# Final Sati project.pdf

*by* Ved Prakash Chaubey

**Submission date:** 02-May-2024 10:25PM (UTC+0530)
**Submission ID:** 2368984862
**File name:** Final_Sati_project.pdf (769.88K)
**Word count:** 3649
**Character count:** 20516

# Satellite Image Segmentation

Chebrole h v n v s uday kiran
*Lovely Profesional University,*
Punjab, India
Chebroleuday@gmail.com

Devaraj goli
*Lovely Professional University,*
*Punjab, India*
Devagoli66@gmail.com

Sai Shankar Gattu
*Lovely Professional University,*
*Punjab, India*
saishankar2748@gmail.com

Gadge Ajay
*Lovely Professional University,*
*Punjab, India*
saishankar2748@gmail.com

Ved Prakash Chaubey
*Upgrad Campus,upGrad*
*EducationPrivate Limited,*
*, India*
vedeprakesh05@gmail.com

***Abstract*--** **According to projections, India's urbanization rate will reach 35.9% by 2022. In the present situation 45.23% is from the expansion happening in Maharashtra, and this state is third most expanded state in India, even after Tamil Nadu and Kerala. The categorization of land cover constituted the metropolitan areas' object of remote sensing over the last decades. Featured with complex architecture and insufficient labeled data, this task turns out to be quite difficult while classifying buildings from almost very high-resolution (VHR) satellite images of the metropolitan area. Creating a classification model conventionally consists of hand-engineered features of transfer-learning strategies. This is done in multiple ways that have difficulties with the diverse types of buildings (shape, point of view), resulting in less accuracy in large metropolitan sector and bad functioning with great-resolution satellite images. A deep learning-based framework with U-Net architecture is employed to perform semantic segmentation for the building classification. Spatial data of urban area images in 0.5m resolution was computed by the SASPlanet Model. One hot encoding is being applied to classify buildings by category. U-Net is trained on target data, which is in an encoded format. Based on that Indian dataset, particularly the dense populated locations of Nashik city in Maharashtra, the most accuracte results are 60% and 85% of the building detection. Change detection is carried out by the comparison between the two temporal image frames. The GIS maps are recreated to indicate building shifts, displayed by different colors for buildings under construction, maintained ones, and destroyes ones, respectively.**

*Keywords - Feature Extraction, VHR images, Semantic Segmentation, U-net, Building Classification, Geo Referencing*

## I. INTRODUCTION

The structures are the main parts that determine the form of urban areas, affecting the types of land use, the levels of population density, and the sustainability of the city. The building types, distribution, and conditions comprehended by the urban planners are necessary for good and effective decision-making. The conclusion of buildings from satellite images by means of deep learning algorithms gives, therefore, important information on the type and structure of urban areas. U-Net, a popular deep learning network for image segmentation that is designed specifically for this task, is one such architecture. Thanks to a simple symmetric encoder and decoder design, U-Net can abstract huge level features right from the input images and decode them to generate the segmentation masks. It means that buildings can be reliably distinguished and classified based on such factors as roof shape, size, and color. U-Net is a compound structure that has contracting and expanding paths on both sides of the middle layer, with the narrowest one. The encoder expands essential features from the input photos and the decoder recalls the segmentation map using these method. This strategy has demonstrated its effectiveness in several areas, such as building classification from satellite photography. Through the use of U-Net, it is possible for urban planners to perform more accurate building density analysis, which indeed makes it possible to make better decisions regarding land use planning and infrastructure development. This technology provides a good possibility for building classification from the satellite images, and thus it is useful in urban planning procedures and contributes to the creation of smart and sustainable city plans.

## II. LITERATURE REVIEW

In this paper (Reda and Kedzierski, 2020)[20], the Urban and Object Net, or U-Net, was used to determine models in dense populated and residential areas. A most used R-CNN model architecture was employed that contains both a feature extractor and an RPN layer for making structure hypotheses. While they presented the new method for creating border adjustments, they also introduced other manufacturing improvements, such as the shape they specified for the boundaries of triangles and squares. Data collection is an important element of our study. Therefore, author made a database of 500 photos that featured city with different shape and architecture. In this aspect, the Adam and RMSProp optimizations could be responsible for the best results obtained for structure extraction and classification. Yet the model displayed difficulties in correctly detecting garages.

