

A U-Net Model for Urban Land Cover Classification Using VHR Satellite Images

Mohamed Fawzy^{1,2*}, Arpad Barsi¹

¹ Department of Photogrammetry and Geoinformatics, Faculty of Civil Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, H-1111 Budapest, Hungary

² Civil Engineering Department, Faculty of Engineering, South Valley University, 83523 Qena, Egypt

* Corresponding author, e-mail: mohamed.fawzy@emk.bme.hu

Received: 04 June 2024, Accepted: 13 September 2024, Published online: 26 September 2024

Abstract

Urban settings are dynamic, constantly changing, and presenting a wide range of surface materials with high diversity in both spatial and spectral variation. As a result, mapping urban growth, evaluating infrastructure, managing water resources, and monitoring natural land cover become more complex tasks. Urban applications have made considerable progress thanks to the abundance of VHR orbital data and the recent development of artificial intelligence strategies especially neural networks. Convolutional neural networks have the potential to significantly enhance the analysis of urban land cover by addressing the limitations of traditional techniques. U-Net is a popular neural network for land cover analysis in remote sensing images. The current research presents a CNN model employing U-Net for image semantic segmentation in urban study area using both spectral and spatial context of VHR satellite data. The proposed model is trained, validated, and tested for VHR satellite image classification into five urban classes: water, vegetation, bare soil, road, and building. The CNN semantic segmentation results are compared to maximum likelihood image classification outcomes for validation and stability evaluation. A confusion matrix is applied to the classified scenes to determine the overall accuracy, producer's and user's accuracy, and Kappa coefficient using 400 random points with their corresponding ground truth. The U-Net image semantic segmentation technique achieved an overall accuracy of 87.50% and Kappa coefficient of 0.8395 which outperforms the maximum likelihood classification method with an overall accuracy of 83.25% and Kappa coefficient of 0.7812.

Keywords

VHR satellite images, convolutional neural networks, land cover classification, urban environments

1 Introduction

Remote sensing techniques have a wide range of applications for sustainable urban development by offering up-to-date information to cover different facets of planning and metropolitan management [1]. Recent Earth observation satellites provide a wide range of valuable spatial and quantitative information, large geographic coverage, and real-time monitoring, which successfully address the challenges related to urban environments [2]. Monitoring urban changes at both spatial and temporal scales is necessary to improve our understanding of how urban development impacts natural resources and environmental systems. Using satellite imageries, for urban land cover analysis, has several benefits such as the reduced time and effort requirements, the inexpensive data collection, and the up-to-date land cover views. Very High-Resolution (VHR) satellite imagery is an effective source

for tracking the expansion of cities, detecting changes in Land Use and Land Cover (LULC), and analyzing urban growth [3]. VHR image interpretation faces several challenges in built-up regions due to the complexity of urban landscapes, the diversity of urban land covers, the disparities in the sizes and shapes of urban objects, the shadows of trees and buildings, and the spectral similarities between different classes [4]. These limitations complicate the accurate land cover classification and the identification of specific components in the metropolitan environments using traditional classification methods. Fortunately, the latest advance in Artificial Intelligence (AI) and remote sensing strategies have offer practical solutions for these concerns. Integration of remote sensing and AI approaches provide affordable tools for gathering and analyzing large amounts of data in comparison to resource-intensive

traditional methods like human based surveying and field observation [5]. AI techniques offer several advantages for effectively managing satellite images in the context of big data analytics such as speed, precision, scalability, automation, and enhanced visual insights [6]. AI enhances image processing accuracy and efficiency, providing valuable information for urban development, infrastructure management, and environmental surveillance. AI models automate the feature extraction process to identify urban items like buildings, roads, and other land cover features using different image forms and patterns [7]. AI tools interpret VHR satellite images employing cutting-edge CNNs to identify relevant urban characteristics and overcome urban environment challenges [8]. CNNs apply semantic segmentation for complex urban scenes based on the contextual information in VHR satellite images [9]. Numerous studies have demonstrated the versatility and effectiveness of deep learning models such as U-Net, GoogLeNet, AlexNet, and ResNet50, for urban image interpretation [4, 10–12]. Consequently, CNN-based image segmentation methods facilitate a variety of urban applications, including building classification, smart city initiatives, road dataset creation and update, transportation management, and land cover features analysis providing rich information for decision-makers [10, 13].

2 Methodology

The current research addresses using an adjustable model to deal with multi-spectral images with VHR spatial resolution to extract detailed urban features. The proposed model is distinct from previous works and surpasses the traditional models based on the versatility to fit different types of input data, different spectral and spatial resolution, and various urban land covers. The presented work aims to explore the convolutional neural networks, like U-Net, to perform semantic segmentation of VHR satellite images for urban areas. To achieve the research objective, a U-Net model is developed, trained, validated, and applied to a VHR satellite image for the semantic segmentation of an urban study area using the following methodology (Fig. 1):

- Determining the required study area and selecting the appropriate satellite image.
- Implementing the preprocessing techniques to provide the raw satellite image in a more relevant format for the classification process.
- Applying a CNN model using U-Net by selecting label data, training the network, and segmenting the input image into the target classes.

- Employing the Maximum Likelihood (ML) pixel-based classifier, as a standard approach, to verify the obtained results and assess the presented model stability.
- Assessing the performance of the proposed CNN and ML models for image classification using a confusion matrix based on ground truth points.

3 Experimental works

3.1 Study area

Egypt is a prominent Middle Eastern country with varying degrees of urban areas. Qena Governorate is one of the ancient provinces, within Upper Egypt, with several land cover classes. It occupies a location between latitude 25°42'24.45", 26°43'40.4" N and longitude 31°52'4.56", 33°36'0.64" E encompassing over 11389 square kilometers. The defined research area is represented by five land cover classes: water, vegetation, bare soil, buildings, and roads (Fig. 2 (a)).

3.2 Data used

The WorldView-2 imaging and environmental monitoring satellite was launched to meet the increasing commercial demand for high-resolution satellite imagery. A WorldView-2 image is used comprises a panchromatic band with 0.5m spatial resolution and 8 multi spectral bands (coastal blue, blue, green, yellow, red, red edge, NIR1, and NIR2) with 2.00 m spatial resolution (Fig. 2 (b)). The WorldView-2 sensor captures higher spectral and spatial resolution compared to other VHR sensors with relatively narrow ranges in the visible and near-infrared spectrum (Fig. 3) [14]. The used image datum is World Geodetic System (WGS84) and the map projection is Universal Transverse Mercator (UTM) zone 36.

3.3 Image preprocessing

Pre-processing techniques are necessary to enhance the satellite raw images for a more efficient image analysis tasks. The applied pre-processing procedures focus on two main processes: data fusion and shadow correction [15].

