, https://orcid.org/0000-0003-1060-5938

1. Introduction

The imminent urban growth that we witness today has become one of the major causes of the negative impact we have on the environment. According to an estimate that was made by the United Nations, the level of urbanization will grow from 55% (today) to 70% (2050). The excessive consumption of soil causes an imbalance in the natural ecosystem; one of the most important consequences of this is soil and riverbank erosion [1]. Another consequence of our soil consumption is certainly the increase in the amount of damage to cultural heritage sites due to the rapid urban growth around them. It is therefore appropriate to develop a new methodology that is able to provide the greatest amount of information for environmental monitoring and for a possible environmental risk-prediction model. For this reason, a combination of remote-sensing techniques and deep-learning algorithms can be the right answer to this need. However, the problem of managing a large amount of data to be processed always remains. Fortunately, recent advances in the field of computational power combined with new developments in machine-learning and statistical model techniques offer new opportunities in the management of Earth system data. AI gives many opportunities in terms of accessibility, preservation, and dissemination in the cultural heritage (CH) fields (particularly for museums and archaeological sites). Thanks to the technological progress of the last decades, it also allows for the possibility of processing large amounts of data thanks to the use of neural networks [2]. At the same time, great interest has emerged in identifying urban growth by using remote sensing images. Nowadays, the accurate identification of building footprints involves many areas of application; from disaster-event monitoring to new urban-planning strategies [3]. An example of how the combination of machine-learning algorithms could help in the rapid detection of vulnerabilities on existing buildings was provided by [4, 5]. It is possible to integrate this research work with a GIS platform. In this way, it can be used for the detection of unknown areas; for example, to identify houses outside a land register or even roads outside a land register. This would thus become a highly exploitable methodology. CNNs therefore demonstrate great success in various application sectors, which is why researchers are transferring increasingly higher performance levels to these deep-learning algorithms. In [6], there is a detailed analysis on recent deep-learning developments for semantic segmentation. Deep-learning semantic-segmentation algorithms are widely used in many urban-change-detection and building-footprint-detection applications [7]. Thanks to the enormous accessibility of satellite data, these remote-sensing techniques combined with machine learning can be exploited in environmental remote-sensing applications [8]. This is because it is necessary to only have a pre-trained model and labeled images to enter into a deep-learning segmentation algorithm. This is different from the machine-learning approach, where images are designed manually. In this paper, a combination of a neural network and a semantic segmentation on HRV images is proposed in order to detect the building footprint

in the surrounding archaeological area of Chan Chan, Peru. In the second section, there is a brief introduction to the state of the art. In the third section, there is a short presentation of the area of study, followed by a description of the materials and methods that were used. We will pass from remote sensing, geospatial data, artificial intelligence, and the UNet architecture and go through a presentation of the labeled public data sets until we arrive at the workflow that we have developed here. In the fourth section, there is the part that is dedicated to the obtained results; here, the results refer to three different scenarios. The last section concerns the conclusion and the future improvements that the authors expect to achieve in the near future.

2. State of the Art

In recent years, many researchers have attempted to develop new methodologies that are capable of predicting, monitoring, and identifying the phenomenon of urban growth. This is a constantly growing phenomenon, and it is evident in every city in the world. Sometimes, this new urbanization is due to the migrations of populations from rural areas. Nowadays, it has become important to know this phenomenon in order to readily formulate urban-planning strategies that are compliant to the rules of environmental sustainability [9]. This also concerns the sphere of cultural and archaeological heritage sites that need to be protected and safeguarded now more than ever. Together with inappropriate site management, uncontrolled urban development and tourism growth are threats to CH sites [10].

There are many applications in the state of the art where, when evaluating urban growth, researchers have applied a combination of remote sensing and GIS. This type of approach was proposed in [11] in order to evaluate urban grown at a micro-level scale. In [11], the authors collected their multi-temporal database using satellite images (Landsat-1, Landsat-3, Landsat-5, and Landsat-7) with different spatial resolution levels and with the use of population data that came from a district census handbook. A supervised classification was performed using the maximum-likelihood classifier (MLC) algorithm for LULC classification for acquired images from 1972, 1980, 1990, 2001, 2010, and 2016.

Based on the same combined technique, there was a case study that was proposed in [12] regarding the area of Kerala, India. The authors analyzed the area's dynamic urban growth during the period of 1991–2018 by using a collection of satellite images (Landsat-5, Landsat-7, and Landsat-8). They used the derived built-up index (IDBI) to extract automatically built-up features from the satellite images. To quantify the urban growth, they extracted the urban area from each year's data and then used it for Shannon's entropy calculations.

In [9], a data set that was composed of satellite images (Landsat-5, Landsat-7, and Landsat-8) were used to evaluate urban growth – selecting images from 2000 through 2010. They proposed a new method based on mapping the impervious

surface percentage (ISP) year-by-year in the area of Guangzhou in China. As a result, they were able to capture the spatial variation and urban growth throughout the period. To validate their results, they compared them to Google Earth images; thus, the continuous monitoring of urban growth by a thematic map (civil map) is necessary nowadays. The problem that occurs is a great commitment from economic and temporal points of view for the collection of data. So, an automatic approach is required to make this detection faster and cheaper. In this case, AI can come to the rescue. This type of civil map is also useful in the case of a post-disaster event.

Building detection was proposed in [13] based on the use of the UNet convolutional neural network (CNN) [14] using the SpaceNet data set. The authors used three different approaches in order to gain improvements with their UNet implementation. In their first method, they used an algorithm that was able to place each pixel into one of three classes that they had selected: border, inside a building, or outside a building. This was based on a random-forest-based classification. In the second method, the authors used a classification CNN when considering the same three classes. The output of this CNN was a map that converted into a polygon footprint mask. The last approach was on the basis of a cascade approach. Therefore, they concluded by saying that their UNet network had very good performance with respect to the three methods based on the use of the SpaceNet building detection and that it seemed to be a good solution in terms of the method for segmentation. However, it had some limitations; for example, their UNet network had difficulty in detecting buildings that were too close to each other. A similar approach that was based on a combination of the effectiveness of EfficientNetV2 T as an encoder and the convolutional layers of UNet as a decoder was proposed in [15]. EfficientNetV2 T was trained by the ImageNet data set. The authors used two data sets to evaluated the combined network. The first data set was composed of timeseries (2017–2021) Sentinel-2 satellite images. The second data set was the Onera Satellite Change Detection data set (OSCD), which is a specific data set for urban-change detection. The authors achieved an overall accuracy of 97.66%. Another case in the literature that featured the use of the UNet network for segmentation was presented in [16]. They proposed a hybrid optimization of the UNet network, and they evaluated its overall accuracy. They recognized the segmentation errors and spectral and spatial errors due to the noise; thus, they proposed a new approach for detecting changes in multitemporal multispectral satellite images (called Wader hunt optimization – WaHO) in order to reach high accuracy. They pre-processed the satellite images by a median filter and then extracted new features by using seven vegetable indexes. This was followed by the use of the UNet network to carry out vegetation segmentation after being tuned by the WaHO algorithm. Also, the authors of [17] proposed a research methodology whose aim was to detect land-cover changes; in particular, they concentrated it on landfill detection by using a combination of satellite images and a neural network. From their point of view, semantic segmentation with satellite images with the goal of extracting vegetation and urban areas was very useful for