The article (Lloyd et al., 2020)[17] was devoted to the GIS implementation of the semi-automatic method that was to be used to create a new global spatial data layer classifier. This model was based on the retrieval of simplistic classifications and relied on photo extraction algorithms to construct the footprint data. The workflow had set up the integration of diverse data sources, which were utilized to refine and improve the model's operational precision. Our ensemble model allowed us to classify the type of housing with an accuracy of 85% to 93%, choosing between residential and non-residential. Nevertheless, the model's transferability at the local or regional level has not been explored.

Zhao et al. (2018)[24] used Deep Neural Network (DNN) and Incremented-Extended Profile (IEP) IDs for style extraction purposes. It is also used for assessments of the similarities and dissimilarities in the styles of architecture. In the building of a model, to enchance module data google net inception techniques for the purpose of lessening computing costs and stopping model overfitting. Another impressive feature of AMILE is its performance, which achieved 25 different styles of architecture categorization. It has achieved a promising 98.57 percent accuracy in recognizing architectural styles. Nevertheless, the applicability of this model of recognition works just for Greek and Georgian styles of architecture.

Kang et al. (2018)[13] have thusly thought up a structure for the classification of land use by building type. Here, GIS maps were used for the extraction of building footprints, with outliers removed, and what is called U-Net for building categorization, where in the first place the public VGG16 model was used. The developed model showed a great deal of predictive power in assigning a certain land use to individual buildings. However, challenges arose when metropolitan building plots sometimes necessitated the development of other means of data retrieval from the GIS maps.

Another human (Huang et al., 2017)[10] gave us a method that is able to classify different types of buildings using the information gathered by high-resolution remote sensing images and LiDAR data. The LiDAR data on the height of buildings, in addition to the image data, as well as spatial and autocorrelational relationships, were considered. Application of this method encompassed spatial features and autocorrelation, with a more robust classifier performance. The trials was not comprehensive for validation and complete testing either in the end, it did not manage to pinpoint the contribution of landscape to the results.

In scenario case (Wurm et al., 2015),[23] LDA was used to do the classifications using Plots of building area can be obtained by utilizing real estate cadastral data to search digital surface models that are produced from aerial pictures. It facilitates the grouping of buildings into several categories. They came up with a set of 26 shape-based attributes that have the potential to yield accurate photo matching, where thickness, sphericity, and 2-D assessments of shortness max were the most relevant. The advantages of this approach lie in the possibility of carrying out elaborate research (1000 runs) to decrease favourable in data sets as well as to finalize high sensitivity and excellent precision. Hence, the LDA comes with a caveat that restricts its applicability, and the complex features, like blocks, which are frequently similar in form and design, such as circumference block development are sometimes difficult to differentiate.

Author's article itself (Kaichang et al., 2000)[12] emphasized two learning approaches from spatial data based on inductive instruction. One of the approaches outlined involved using two learning levels, namely pixels and spatial objects. It defined rules for the classification of pictures by means of spectrum, position, and altitude. A merit was the fact that overall accuracy increased considerably by approximately 11%, especially in land, in categories like residential areas, paddy fields, irrigated fields, vegetable fields, and water about 94.4% accuracy were obtained. GIS data used for image categorization and data mining-based techniques is one of the factors that contribute to this accuracy rate. Although such an integration did arise, the precise detection of forest shadows as streams successfully remained a long-standing problem. According to the study by Goldblatt et al. (2016)[7], there are three ways to classify images: collection, selection, and pre-processing of data; scene search; and pixel-level classification for built-up region detection. The classifier is observed to sustain its consistency of performance with various land cover factors both in the train and test parts, finally yielding an accuracy of around 87%. however, there is no information provided about class variables, physical features, or location data, which could lead to an increase in the accuracy of the classifier.

## III. METHODOLOGY

The delineation and classification process flowchart for structures into appropriate categories in metropolitan areas is outlined in fig. Comprises the evaluation of model and training and testing process.

To start, a dataset of 288 photos was randomly chosen from SASPlanet for the sake of preparation.

a. In terms of pre-processing, different methods, such as resizing, enhancing, and removing noise, are performed before the processed data.

b. For instance, red was used to represent building objects, blue for industrial industries, and green for holy religious sites. The developed map was done in QGIS, where about 288 images were masked.

c. The sequence of the actions, including semantic segmentation, feature extraction, object detection, and categorization of the built-up area into residential, industrial, and holy contrast, in short.

d. That is, out of the total 256 photos, 0.8% were used for training, while 0.2% were used for testing.