3.3.1 Data fusion

Optical satellite images feature both Panchromatic (PAN) and Multi-Spectral (MS) bands. The PAN band provides a high spatial resolution meanwhile the MS bands offer high spectral with low spatial resolution [16]. The data fusion (pan-sharpening) process is applied to integrate both high spatial and multi spectral resolution in one pan-sharpened

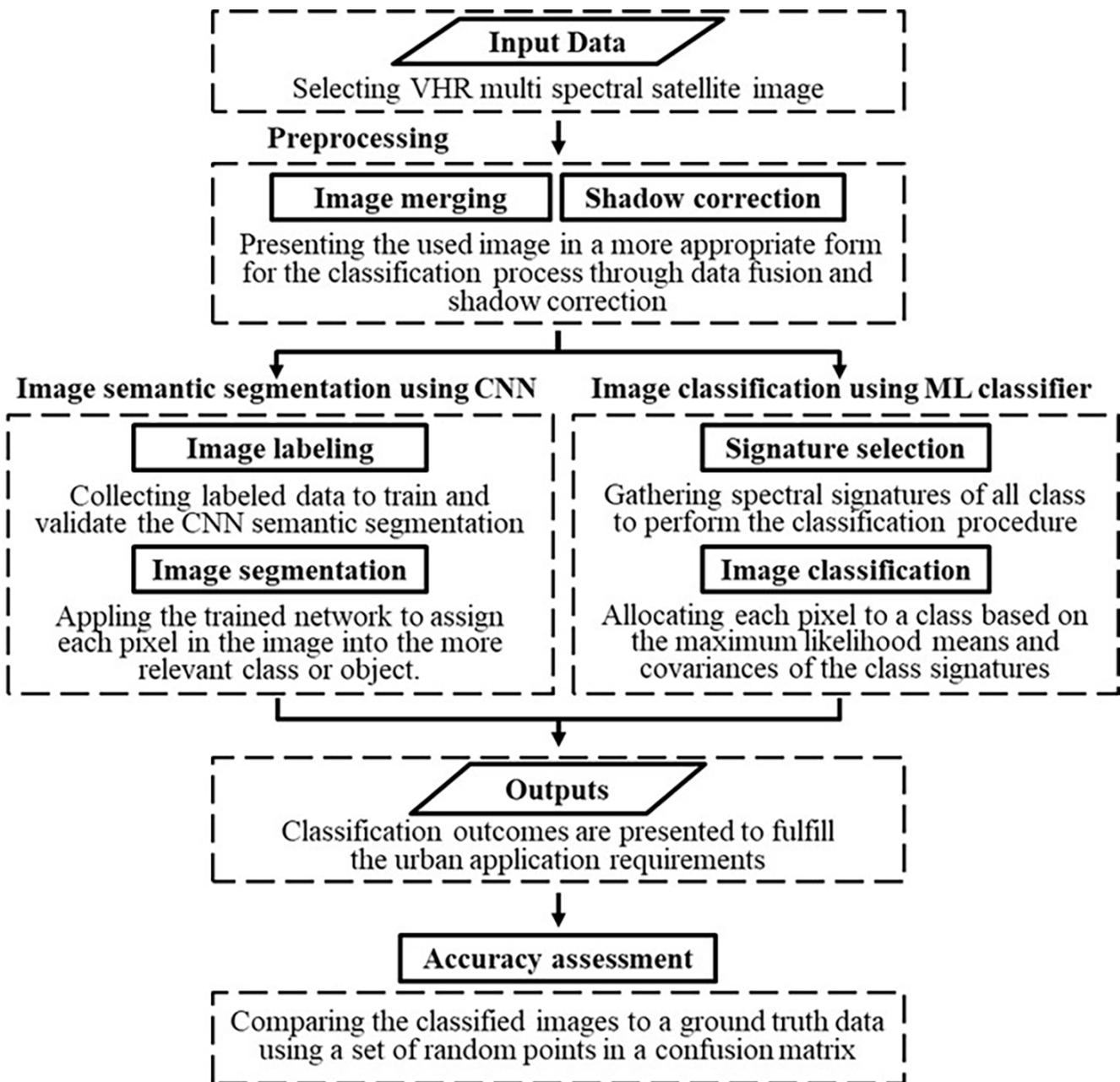


Fig. 1 Flow chart of the proposed procedures

image for land use land cover applications. Several image fusion techniques, such as Principal Component Analysis (PCA), Intensity Hue Saturation (IHS), and Brovey Transform (BT), have been developed to enhance the spatial resolution of the MS images by incorporating with the high spatial resolution of the pan image [17]. The PCA technique is the most effective method for fusing the eight bands of the WorldView-2 image as it maintains the same number of bands before and after the fusion process [18].

3.3.2 Shadow correction

Shadows deteriorate the quality of the accessible information in the VHR satellite images, posing challenges for the

optimal use for urban applications [19]. Despite the low reflectance collected in shadowed regions, valuable information can still be extracted through shadow restoration techniques [20]. Shadow correction process is applied to enhance the satellite image effectiveness by detecting and compensating the shadow areas. Shadow detection focuses on identifying and extracting the shaded pixels from the VHR image without combining with dark or water pixels. Shadow pixels are isolated on the image histogram of the convenient index like Optimized Shadow Index (OSI) [21], Shadow Detector Index (SDI) [22], or Saturation Intensity Shadow Detection Index (SISDI) [19]. After applying the adequate index, a suitable threshold is required to isolate

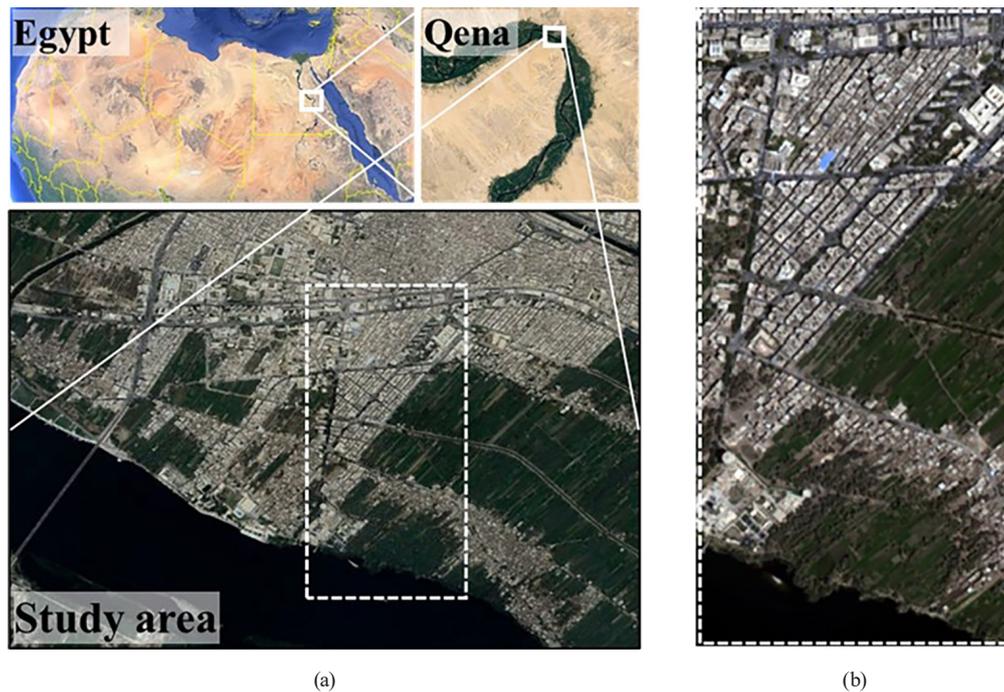


Fig. 2 (a) A satellite view shows the location of the study area in Qena city-Egypt, (b) WorldView-2 image (RGB view)