providing good support for sustainability. To carry out their methodology, they used a modified version of UNet named Deep UNet, with a pre-processing phase using FAAGKFCM and SLIC Superpixel. Comparing this method to others (for example, SegNet and UNet), they demonstrated their higher accuracy in their new method.

Another comparative set of experiments was conducted in [3]. Starting from the assumption that obtaining precision edges in segmentation was still an open challenge, the authors proposed their own methodology that was capable of identifying a building's footprint, thus improving the boundaries of the segmentation masks. The proposed method that they adopted was named holistically nested edge detection (HED), and it was able to reach improved performances when extracting building footprints. In this case of study, the aim of the building footprint was to preserve and protect archaeological areas for future generations by exploiting AI. In our case, the surrounding area of Chan Chan will be analyzed from the point of view of the urban growth that is happening around it.

3. Materials and Methods

3.1. Study Area

The case study refers to the surrounding archaeological area of Chan Chan, which covers 20 km^2 (more or less) if the completely urbanized territory is considered (Fig. 1).

Fig. 1. Aerial view of monumental site of Chan Chan

Source: [14]

In the central area, there are nine monuments and five *huacas* (Fig. 2). Since 1986, it has been on the UNESCO World Heritage List [18]. For this reason, it is of considerable importance to be able to protect the site in a continuous and non-invasive manner. The application of remote-sensing techniques with the support of artificial intelligence algorithms is a valid non-invasive, inexpensive, and immediate solution. Despite its notable importance from a historical/cultural point of view, the archaeological site of Chan Chan is not well-known in the world. Over the years, much geospatial data has been collected on it; a storytelling of the site with the aim to disseminate and promote the site at an international level was proposed in [19].

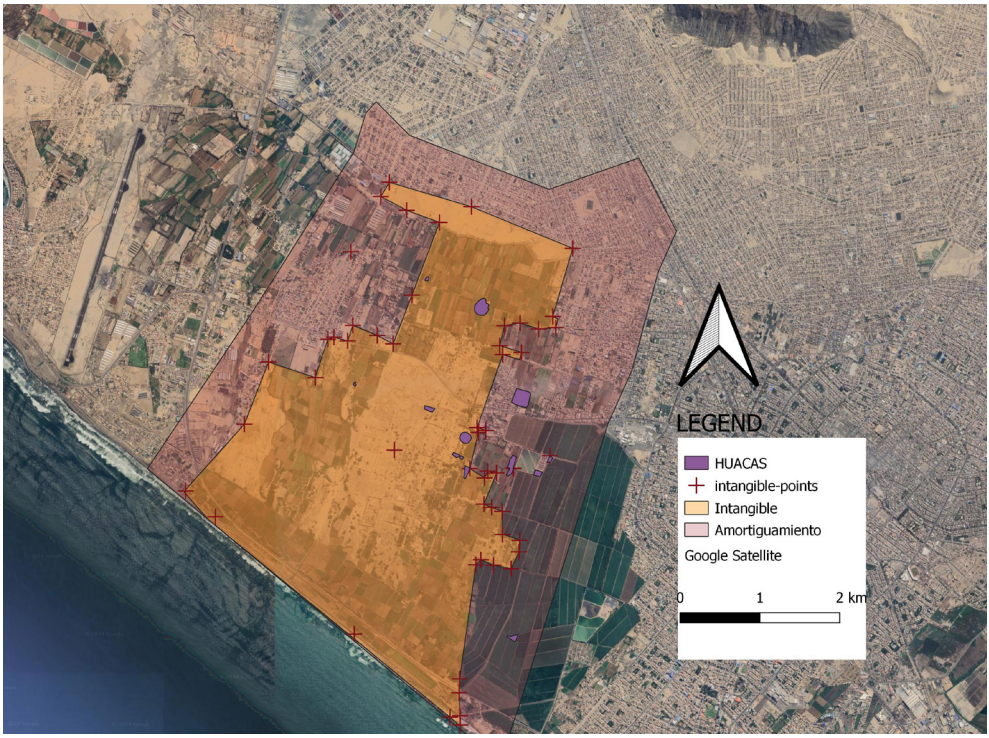

Fig. 2. Satellite view of archaeological site of Chan Chan

3.2. Methods

The goal of this work is to identify buildings through the semantic-segmentation technique. For this type of work, support for providing quite good results is provided by a combination of geospatial data and artificial intelligence (GeoAI). Artificial intelligence (AI) includes machine-learning (ML) methods that include deep learning (DL). AI allows us to simulate human brain processes using numerical algorithms. In particular, DL algorithms are faster with respect to ML algorithms; this

is one of the reasons that researchers rely on them. Other reasons are the fact that there is sometimes much data to process, or perhaps there is not prior knowledge of a problem. The process of semantic segmentation (SS) is a classification of images in which each pixel is associated with a label (class). In the SS process, a label is assigned to detect the edges of an object; so, it is different from the object detection that is done in which the identification is made by a rectangular bounding box. So, a good level of SS results is much more difficult to achieve. DL is based on the artificial convolutional neural networks (CNN) that are used to achieve image segmentation. In this case of study, a CNN that was based on the UNet architecture was applied.