The true highlight of my task was that it consisted of two parts: training and testing of image recognition. We have done, among other tasks, the resizing of the image to 224 x 224 and also applying the median filter to remove noise. Through the QGIS mask tool, these images were masked for better visualization.

The process of detection was carried out by means of semantic segmentation, and based on the unlabeled datasets, some information was revealed. In this connection, U-NET segmentation was the method used. The model development and model diagnosis processes were made easier with the help of a labeled data set. The findings were then ultimately verified. The model got the training done for 25 epochs; they used the satellite imagery data set in high resolution. The batch was composed of 16 elements. (1) The target machine was one single GPU with high performance. These entities themselves consumed the 8 GB of graphic processing unit memory and system memory. Lastly training process takes around 4 hours. It comprises 213 MB, resulting from 28 million parameters. That was about 89% of the model efficiency.

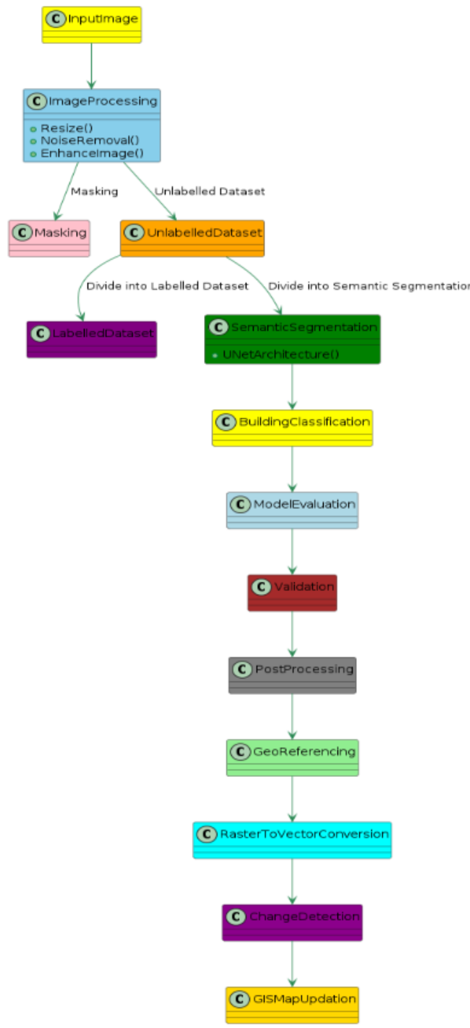



**Fig.1**.Methodology flow chart of the project.

Image resolution: 224 x 224 px.

Validation data: we used the Mumbai, India, dataset as the validation model with dimensions of 6097 x 9058 pixels. The picture is scaled to 224 x 224. Following the patching of our images, we will end up with 288 images of size 224px x 224px.

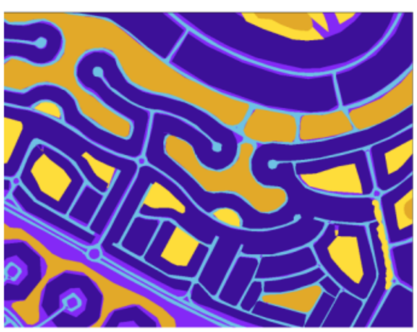
Evaluation data: dataset types like LEVIR-CD (Mohammad et al., 2022), SpaceNet (Van Etten et al., 2018), and WHU Buildings (Ji et al., 2018) are employed for testing purposes.

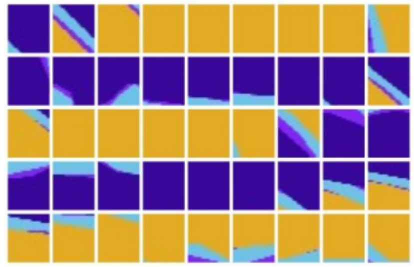
## IV. RESULTS

Actually images were downloaded from SASPlanet and also the masking is done.