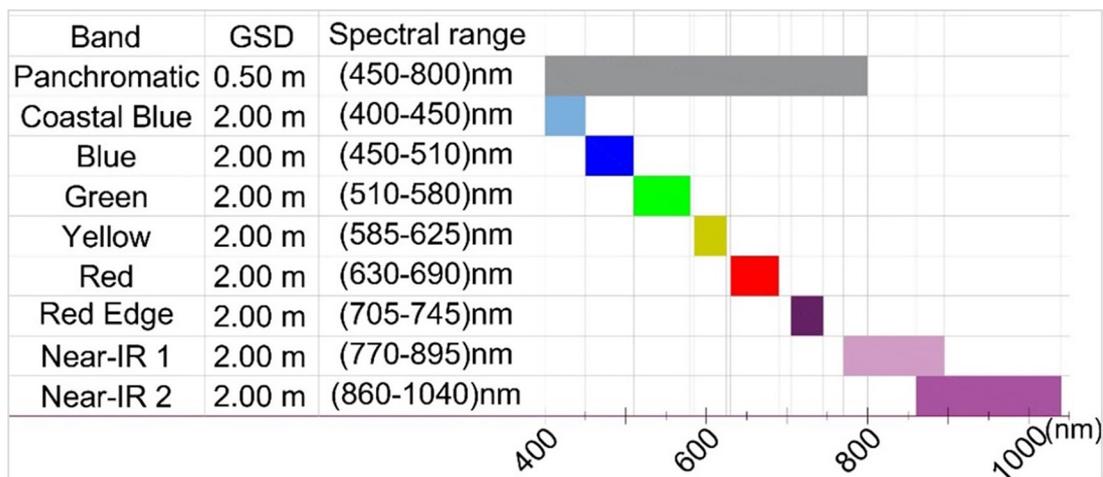


Fig. 3 WorldView-2 band specifications [14]

the shadow values on the resulting index histogram. The OSI is considered to detect the shadow areas in the given image using X-threshold, which separates shadow pixels from non-shadow ones, including dark water pixels. Shadow compensation seeks to adjust the brightness disparities between shadow and non-shadow areas using spectral information of the nearby pixels through mathematical models. The Linear Correlation Correction (LCC) function is derived from multiple pairs of pixel values in both shadow and non-shadow conditions [23]. LCC provides meaningful information for the corrected pixels despite the poor signals in the shadowed areas.

3.4 Image classification

Numerous image classification models are developed to detect and categorize different kinds of land cover classes for urban applications [4]. Effective classification of VHR satellite imagery can be achieved through a variety strategies including AI-based techniques, like convolutional neural networks, as well as standard approaches, like pixel and object-based classifiers [24]. This study presents applying and evaluating the convolutional neural network and the maximum likelihood approaches for VHR image classification in urban areas.

3.4.1 Image semantic segmentation using U-Net model

Urban land cover classification achieves high accuracy using deep learning techniques like convolutional neural networks [24]. CNN models are well-suited for identifying the complex patterns and fine features in VHR satellite data to meet the needs of urban applications. CNN are appropriate and adaptable for pixel-by-pixel classification in complex image segmentation tasks [25]. Classification strategies based on CNN and VHR satellite data objective the identification and categorization of various urban land cover features like buildings, roads, bare soil, vegetation, and water. Image segmentation using CNN labels each pixel to efficiently divide images into convenient categories with more detailed information than just the object-based identification or pixel-based classification. The CNN model is trained and validated using predefined labeled data to segment images accurately into meaningful classes [26, 27]. A transfer learning U-Net is proposed to create a VHR multi-spectral image segmentation model. The developed model workflow (Fig. 4) considers:

1. reading and viewing the VHR image,
2. labeling the image by selecting the appropriate samples for each class corresponding to the real world land cover,
3. using the transfer learning strategy to build, train, validate, and test the U-Net, and
4. applying the trained network to predict the class for new inputs during the testing phase or for real scenarios [28].

Image labeling

VHR satellite image analysis tasks, involving image semantic segmentation, image classification, and LULC mapping, perform better when more useful information is supplied [29]. Since deep learning relies on possessing a large amount of training data, the procedure for developing

a CNN model for semantic segmentation necessitates pixel-by-pixel labeling of the training data sets. Using the relevant labeled data, CNNs are trained, tested, or verified for image semantic segmentation. Images can be manually or automatically labeled using the appropriate algorithm. Accurate labeling is a crucial step of the entire semantic segmentation process to enhance the modelling performance [30]. The proposed model uses MATLAB Image Labeler to create the training and validation datasets [31].

Network architecture

The used U-Net architecture follows the typical structure of convolutional neural networks consisting of three main sections: an encoder, a mid-layer, and a decoder:

- *The encoder section* has four blocks each with a 3×3 convolutional layer with a rectified linear unit (ReLU) activation layer, then again, a 3×3 convolutional layer with a ReLU layer and finally a 2×2 maxpooling layer. The last encoder block has additionally a dropout layer with 50% probability value before the maxpooling. The encoder sizes are 64, 128, 256 and 512 respectively.
- *The mid-layer section* is built of a 3×3 convolution layer with a corresponding ReLU layer, then a second 3×3 convolution layer with ReLU layer, and finally a 50% dropout layer. The sizes are here 512 and 1024.
- *The decoder block* has the same number of decoders as encoders, i.e. four decoders are taken. Each decoder has a 2×2 upsampling convolutional layer with a ReLU layer, a concatenation point, a 3×3 convolution layer with a ReLU layer, then a second 3×3 convolution layer with a ReLU layer. The block sizes are 1024, 512, 256 and 128 respectively. After the last decoder block, $64 \times 1 \times 1$ final convolution layers and $5 \times 1 \times 1$ softmax layers are connected.