In Figure 3, the workflow scheme is represented. The first step was to create an image collection based on Google Earth Pro images over the time interval from 2003 through 2023. Once this step was completed, it was time to identify two public data sets to train the network. Then, the two public data sets were equalized, and the UNet was trained, validated, and tested with the EfficientNet-B0 encoder. At the end, the final test was taken on a 2023 Google Earth Pro image.

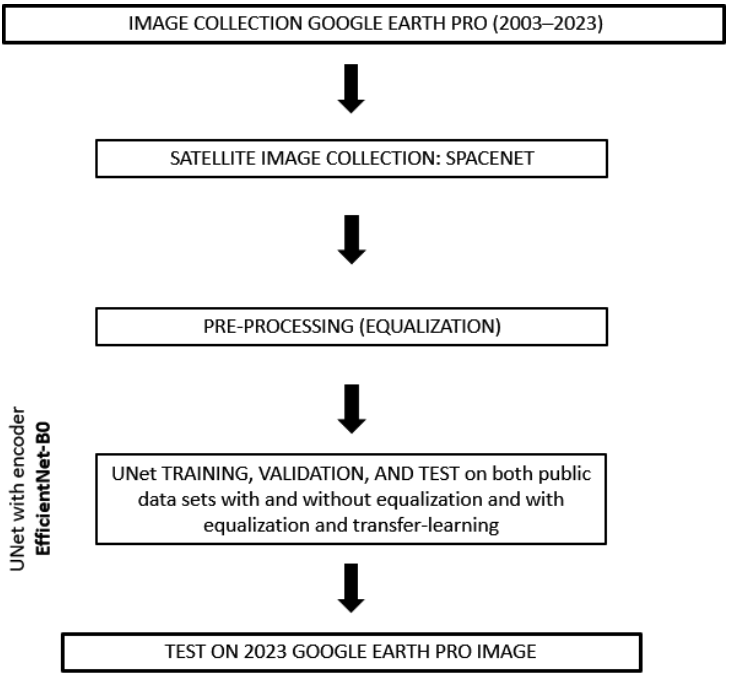

Fig. 3. Workflow scheme

SpaceNet data sets. The SpaceNet data set is a public data set that was inspired by the ImageNet model [20] divided in eight challenges [21]. The first two SpaceNet challenges focus their attention to building footprints. Challenge 1 of SpaceNet is the building footprint in Rio de Janeiro. Challenge 2 includes four areas: Las Vegas,

Paris, Shanghai, and Khartoum (the authors focused their attention on the Shanghai area). For this work, both the public data sets labeled SpaceNet 1 (Rio) and SpaceNet 2 (Shanghai) were considered for the building-footprint-detection task. These were both composed of TIFF images, but SpaceNet 1 was composed of WorldView2 images, and SpaceNet 2 was composed of WorldView3 images (Fig. 4).

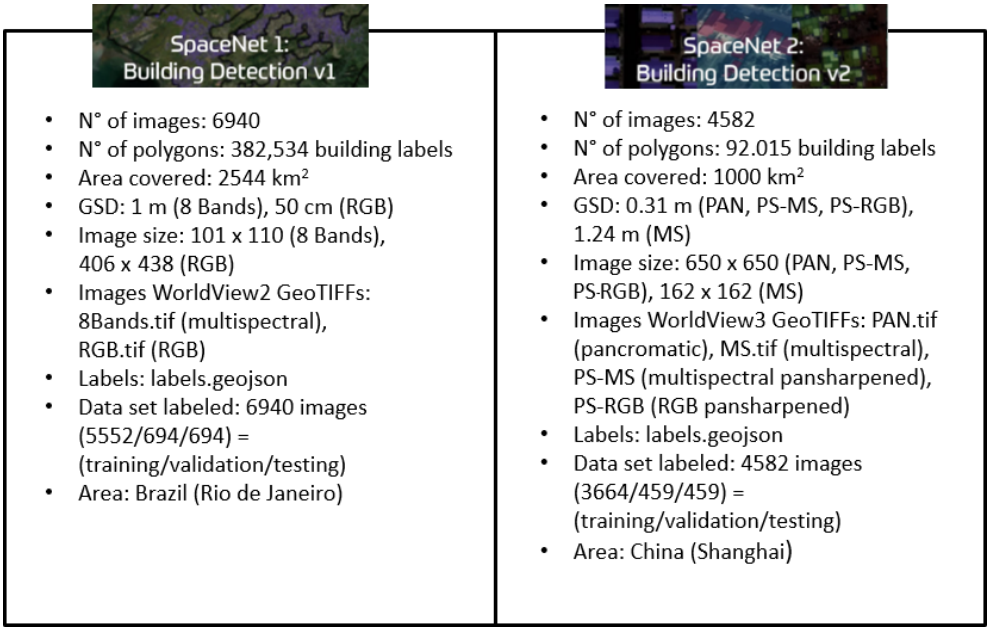

Fig. 4. List of public data set property

The SpaceNet 1 (Rio) data set was composed of 382,534 building-label polygons, with a land cover of 2544 km^2; a single tile covered 200 m × 200 m. The SpaceNet 2 (Shanghai) data set was composed of 92,015 building-label polygons and had a cover area of 1000 km^2. Each image was a 200 m × 200 m tile (Fig. 4). A division of both data sets was established in order to have three partitions in percentages: training (80%), validation (20%), and testing (20%). In SpaceNet 1 (Rio), this partition amounted to 5552/694/694 (6940 total images); in SpaceNet 2 (Shanghai), it amounted to 3664/459/459 (4582 total images).

Pre-Processing. Because these were at different resolutions, an equalization process was necessary (Fig. 5). Only RGB images were considered in this workflow. To conduct the equalization, an RGB image size that was common to both data sets was chosen in order to avoid overfitting. The RGB image size of SpaceNet 1 was 406 × 438 pixels, and the RGB image size of SpaceNet 2 was 650 × 650 pixels. The common size that was chosen for both data sets was 320 × 320 pixels; in this way, the value of the ratio between centimetres and pixels was about 60 for both data sets.