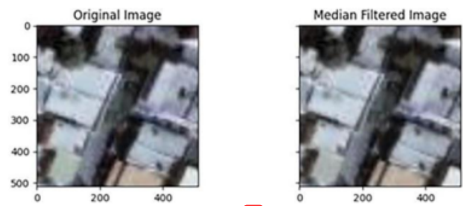
Patchify and mask operations that split each images with the size of 12025 x 5878 are patched into size of 224 x 244 are the spatial. Images are cut into patches with size 224 x 224px.

i. Images from SASPlanet and its corresponding patches.





ii. Removing useless images.




iii. Model displays the test image, the test labels and the prediction.

**Fig.2** The flow chart of working process of the model.

MSE represents mean squared error of the two pictures, and MAX represents maximum pixel value of this picture which are shown above.

Through the phase of training when then buildings, the white area, and other items are well represented. The disruption, for instance, of that project, disrupts the data, which exactly that initially coded position or coordinate information is not available. The injuring ad is localized and this is followed by spatial data restoration. This can be done by photos superimposing on a map or by displaying the images plus other geographic information from the geographic information system (GIS).

In this way this goal can be achieved either by laying of photo on a map or showing pictures and this also requires the support of geographical information system (GIS).One can accomplish this goal simply through the use of a photo superimposed onto a map or co-displayed with other geographic information in the geographic information system (GIS).

This can be done by photos superimposing on a map or by displaying the images plus other geographic information from the geographic information system (GIS).Such purpose can be achieved through two means, photo projection on a map or visualization of an image with other geographical information in the geographical information system (GIS).This goal can be achieved through image overlapping on a map or presentation of a photo with other geographic information in the geographical information system (GIS).

The GeoTIFF format is used to structure the scan pattern datasets of satellite images, and GeoJSON format is used for direction datasets that alter the existing GIS maps to display the change detections of infrastructure. After that process of locating coordinates and overlaying pixels on a GIS base map, a GeoTIFF file will be obtained, which is the raster picture. Base vectors are inherent in vector-based GIS maps. So, consequently, the GeoTIFF file (raster image) gets incorporated into the GeoJSON file (geospatial vector format).

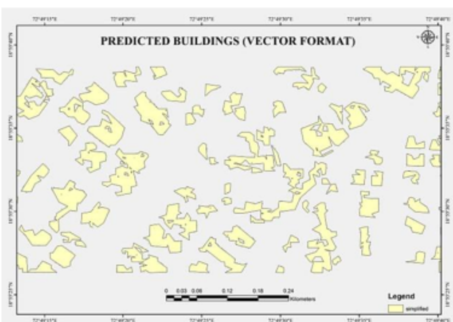



**Fig.3** Dataset distribution.

Weaving mixed-use buildings with residential category in India could appear detached or contrary from the satellite view. And this is why we only recognize three kinds of infrastructures as residential, manufactures and holy ones. Masking is an operation that can be realized using QGIS software. Fig presents a satellite image of Bombay in 2020 and the binary mask for the same year. The next step is going to be training the model using given image as input. We will be checking which model performed the best after training the model with all the images which are present in the dataset.

We will also be checking if the model is performing accurately when it comes to unseen data so that we can evaluate the model performance at every stage.


**Fig.4**.Satellite image from2023

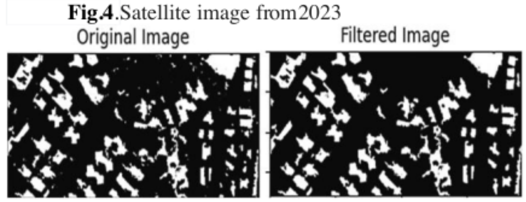



**Fig.5**.Binary mask of buildings.

Fig. exhibited the removal of the white area applying the thresholding approach. At the end, a set of objects from limited area defines the criteria (10,000 pixels). Pixels greater than 10000 are saved, nonetheless pixels smaller than 10000 may be deleted.

## V. Performance analysis

The train Loss vs Val loss in the Training epochs shown in graph. The Training loss line is coloured yellow and val loss is blue. From the training graph we can see that how our model performed while training and how did I all go. The graphs which we have plotted had showed us that our models had worked well in comparison to the Unet model which we have trained.