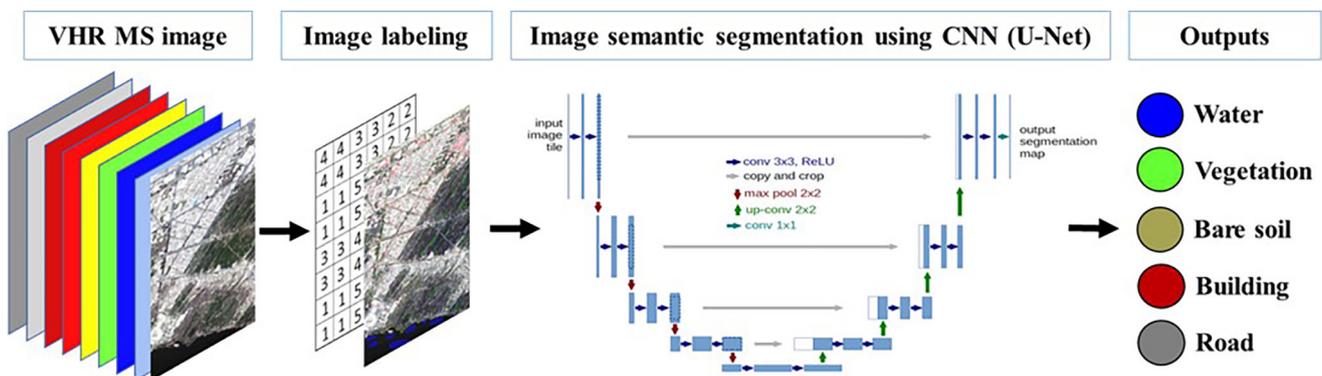


Fig. 4 VHR image classification model using convolutional neural network

- *Finally, the model incorporates cross-connections* where the first encoder connected to the fourth decoder, the second encoder to the third decoder, the third encoder to the second decoder and the fourth encoder to the first decoder blocks (Fig. 5) [32]. The current semantic segmentation network considers the properties of incoming 8-band multispectral image, as well as the number (5) and type of resulting classes. The weights and biases for the convolutional layers were initially set using a random number generator that produces a normally distributed random number.

Network training

Convolutional neural network training on large amount of data and different forms of features consumes high memory usage and computational power. Consequently, the effective training of CNN model on devices with limited resources seems to be a significant challenge. The neural network model is trained using raw data, and the cross-entropy loss function is calculated by contrasting the predictions with the corresponding real labels. All parameter and weights changes are adjusted during the training procedure layer by layer. The training process was conducted using the Stochastic Gradient Descent with Momentum (SGDM) optimizer. SGDM is one of the most popular optimization algorithms which helps accelerate gradients vectors in the correct directions facilitating faster convergence during training. The learning rate and momentum were set to 0.05 and 0.9 respectively. The training was extended by a validation step. The training deals to adjust the model parameters, while the validation

provides an unbiased evaluation of the model during the training phase. For training and validation, two data sets were selected comprising 128×128 pixel sized patches. The training was performed using 16000 patches, and the validation was computed based on 2000 patches. The complete computation was organized into minibatches with the size of 16. A minibatch is a subset of the training set to evaluate the gradient of the loss function and update the weights. The maximum number of epochs (full passes of the data) for training was specified as 150. Adding a regularization term for the weights to the loss function is one way to reduce overfitting. The regularization term is also called weight decay. The L2 regularization technique was used with a factor of 0.0001. To accelerate the training, the algorithms have run on a GPU. The segmentation network parameter adjustment process was monitored by a training and validation progress monitor chart (Fig. 6). It clearly shows the training accuracy and loss progress for the CNN model with final validation accuracy of 99.00%.

Image classification using maximum likelihood

Maximum likelihood is one of the most effective pixel-based techniques to classify each pixel in the image according to the highest probability (likelihood) that a pixel belongs. The ML classifier focuses on the optical values of individual pixels in the VHR satellite images to precisely classify different land cover types in urban scenes based on their spectral signatures. Based on the likelihood for each class, each pixel is assigned to the class with the highest probability of membership [33]. This technique is advantageous for classifying VHR multi-spectral images with detailed information. ML method has the capacity to