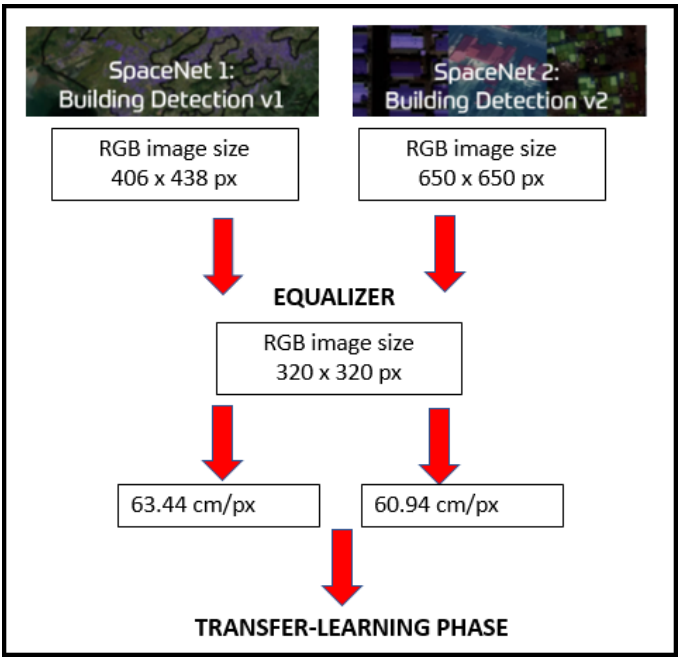

Fig. 5. Equalization process

In addition, the following strategies were applied in the training partitions:

- One of three strategies was applied at a time in order to obtain 320 × 320 pixels images (RandomCrop of 406 × 406 pixels, followed by resizing to 320 × 320 pixels). A random part of the input was cropped and rescaled to 320 × 320 pixels (RandomSizedCrop). A rotation (Rotate) of 15° was applied, followed by a centered crop (CenterCrop) of 320 × 320 pixels.
- Randomly, the input was flipped horizontally (HorizontalFlip) around the y-axis, vertically (VerticalFlip) around the x-axis, and rotated 90 sexagesimal degrees, with a 50% probability of occurrence for each case.
- Random variations in brightness and contrast (RandomBrightnessContrast) were applied in addition to changes in the RGB (RGBShift) and HSV (HueSaturationValue) color representations. This had a probability of 25%.
- Blur, Gauss-Noise, and RandomGamma were applied, with a probability of 25%.

Network Architecture. In the case of this study, a CNN based on the UNet architecture was applied. Each CNN consisted of an input layer and an output layer. A combination of an encoder (to reduce the input image) and a decoder (to amplify the output image) composed the UNet architecture. A scheme of the UNet architecture is shown in Figure 6. The encoder was pre-trained on the ImageNet data set that was organized according to the WordNet hierarchy [22]. The used

encoder corresponded to the EfficientNet B0 architecture and was produced through a multi-objective search neural architecture that optimized both accuracy and efficiency [23].

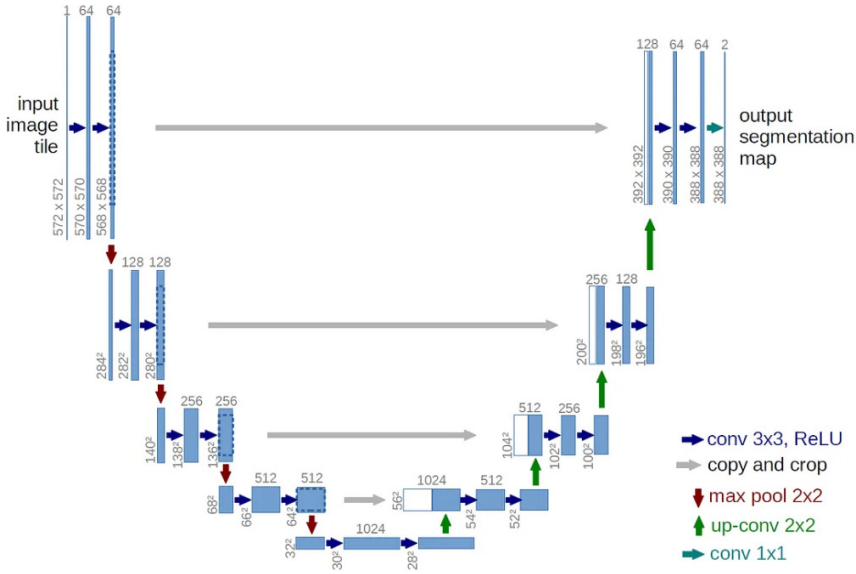

Fig. 6. UNet Architecture

Source: [9]

Experimental design. Two experiments were conducted in order to train UNet. This phase was the transfer-learning phase in which in the first experiment (the first training of the network) was done on Data Set 1 (SpaceNet 1-Rio); then, it was re-trained on Data Set 2 (SpaceNet 2-Shangai). The second experiment was the opposite order (Fig. 7).

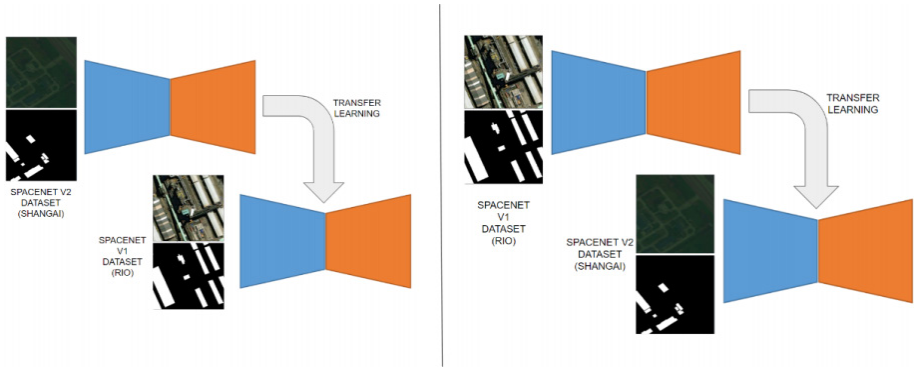

Fig. 7. Training work scheme

Each model was trained with a maximum of 300 epochs and a batch size of 16. Adam was chosen as the optimizer, with default beta 1 and beta 2 (0.9 and 0.9999) and an initial learning rate of 0.001. To avoid overfitting, a ReduceLROnPlateau organizer was used with a reduction factor of 0.1 and a patience of 10 epochs, and an early stop was used with a patience of 20 epochs. The training-optimization process was guided by the weighted combination of the loss functions (cross entropy, Dice, and active contours) [24].

The experimentation was performed using Version 1.9 of the Pytorch deep-learning framework and Nvidia GeForce RTX2800 Ti GPU graphics cards.