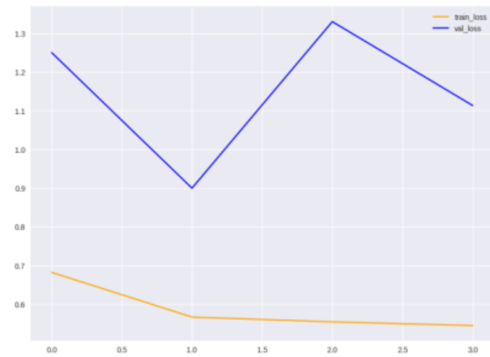



**Fig.6** Train loss vs Validation loss**.**


**Fig.6.1** Train accuracy vs val accuracy

## VI. CONCLUSION AND FUTURE WORK

Life on earth interminably manifests these ongoing patterns. The pace at which institutional memory of urban development processes such as disaster response, environment monitoring, security, and property management systems is constituted and then retrieved from records is too rapid. Our accuracy of the model, which amounts to according (scratch) 73%, we have used pre trained model which was used for extracting the weights from the image and the weights were used by unet model which gave us accuracy of 77%. Although, in Maharashtra, the dwelling homes and manufacturing buildings look similar, the model did not perform well enough to classify them correctly.

This evaluation shall be the basis on which planners will rely on concrete facts to come up with valid choices that will result from data collection involving infrastructure, zoning, and resource allocation. Ally the community by having it involved and participating actively in these community development projects using this environmental information. There will also be the mentoring of local communities and the respective public, researchers and organizations concerning the development, resulting in sound urban development.

This complexity design poses a challenge and results in low performance when complex geometries, such as non-rectangular or irregular, are presented to the model. Terrain aspect orientation is another factor that can be determinant as to the level of accuracy of the models, and the axis distribution of the objects on the same site can have an impact on the spotting of the buildings with slanted orientations from a satellite. Highly populated dense cities, complex spatial arrangements of skyscrapers, and the accumulation of urban construction pose obstacles to the generalizing power of the generative adversarial network in an attempt to trick a network. We do our job with determination, not at the mercy of these hindrances. On the one hand, the categories of buildings that include mosques, churches, and informal settlements in urban areas that admix used buildings also have to be included. The same breakdown will be used for the other sub-groups too. Furthermore, it is planned to make the integrated model powerful enough to include more shape parameters.

## VII. 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.

## VIII. REFERENCES

[1] Zhao, Peipei, et al. "Architectural style classification based on feature extraction module." *Ieee Access* 6 (2018): 52598-52606.

[2] Akçay, H. Gökhan, and Selim Aksoy. "Detection of compound structures using hierarchical clustering of statistical and structural features." *2011 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2011.

[3] Alsabhan, Waleed, Turky Alotaiby, and Basil Dudin. "Detecting buildings and nonbuildings from satellite images using U-Net." *Computational Intelligence and Neuroscience* 2022 (2022).

[4] Alsabhan, Waleed, and Turky Alotaiby. "Automatic building extraction on satellite images using Unet and ResNet50." *Computational Intelligence and Neuroscience* 2022 (2022)..

[5] Ansari, Mohd Aquib, Diksha Kurchaniya, and Manish Dixit. "A comprehensive

analysis of image edge detection techniques." *International Journal of Multimedia and Ubiquitous Engineering* 12.11 (2017): 1-12.

[6] Erdem, Fırat, and Uğur Avdan. "Comparison of different U-net models for building extraction from high-resolution aerial imagery." *International Journal of Environment and Geoinformatics* 7.3 (2020): 221-227.

[7] Goldblatt, Ran, et al. "Detecting the boundaries of urban areas in india: A dataset for pixel-based image classification in google earth engine." *Remote Sensing* 8.8 (2016): 634.

[8] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[9] Vasavi, S., Hema Sri Somagani, and Yarlagadda Sai. "Classification of buildings from VHR satellite images using ensemble of U-Net and ResNet." *The Egyptian Journal of Remote Sensing and Space Sciences* 26.4 (2023): 937-953.

[10] Huang, Yuhan, et al. "A novel building type classification scheme based on integrated LiDAR and high-resolution images." *Remote Sensing* 9.7 (2017): 679.

[11] Ji, Shunping, Shiqing Wei, and Meng Lu. "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set." *IEEE Transactions on geoscience and remote sensing* 57.1 (2018): 574-586.