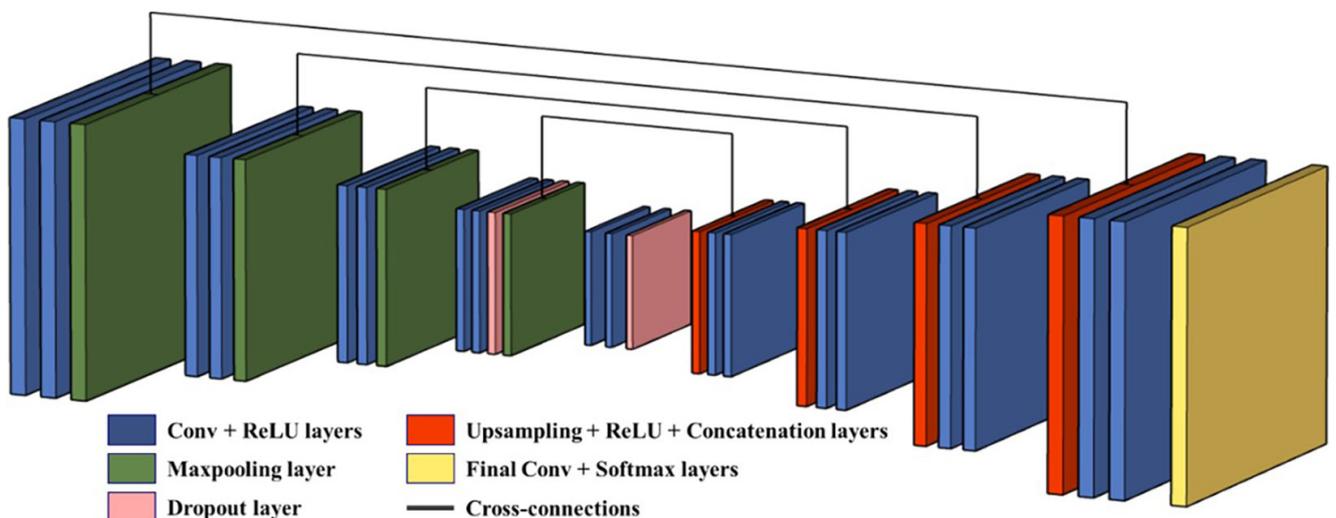


Fig. 5 The applied U-Net architecture

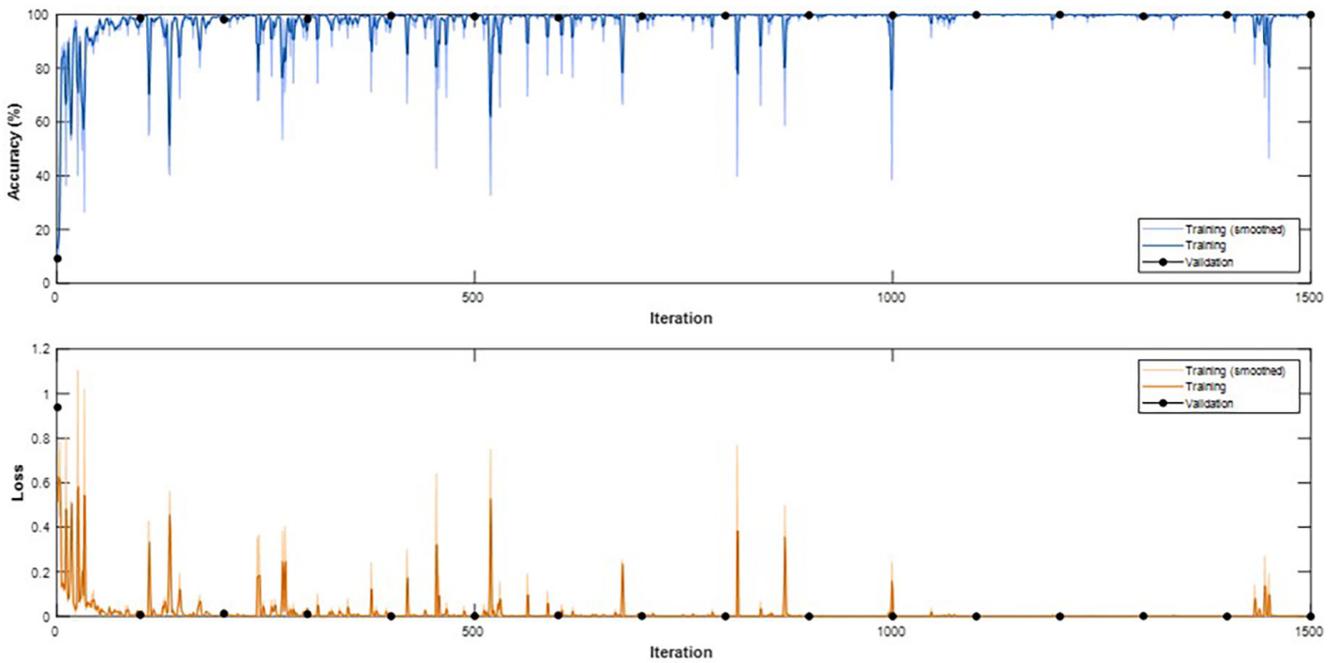


Fig. 6 Training accuracy and loss progress for the CNN model

handle multi-dimensional data which demonstrates consistent performance in the challenging multi-label image classification tasks [34]. A maximum likelihood classification model is presented to extract urban features from VHR satellite images. The model (Fig. 7) utilizes a multispectral VHR satellite image as input, selects effective and sufficient signatures for each class, and applies a maximum likelihood strategy to assign each pixel in the image to a target class based on the means and covariances of the collected signatures [35].

4 Results and discussions

4.1 Classification outcomes

The CNN semantic segmentation process is applied using the collected image labels (Fig. 8 (a)). Results are qualitatively evaluated by comparing the extracted features (Fig. 8 (b)) with real-world land cover classes in the RGB

image (Fig. 2 (b)) to ascertain the size, location, and errors of the classification process. As a visual assessment, the classified image shows significant detection and separability for vegetation and water classes. Additionally, roads, buildings, and bare soil are clearly extracted; however, some misclassified pixels can be observed within each class. Furthermore, to perform the qualitative evaluation, the CNN outcomes are compared using the confusion matrix with a maximum likelihood classified image (Fig. 8 (c)).

4.2 Accuracy assessment

The confusion matrix displays the probability that each pixel in the classified image (column values) matches the real-world land cover class (row values) [36]. A confusion matrix is used to calculate the overall accuracy (Eq. (1)), user's accuracy, and producer's accuracy (Eqs. (2), (3)), and Kappa coefficient.