Evaluation metrics. The evaluation of the segmentation quality of the models was based on a calculation of the confusion matrices per image and a cumulative confusion matrix. These matrices related the ground truth (GT) pixels and the predicted pixels and were composed of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The following quality metrics were derived from these matrices:

- FPR: the false-positive rate was calculated as the number of false-positive predictions (FP) divided by the total number of false-positive predictions
  FPR = FP/(FP + TN).
- FNR: the false-negative rate was calculated as the number of incorrect negative predictions (FN) divided by the total number of negatives
  FNR = FN/(TP + FN).
- IoU: the false positives were penalized using intersection over union (IoU) or Jaccard's similarity coefficient [25]. It was calculated as a function of
  IoU = TP/(TP + FP + FN).
- Dice: the Dice similarity coefficient (which is based on the principle of the Søresen-Dice coefficient [26]) was calculated as a function of
  Dice = 2TP/(2TP + FP + FN).

4. Results and Discussion

In Figures 8 and 9, the results that were obtained by the UNet architecture with the EfficientNet-B0 encoder training are represented. The results are shown in three different panels: the first refers to a normal RGB image as input, the second is the case of an equalized RGB image, and the third is the transfer learning (TL) equalized RGB image input. It is possible to see how the values of the IoU (intersection of union) change in the three different cases. In particular, it is notable that there was an improvement in the IoU value after the TL phase in the first experiment that was conducted with the SpaceNet 1 data set. The third panel reached an IoU value of 61.03%, which is an increase when compared to the IoU value of the normal RGB image (53.51%) (Fig. 8). The same consideration cannot be done for the experiment that was conducted with the SpaceNet 2 data set (Fig. 9), in which the IoU value of the third panel was less than the value of the IoU of the normal image input.

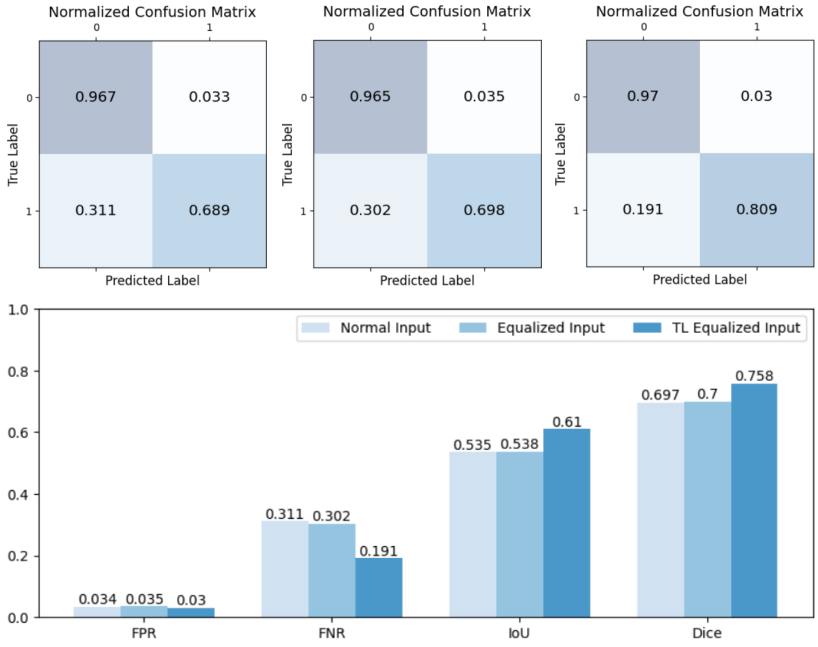

Fig. 8. Model results of UNet architecture with EfficientNet-B0 encoder (SpaceNet 1)

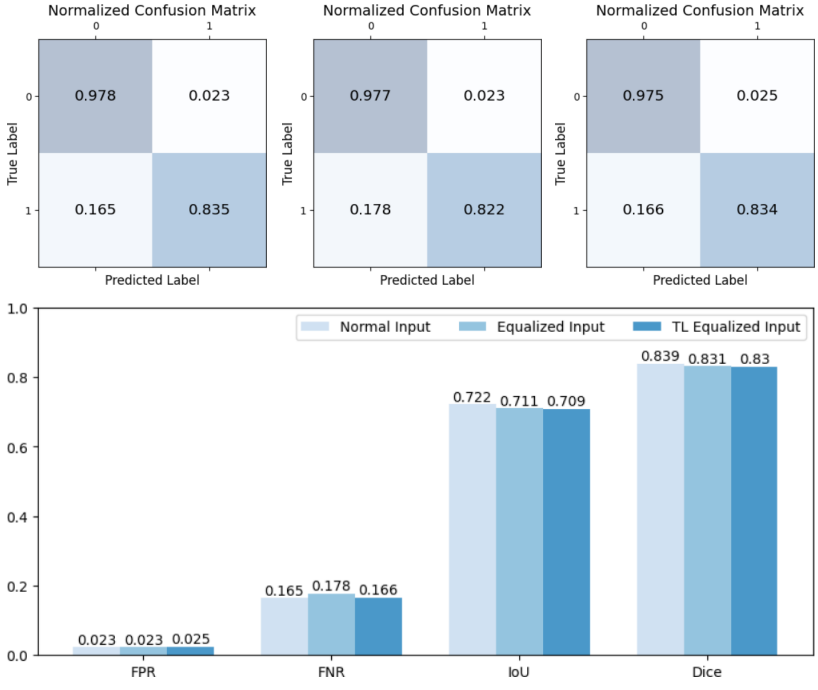

Fig. 9. Model results of UNet architecture with EfficientNet-B0 encoder (SpaceNet 2)

Once the TL was done (and after having tested the UNet on the two different experiments), the choice fell onto the SpaceNet 1 public labeled data set. At this point, the network was tested on a 2023 Google Earth Pro image. Three different scenarios were considered in order to detect in which UNet worked best. The three scenarios were the area with industrial roofs (Fig. 10), the area close to the sea (Fig. 11), and the paved area with high urban density (Fig. 12). Differentiating the three scenarios was useful for seeing where the network worked better and where it experienced more difficulty in its segmentation. It was worthwhile to make some considerations for future improvements. In the black panels (which were the result of the segmentation), four columns of images are present. The segmentation results are in the third column, and the ground-truths (manual annotations) are present in the fourth column.