[12] Kaichang, Di, Li Deren, and Li Deyi. "Remote sensing image classification with GIS data based on spatial data mining techniques." *Geo-spatial Information Science* 3.4 (2000): 30-35.

[13] Kang, Jian, et al. "Building instance classification using street view images." *ISPRS journal of photogrammetry and remote sensing* 145 (2018): 44-59.

[14] Katpatal, Yashwant Bhaskar, Abhijeet Kute, and Deepty Ranjan Satapathy. "Surface-and air-temperature studies in relation to land use/land cover of Nagpur urban area using Landsat 5 TM data." *Journal of urban planning and development* 134.3 (2008): 110-118.

[15] Kay, Susan, John D. Hedley, and Samantha Lavender. "Sun glint correction of high and low spatial resolution images of aquatic scenes: a review of methods for visible and near-infrared wavelengths." *Remote sensing* 1.4 (2009): 697-730.

[16] Livne, Michelle, et al. "A U-Net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease." *Frontiers in neuroscience* 13 (2019): 422707.

[17] Lloyd, Christopher T., et al. "Using GIS and machine learning to classify residential status of urban buildings in low and middle income settings." *Remote Sensing* 12.23 (2020): 3847.

[18] Mohammad, Afrose, et al. "Updating of GIS maps with Change Detection of Buildings using Deep Learning techniques." *2022 International Conference on Futuristic Technologies (INCOFT)*. IEEE, 2022.

[19] Ansari, Mohd Aquib, Diksha Kurchaniya, and Manish Dixit. "A comprehensive analysis of image edge detection techniques." *International Journal of Multimedia and Ubiquitous Engineering* 12.11 (2017): 1-12.

[20] Reda, Kinga, and Michal Kedzierski. "Detection, classification and boundary regularization of buildings in satellite imagery using faster edge region convolutional neural networks." *Remote Sensing* 12.14 (2020): 2240.

[21] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer International Publishing, 2015.

[22] Van Etten, Adam, Dave Lindenbaum, and Todd M. Bacastow. "Spacenet: A remote sensing dataset and challenge series." *arXiv preprint arXiv:1807.01232* (2018).

[23] Wurm, Michael, Andreas Schmitt, and Hannes Taubenböck. "Building types' classification using shape-based features and linear discriminant functions." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.5 (2015): 1901-1912.

[24] Zhang, Tao, and Hong Tang. "Built-up area extraction from Landsat 8 images using convolutional neural networks with massive automatically selected samples." *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Cham: Springer International Publishing, 2018.

# Final Sati project.pdf

ORIGINALITY REPORT

| 9% | 2% | 9% | 1% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | S. Vasavi, Hema Sri Somagani, Yarlagadda Sai. "Classification of buildings from VHR satellite images using ensemble of U-Net and ResNet", The Egyptian Journal of Remote Sensing and Space Sciences, 2023<br>Publication | 5% |
| 2 | pubmed.ncbi.nlm.nih.gov<br>Internet Source | 1% |
| 3 | Badavath Srikanth, Shamneesh Sharma, Ved Prakash Chaubey, Aman Kumar. "Forecasting the Prices using Machine Learning Techniques: Special Reference to used Mobile Phones", 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), 2023<br>Publication | 1% |
| 4 | doaj.org<br>Internet Source | 1% |
| 5 | thesai.org<br>Internet Source | <1% |

| | | |
|---|---|---|
| 6 | Anastasios Temenos, Eftychios Protopapadakis, Anastasios Doulamis, Nikos Temenos. "Building Extraction from RGB Satellite Images using Deep Learning: A U-Net Approach", The 14th PErvasive Technologies Related to Assistive Environments Conference, 2021<br>Publication | <1 % |
| 7 | dspace.dtu.ac.in:8080<br>Internet Source | <1 % |
| 8 | res.mdpi.com<br>Internet Source | <1 % |
| 9 | H. Taubenböck, A. Droin, I. Standfuß, F. Dosch, N. Sander, A. Milbert, S. Eichfuss, M. Wurm. "To be, or not to be ‘urban’? A multi-modal method for the differentiated measurement of the degree of urbanization", Computers, Environment and Urban Systems, 2022<br>Publication | <1 % |

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | < 3 words |
| Exclude bibliography | On | | |