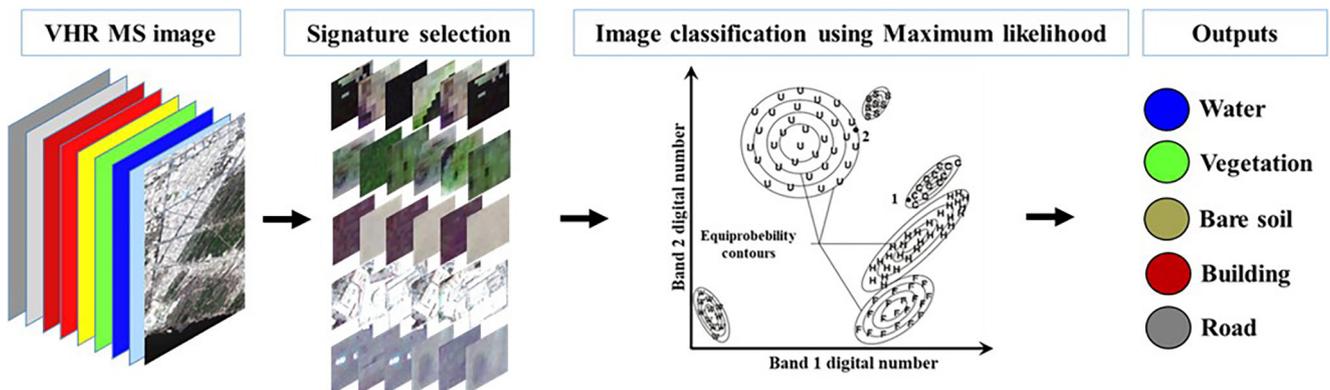


Fig. 7 VHR image classification model using maximum likelihood

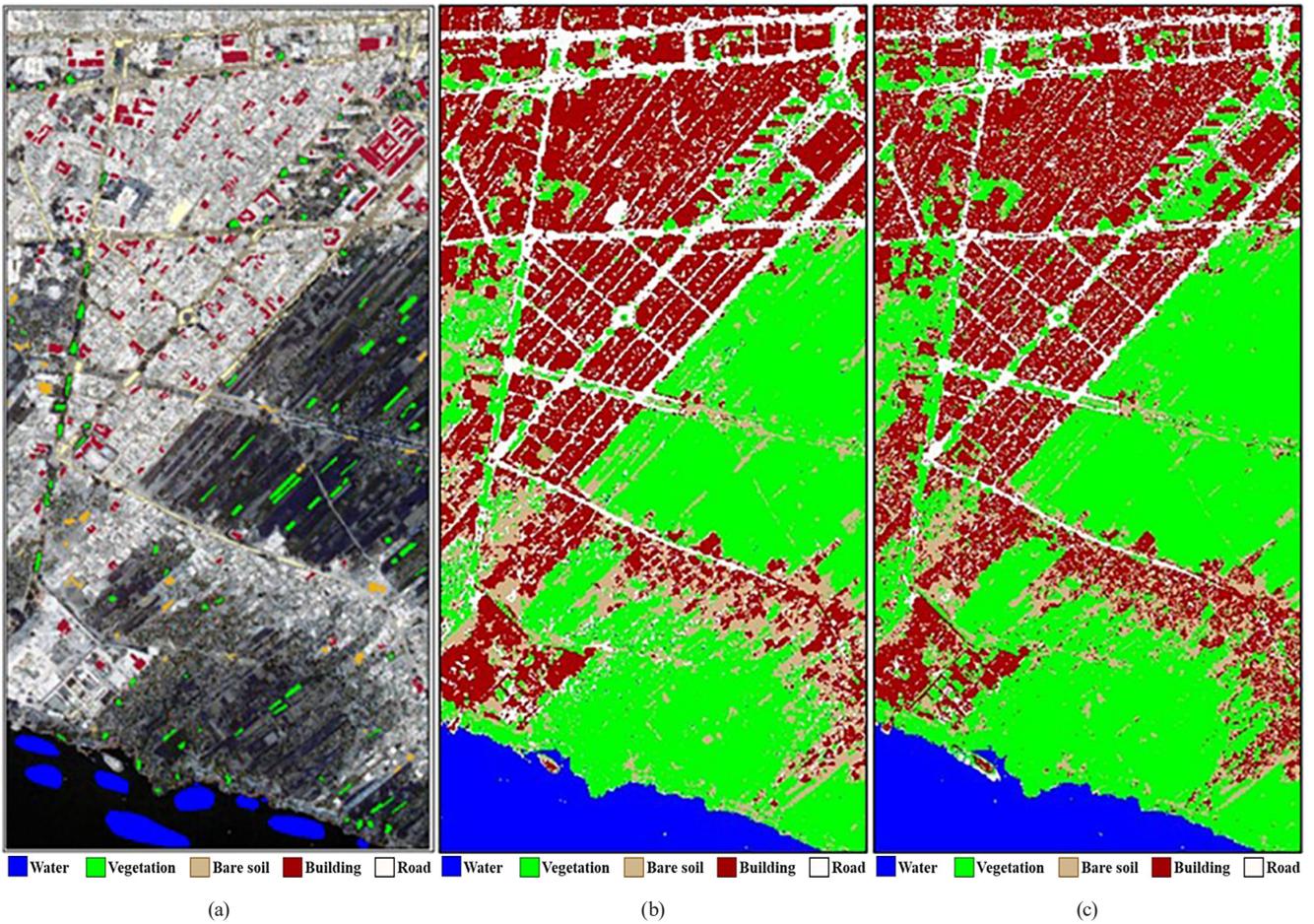


Fig. 8 (a) Image labels, (b) CNN and (c) ML classification results

$$\text{Overall Accuracy: } \tau = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Producer's Accuracies: } \eta = \frac{TP}{TP + FN} \quad (2)$$

$$\text{User's Accuracies: } p = \frac{TP}{TP + FP} \quad (3)$$

Where True Positive (TP) indicates the number of class pixels accurately match the real-world ones, False Positive (FP) refers to the number of non-class pixels that are detected as class ones, True Negative (TN) represents the number of non-class pixels which are classified correctly, and False Negative (FN) shows how many class pixels that are misclassified as non-class ones [37]. The confusion matrix is generated for both the CNN and ML classified images using 400 random points distributed across all classes with a minimum of 50 reference points per class (Tables 1 and 2) [38]. points were assigned to each class based on its relevance to the application's requirements and its representation within the entire scene.

The proposed convolutional neural network model for image semantic segmentation achieved an overall

Table 1 Confusion matrix for image semantic segmentation using CNN

Classified data	Water	Vegetation	Bare soil	Buildings	Roads	Total row	Producer's accuracy	User's accuracy
Water	66	0	0	0	0	66	97.06%	100.00%
Vegetation	1	113	5	0	0	119	86.92%	94.96%
Bare soil	1	6	38	2	0	47	76.00%	80.85%
Buildings	0	0	4	83	4	91	84.69%	91.21%
Roads	0	11	3	13	50	77	92.59%	64.94%
Total column	68	130	50	98	54	400	–	–

Overall accuracy **87.50%** Overall Kappa statistics **0.8395**

accuracy of 87.50% and Kappa coefficient of 0.8395 outperforming the maximum likelihood classification findings with an overall accuracy of 83.50% and Kappa coefficient of 0.7812 (Fig. 9). The CNN model presents high stability and promising performance for producer's and user's accuracy of each class (Fig. 10). Notably, accurate classifications are observed in the water, vegetation, and building classes; meanwhile, due to signature and color

Table 2 Confusion matrix for image classification using maximum likelihood

Classified data	Water	Vegetation	Bare soil	Buildings	Roads	Total row	Producer's accuracy	User's accuracy
Water	64	0	0	0	0	64	94.12%	100.00%
Vegetation	2	120	15	1	2	140	92.31%	85.71%
Bare soil	0	4	29	1	0	34	58.00%	85.29%
Buildings	0	2	6	81	13	102	82.65%	79.41%
Roads	2	4	0	15	39	60	72.22%	65.00%
Total column	68	130	50	98	54	400	–	–