From a mere qualitative point of view, comparing all the scenarios, the building segmentation that is linked to the presence of the shadows seems to be a problem. In the first scenario where the detection was in an area with plenty of industrial roofs, the quality of the results seemed to be better with respect to the others. In this area, the buildings (being industrial in nature) were bigger than the urban buildings, so the detection met fewer difficulties. In the second scenario, the buildings were closer to each other with respect to Scenario 1 (being in an area that was close to the sea). In the last scenario (in a condition that was very similar to the second scenario), the segmentation seems to have not been very good (especially in the building contours), as the single buildings were too close to each other and the distinctions of the edges of each single building were not evident.

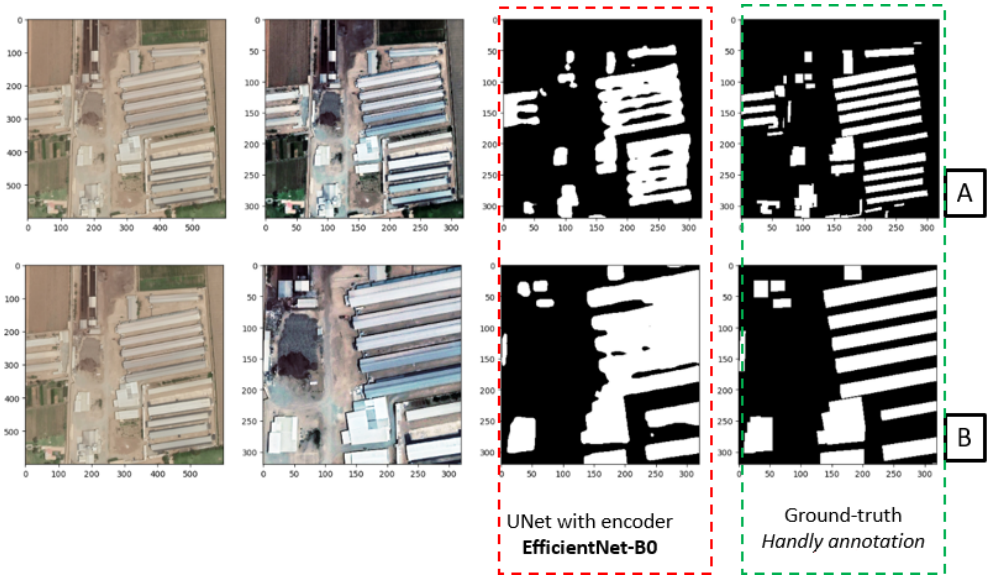

Fig. 10. Results of best evaluated model – area with industrial roofs (Scenario 1)

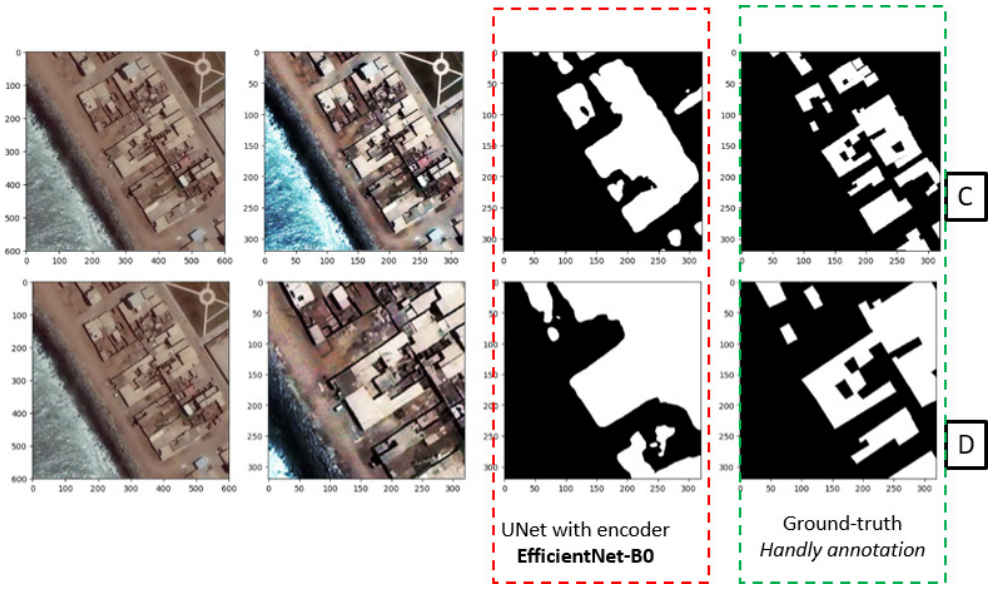

Fig. 11. Results of best evaluated model – area adjacent to sea (Scenario 2)

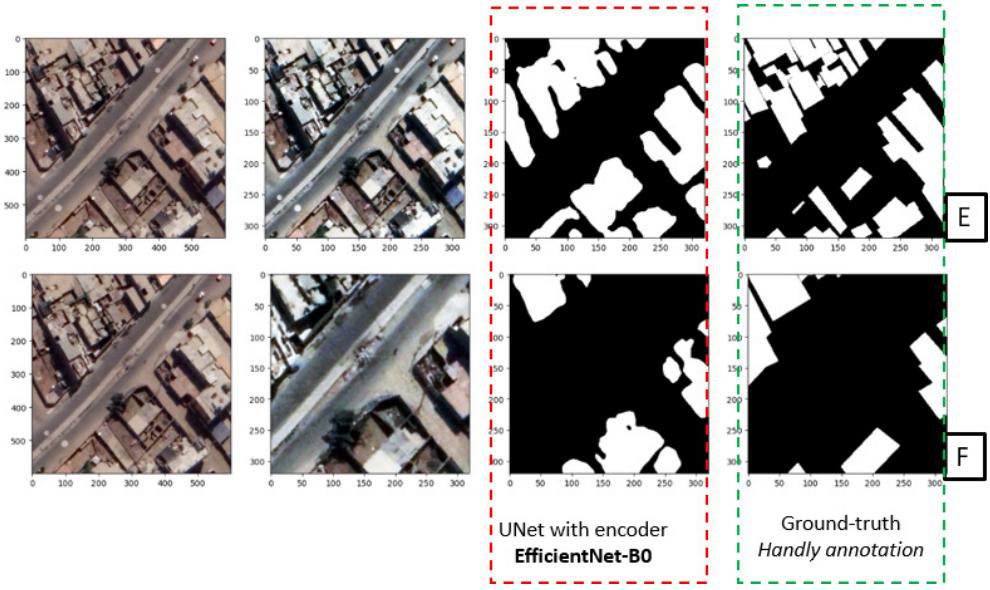

Fig. 12. Results of best evaluated model – paved area (Scenario 3)

From a quantitative point of view, it is possible to validate what was said for the qualitative results based on the IoU values. In Table 1, the IoU values of each scenario (1–3) are shown. It is visible how the IoU decreased when passing from Scenario 1 (A, B) to Scenario 3 (E, F).