Overall accuracy **83.25%** Overall Kappa statistics **0.7812**

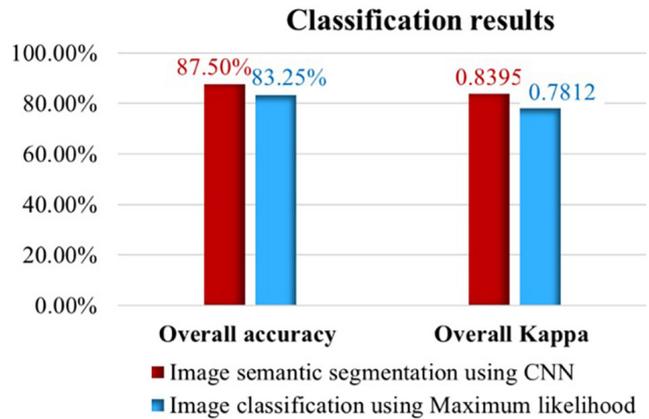


Fig. 9 CNN and ML overall accuracy

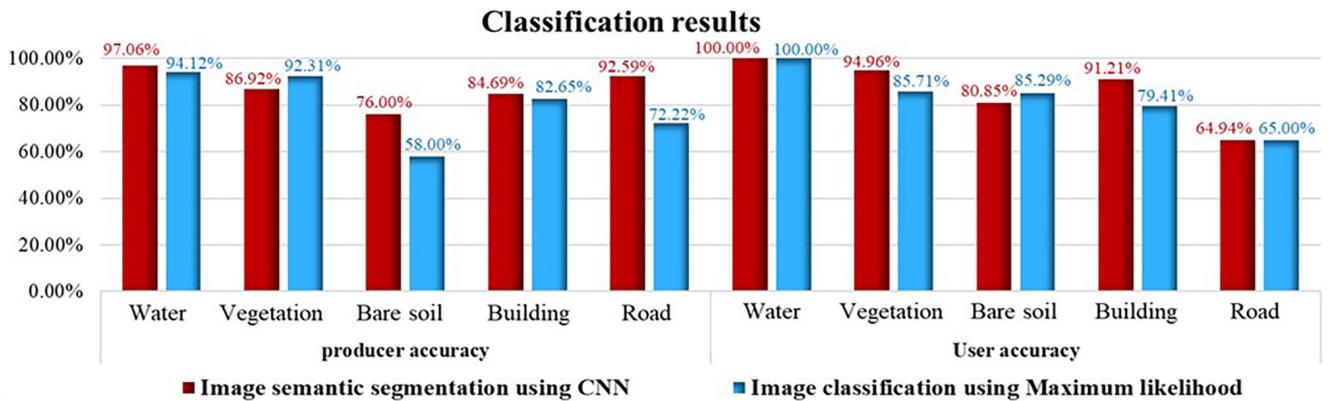


Fig. 10 CNN and ML producer's and user's accuracy

similarities, some misclassified patches are noticeable in the bare soil and road areas. CNNs use both spatial and spectral features to minimize the misclassifications between the road and bare soil classes which are more challenging for the maximum likelihood classifier due to their similar spectral characteristics.

5 Conclusions

The current work explores the effectiveness of convolutional neural networks for VHR data analysis in urban areas focusing on the application of a U-Net model for image semantic segmentation. The proposed U-Net model has demonstrated encouraging results in terms of optimizing the feature extraction, object detection, and LULC analysis for urban uses. The U-Net has a considerable capability to produce accurate classification outcomes for Earth observation applications; However, the network

is constrained by a wide variety of network parameters. The achieved accuracy of the U-Net semantic segmentation model surpasses the maximum likelihood result mainly due to integrating both spatial and spectral features during the classification process.

Acknowledgements

The research reported in this paper is part of project no. BME-NVA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development, and Innovation Fund, financed under the TKP2021 funding scheme.

Civil Engineering Department, Faculty of Engineering, South Valley University, Qena, Egypt is gratefully acknowledged for providing the WorldView-2 satellite image of Qena City.

References

- [1] Esch, T., Taubenböck, H., Heldens, W., Thiel, M., Wurm, M., Geiß, C., Dech, S. "Urban remote sensing – How can earth observation support the sustainable development of urban environments?", In: Schrenk, M., Popovich, V., Zeile, P. (eds.) Real CORP 2010 Proceedings, Wien, Österreich, 2010, pp. 837–847. [online] Available at: <https://elib.dlr.de/64172/> [Accessed: 20 July 2024]
- [2] Yu, M., Yang, C., Li, Y. "Big data in natural disaster management: A review", *Geosciences*, 8(5), 165, 2018. <https://doi.org/10.3390/geosciences8050165>
- [3] Fawzy, M. "Urban Feature Extraction from High Resolution Satellite Images", MSc Thesis, South Valley University, 2020. <https://doi.org/10.13140/RG.2.2.11702.93760/1>
- [4] Vasavi, S., Sri Somagani, H., Sai, Y. "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), pp. 937–953, 2023. <https://doi.org/10.1016/j.ejrs.2023.11.008>
- [5] Ouma, Y. O., Keitsile, A., Nkwae, B., Odirile, P., Moalafhi, D., Qi, J. "Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach", *European Journal of Remote Sensing*, 56(1), 2173659, 2023. <https://doi.org/10.1080/22797254.2023.2173659>
- [6] Deshpande, A., Kumar, M. "Artificial intelligence for big data: complete guide to automating big data solutions using artificial intelligence techniques", [e-book] Packt Publishing Ltd., 2018. ISBN 978-1-78847-217-3 [online] Available at: https://books.google.hu/books/about/Artificial_Intelligence_for_Big_Data.html?id=MpbhtAEACA&redir_esc=y [Accessed: 20 July 2024]
- [7] Tian, T., Li, C., Xu, J., Ma, J. "Urban area detection in very high resolution remote sensing images using deep convolutional neural networks", *Sensors*, 18(3), 904, 2018. <https://doi.org/10.3390/s18030904>
- [8] Fawzy, M., Szabó, G., Barsi, A. "A Shallow Neural Network Model for Urban Land Cover Classification Using VHR Satellite Image Features", *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-1/W1-2023, pp. 57–64, 2023. <https://doi.org/10.5194/isprs-annals-X-1-W1-2023-57-2023>
- [9] Kadhim, M. A., Abed, M. H. "Convolutional neural network for satellite image classification", In: *Intelligent Information and Database Systems: Recent Developments (ACIIDS 2019)*, Yogyakarta, Indonesia, 2020, pp. 165–178. https://doi.org/10.1007/978-3-030-14132-5_13
- [10] Arulananth, T. S., Kuppusamy, P. G., Ayyasamy, R. K., Alhashmi, S. M., Mahalakshmi, M., Vasanth, K., Chinnasamy, P. "Semantic segmentation of urban environments: Leveraging U-Net deep learning model for cityscape image analysis", *Plos One*, 19(4), e0300767, 2024. <https://doi.org/10.1371/journal.pone.0300767>
- [11] Gonzalez, D., Rueda-Plata, D., Acevedo, A. B., Duque, J. C., Ramos-Pollán, R., Betancourt, A., García, S. "Automatic detection of building typology using deep learning methods on street level images", *Building and Environment*, 177, 106805, 2020. <https://doi.org/10.1016/j.buildenv.2020.106805>
- [12] Ronneberger, O., Fischer, P., Brox, T. "U-net: Convolutional networks for biomedical image segmentation", In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Munich, Germany, 2015, pp. 234–241. ISBN 978-3-319-24573-7 https://doi.org/10.1007/978-3-319-24574-4_28
- [13] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., Farhan, L. "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions", *Journal of Big Data*, 8, 53, 2021. <https://doi.org/10.1186/s40537-021-00444-8>
- [14] The European Space Agency "WorldView-2 Instruments", [online] Available at: <https://earth.esa.int/eogateway/missions/worldview-2> [Accessed: 20 July 2024]
- [15] Fawzy, M., Mostafa, Y. G., Khodary, F. "Automatic Indices Based Classification Method for Map Updating Using VHR Satellite Images", *JES. Journal of Engineering Sciences*, 48(5), pp. 845–868, 2020. <https://doi.org/10.21608/jesaun.2020.114602>
- [16] Mishra, R. K., Zhang, Y. "A comparison of commercial Pan-sharpening techniques for HR Satellite imagery", [pdf] presented at 2013 Esri International User Conference, San Diego, CA, USA, Jul., 8–12, 2013. Available at: https://proceedings.esri.com/library/userconf/proc13/papers/470_149.pdf [Accessed: 20 July 2024]
- [17] Pushparaj, J., Hegde, A. V. "Evaluation of pan-sharpening methods for spatial and spectral quality", *Applied Geomatics*, 9(1), pp. 1–12, 2017. <https://doi.org/10.1007/s12518-016-0179-2>
- [18] Pohl, C., van Genderen, J. "Remote sensing image fusion: A practical guide", CRC Press, 2016. ISBN 978-1-4987-3002-0 <https://doi.org/10.1201/9781315370101>
- [19] Mostafa, Y., Yousef, M., Mostafa, F. "A new shadow detection index for 8-band very high-resolution satellite images", *International Journal of Remote Sensing*, 41(2), pp. 420–432, 2020. <https://doi.org/10.1080/01431161.2019.1641760>
- [20] Mostafa, Y., Abdelwahab, M. A. "Corresponding regions for shadow restoration in satellite high-resolution images", *International Journal of Remote Sensing*, 39(20), pp. 7014–7028, 2018. <https://doi.org/10.1080/01431161.2018.1471541>
- [21] Mostafa, Y., Abdelhafiz, A. "Shadow identification in high resolution satellite images in the presence of water regions", *Photogrammetric Engineering & Remote Sensing*, 83(2), pp. 87–94, 2017. <https://doi.org/10.14358/PERS.83.2.87>
- [22] Mostafa, Y., Abdelhafiz, A. "Accurate shadow detection from high-resolution satellite images", *IEEE Geoscience and Remote Sensing Letters*, 14(4), pp. 494–498, 2017. <https://doi.org/10.1109/LGRS.2017.2650996>
- [23] Sarabandi, P., Yamazaki, F., Matsuoka, M., Kiremidjian, A. "Shadow detection and radiometric restoration in satellite high resolution images", In: *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, Vol. 6, Anchorage, AK, USA, 2004, pp. 3744–3747. ISBN 0-7803-8742-2 <https://doi.org/10.1109/IGARSS.2004.1369936>

- [24] Mboga, N., Georganos, S., Grippa, T., Lennert, M., Vanhuyse, S., Wolff, E. "Fully convolutional networks and geographic object-based image analysis for the classification of VHR imagery", *Remote Sensing*, 11(5), 597, 2019.
<https://doi.org/10.3390/rs11050597>
- [25] Song, J., Gao, S., Zhu, Y., Ma, C. "A survey of remote sensing image classification based on CNNs", *Big Earth Data*, 3(3), pp. 232–254, 2019.
<https://doi.org/10.1080/20964471.2019.1657720>
- [26] Yi, Y., Zhang, Z., Zhang, W., Zhang, C., Li, W., Zhao, T. "Semantic segmentation of urban buildings from VHR remote sensing imagery using a deep convolutional neural network", *Remote Sensing*, 11(15), 1774, 2019.
<https://doi.org/10.3390/rs11151774>
- [27] Neupane, B., Horanont, T., Aryal, J. "Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis", *Remote Sensing*, 13(4), 808, 2021.
<https://doi.org/10.3390/rs13040808>
- [28] Lu, J., Fang, C., Xu, M., Lin, J., Wang, Z. "Evaluations on deep neural networks training using posit number system", *IEEE Transactions on Computers*, 70(2), pp. 174–187, 2021.
<https://doi.org/10.1109/TC.2020.2985971>
- [29] Xu, Z., Su, C., Zhang, X. "A semantic segmentation method with category boundary for Land Use and Land Cover (LULC) mapping of Very-High Resolution (VHR) remote sensing image", *International Journal of Remote Sensing*, 42(8), pp. 3146–3165, 2021.
<https://doi.org/10.1080/01431161.2020.1871100>
- [30] Niu, B. "Semantic segmentation of remote sensing image based on convolutional neural network and mask generation", *Mathematical Problems in Engineering*, 2021, 2472726, 2021.
<https://doi.org/10.1155/2021/2472726>
- [31] MathWorks, Inc. "Get started with the Image Labeler", [online] Available at: <https://nl.mathworks.com/help/vision/ug/get-started-with-the-image-labeler.html> [Accessed: 20 July 2024]
- [32] Pan, Z., Xu, J., Guo, Y., Hu, Y., Wang, G. "Deep learning segmentation and classification for urban village using a worldview satellite image based on U-Net", *Remote Sensing*, 12(10), 1574, 2020.
<https://doi.org/10.3390/rs12101574>
- [33] Saboori, M., Homayouni, S., Shah-Hosseini, R., Zhang, Y. "Optimum feature and classifier selection for accurate urban land use/cover mapping from very high resolution satellite imagery", *Remote Sensing*, 14(9), 2097, 2022.
<https://doi.org/10.3390/rs14092097>
- [34] Balha, A., Singh, C. K. "Comparison of maximum likelihood, neural networks, and random forests algorithms in classifying urban landscape", In: Singh, V. P., Yadav, S., Yadav, K. K., Corzo Perez, G. A., Muñoz-Arriola, F., Yadava, R. N. (eds.) *Application of Remote Sensing and GIS in Natural Resources and Built Infrastructure Management*, Springer, 2023, pp. 29–38. ISBN 978-3-031-14095-2
https://doi.org/10.1007/978-3-031-14096-9_2
- [35] Topaloğlu, R. H., Sertel, E., Musaoğlu, N. "Assessment of classification accuracies of Sentinel-2 and Landsat-8 data for land cover/use mapping", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B8, pp. 1055–1059, 2016.
<https://doi.org/10.5194/isprs-archives-XLI-B8-1055-2016>
- [36] Story, M., Congalton, R. G. "Accuracy assessment: A user's perspective", [pdf] *Photogrammetric Engineering and Remote Sensing*, 52(3), pp. 397–399, 1986. Available at: https://www.asprs.org/wp-content/uploads/pers/1986journal/mar/1986_mar_397-399.pdf [Accessed: 20 July 2024]
- [37] Kulkarni, A., Chong, D., Batarseh, F. A. "5 - Foundations of data imbalance and solutions for a data democracy", In: *Data Democracy*, Academic Press, 2020, pp. 83–106. ISBN 978-0-12-818366-3
<https://doi.org/10.1016/B978-0-12-818366-3.00005-8>
- [38] Congalton, R. G. "A review of assessing the accuracy of classifications of remotely sensed data", *Remote Sensing of Environment*, 37(1), pp. 35–46, 1991.
[https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)