Table 1. Quantitative results of evaluated model: A, B – industrial roof area (Scenario 1); C, D – area adjacent to sea (Scenario 2); E, F – paved area (Scenario 3)

Metrics (Batch) – A	Metrics (Batch) – B	Metrics (Batch) – C
FPR: 0.1059	FPR: 0.1404	FPR: 0.1348
FNR: 0.1121	FNR: 0.0488	FNR: 0.0933
IoU: **0.6492**	IoU: **0.7480**	IoU: **0.5897**
Dice: 0.7873	Dice: 0.8558	Dice: 0.7419
Metrics (Batch) – D	Metrics (Batch) – E	Metrics (Batch) – F
FPR: 0.2541	FPR: 0.2010	FPR: 0.0820
FNR: 0.0906	FNR: 0.2260	FNR: 0.3151
IoU: **0.5811**	IoU: **0.5637**	IoU: **0.4930**
Dice: 0.7351	Dice: 0.7210	Dice: 0.6604

Even though UNet is one of the most-famous and most-used neural networks for segmentation, it has some limitations. For example, it tends to have imprecise edges; and sometimes when two buildings are close to each other, it tends to blend them together. The segmentation mask of a building can therefore be disturbed by the presence of objects in the background: trees, cars, shadows, or anything else that is adjacent to the building [3]. This happens especially when there are very small buildings or there is high urban density. In this case, the pixels in the boundary area of the small building are confused with whatever is in the background, so they are omitted from the detection [27]. This difficulty is evident in this work, considering the three different scenarios mentioned above in which the building density levels grew from Scenario 1 to Scenario 3. Comparing the IoU values of the three different situations, it is possible to note that this value decreased after Scenario 1 (in which the edges of the industrial roofs were quite well-defined). In this scenario, there were no strong concentrations of buildings; furthermore, they were large enough to not risk that the network would not recognize them individually. Passing to Scenario 2 and Scenario 3, the IoU values decreased; this was because the urban areas were quite dense in Scenario 2 and highly dense in Scenario 3. This value passed from the highest IoU value of the industrial roof area (64.92%) to the IoU value of the paved areas (56.37%). It is planned for the future to test other CNNs and train the net with other labeled public data sets in order to achieve better segmentation results and avoid these kinds of limitations.

5. Conclusion

The recent development of remote-sensing and deep-learning techniques have become a topic of strong interest in various research fields. The support of deep-learning algorithms has made possible to noticeably reduce processing times regarding the application of semantic segmentation for object detection. There are various fields of application of semantic segmentation: sea and land segmentation, old city transformations, building mapping, disaster management [4, 5], change detection, urban planning, vegetation-cover assessment, and road extraction [28, 29]. The authors propose the use of a UNet network and a semantic segmentation application to extract building footprints from HRV images. UNet was trained on public data sets from SpaceNet challenges. It can therefore be said that this approach seems to give good results – particularly in one of the three scenarios. Its application could be re-proposed in scenarios with the same building conditions; i.e., a roof material that is quite similar to that of Peru, or a distribution of buildings that is quite similar to the study area of this application. This approach has therefore highlighted some gaps in the conditions of high urban densities and better results where a built-up environment is less dense and where the element to be segmented is rather defined. As future developments, the authors believe that it could be interesting to exploit this application methodology in areas other than that of this study and could exploit other available public data sets.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT Author Contribution

M. C.: conceptualization, methodology, software, validation, data curation, investigation, review and editing, writing – original draft preparation.

E. S. M.: conceptualization, methodology, validation, formal analysis, review and editing, supervision, investigation.

M. S.: conceptualization, methodology, writing – original draft preparation, writing – review and editing, validation, investigation, formal analysis.

R. P.: conceptualization, methodology, validation, formal analysis, review and editing, supervision, investigation.

F. C.: conceptualization, methodology, review and editing.

F. J. L. T.: conceptualization, methodology, software, validation, data curation, investigation, review and editing, supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Use of Generative AI and AI-Assisted Technologies

No generative AI or AI-assisted technologies were employed in the preparation of this manuscript.

References

[1] Chidi C.L.: *Urbanization and Soil Erosion in Kathmandu Valley, Nepal*. [in:] Pradhan P.K., Leimgruber W. (eds.), *Nature, Society, and Marginality: Case Studies from Nepal, Southeast Asia and Other Regions*, Perspectives on Geographical Marginality, vol. 8, Springer, Cham 2022, pp. 67–83. https://doi.org/10.1007/978-3-031-21325-0_5.

[2] Pansoni S., Tiribelli S., Paolanti M., Frontoni E., Giovanola B.: *Design of an ethical framework for artificial intelligence in cultural heritage*. [in:] *2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS 2023): West Lafayette, Indiana, USA, 18–20 May 2023*, IEEE, Piscataway 2023, pp. 1–5. https://doi.org/10.1109/ETHICS57328.2023.10155020.

[3] Jung H., Choi H.-S., Kan M.: *Boundary enhancement semantic segmentation for building extraction from remote sensed image*. IEEE Transactions on Geoscience and Remote Sensing, vol. 60, 2022, pp. 1–12. https://doi.org/10.1109/TGRS.2021.3108781.

[4] Cardellicchio A., Ruggieri S., Leggieri V., Uva G.: *View VULMA: Data set for training a machine-learning tool for a fast vulnerability analysis of existing buildings*. Data, vol. 7(1), 2022, 4. https://doi.org/10.3390/data7010004.

[5] Ruggieri S., Cardellicchio A., Leggieri V., Uva G.: *Machine-learning based vulnerability analysis of existing buildings*. Automation in Construction, vol. 132, 2021, 103936. https://doi.org/10.1016/j.autcon.2021.103936.

[6] Yuan X., Shi J., Gu L.: *A review of deep learning methods for semantic segmentation of remote sensing imagery*. Expert Systems with Applications, vol. 169, 2021, 114417. https://doi.org/10.1016/j.eswa.2020.114417.

[7] Wang L., Li R., Zhang C., Fang S., Duan C., Meng X., Atkinson P.M.: *UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 190, 2022, pp. 196–214. https://doi.org/10.1016/j.isprsjprs.2022.06.008.

[8] Bazila F., Ankush M.: *A comparative study of deep learning and traditional methods for environmental remote sensing*. ITM Web of Conferences, vol. 56, 2023, 03002. https://doi.org/10.1051/itmconf/20235603002.

[9] Fu Y., Li J., Weng Q., Zheng Q., Li L., Dai S., Guo B.: *Characterizing the spatial pattern of annual urban growth by using time series Landsat imagery*. Science of The Total Environment, vol. 666, 2019, pp. 274–284. https://doi.org/10.1016/j.scitotenv.2019.02.178.

[10] Mikrut S., Papuci-Wladyka E., Struś A., Puntos J.K., Głowienka E.: *The use of photogrammetry in archaeology and multimedia open-air performance in the Castle Square of Kato Paphos*. [in:] *BGC-Geomatics 2018: 2018 Baltic Geodetic Congress: 21–23 June 2018, Olsztyn, Poland: Proceedings*, IEEE, Piscataway 2018, pp. 353–358. https://doi.org/10.1109/BGC-Geomatics.2018.00073.

[11] Das S., Angadi D.P.: *Land use land cover change detection and monitoring of urban growth using remote sensing and GIS techniques: A micro-level study.* GeoJournal, vol. 87(3), 2022, pp. 2101–2123. https://doi.org/10.1007/s10708-020-10359-1.

[12] Krishnaveni K.S., Anilkumar P.P.: *Managing urban sprawl using remote sensing and GIS.* The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-3/W11, 2020, pp. 59–66. https://doi.org/10.5194/isprs-archives-XLII-3-W11-59-2020.

[13] Coulter L., Hall T., Guzman L., Kasahara I.: *Satellite Image building detection using U-Net convolutional neural network*. https://www.luisjguzman.com/media/EE5561/building_detection.pdf [access: 31.10.2023].

[14] Ronneberger O., Fischer P., Brox T.: *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: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III*, Lecture Notes in Computer Science, vol. 9351, Springer, Cham 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

[15] Gomroki M., Hasanlou M., Reinartz P.: *STCD-EffV2T Unet: Semi transfer learning EfficientNetV2 T-Unet network for urban/land cover change detection using Sentinel-2 satellite images.* Remote Sensing, vol. 15(5), 2023, 1232. https://doi.org/10.3390/rs15051232.

[16] Vasantrao C.P., Gupta N.: *Wader hunt optimization based UNET model for change detection in satellite images.* International Journal of Information Technology, vol. 15(3), 2023, pp. 1611–1623. https://doi.org/10.1007/s41870-023-01167-0.

[17] Singh N.J., Nongmeikapam K.: *Semantic segmentation of satellite images using deep-unet.* Arabian Journal for Science and Engineering, vol. 48(2), 2023, pp. 1193–1205. https://doi.org/10.1007/s13369-022-06734-4.

[18] Colosi F., Gabrielli R., Orazi R., Malinverni E.S.: *Discovering Chan Chan: Modern technologies for urban and architectural analysis.* Archeologia e Calcolatori, vol. 24, 2013, pp. 187–207. https://api.semanticscholar.org/CorpusID:60929933.

[19] Malinverni E.S., Pierdicca R., Colosi F., Orazi R.: *Dissemination in archaeology: A GIS-based StoryMap for Chan Chan.* Journal of Cultural Heritage Management and Sustainable Development, vol. 9(4), 2019, pp. 500–519. https://doi.org/10.1108/JCHMSD-07-2018-0048.

[20] Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L.: *ImageNet: A large-scale hierarchical image database*. [in:] *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009): Miami, Florida, USA, 20–25 June 2009*, IEEE, Piscataway, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

[21] Van Etten A., Lindenbaum D., Bacastow T.M.: *SpaceNet: A remote sensing dataset and challenge series.* 2018. https://doi.org/10.48550/arXiv.1807.01232.

[22] ImageNet. https://www.image-net.org/ [access: 20.03.2023].

[23] Tan M., Le Q.V.: *Efficientnet: Rethinking model scaling for convolutional neural networks*. [in:] *36th International Conference on Machine Learning (ICML 2019): Long Beach, California, USA, 9–15 June 2019*, Proceedings of Machine Learning Research, vol. 97, International Machine Learning Society, Stroudsburg 2019, pp. 10691–10700. https://doi.org/10.48550/arXiv.1905.11946.

[24] Chicchon M., Bedon H., Del-Blanco C.R., Sipiran I.: *Semantic segmentation of fish and underwater environments using deep convolutional neural networks and learned active contours*. IEEE Access, vol. 11, 2023, pp. 33652–33665. https://doi.org/10.1109/ACCESS.2023.3262649.

[25] Sørensen T.: *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons*. Kongelige Danske videnskabernes selskabs. Biologiske skrifter, bd. 5(4), Ejnar Munksgaard, København 1948.

[26] Rizwan I Haque I., Neubert J.: *Deep learning approaches to biomedical image segmentation*. Informatics in Medicine Unlocked, vol. 18, 2020, 100297. https://doi.org/10.1016/j.imu.2020.100297.

[27] Chen F., Wang N., Yu B., Wang L.: *Res2-Unet: A new deep architecture for building detection from high spatial resolution images*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, 2022, pp. 1494–1501. https://doi.org/10.1109/JSTARS.2022.3146430.

[28] Sun Y., Bi F., Gao Y., Chen L., Feng S.: *A multi-attention UNet for semantic segmentation in remote sensing images.* Symmetry, vol. 14(5), 2022, 906. https://doi.org/10.3390/sym14050906.

[29] Abdollahi A., Pradhan B., Shukla N., Chakraborty S., Alamri A.: *Multi-object segmentation in complex urban scenes from high-resolution remote sensing data.* Remote Sensing, vol. 13(18), 2021, 3710. https://doi.org/10.3390/rs13183710.