International Research Journal of Earth Sciences ISSN 2321 – 2527
Vol. **14(2),** 8-17, August **(2026)** *Int. Res. J. Earth Sci.*

Review Paper

Review on Deep Learning Methods, and their Implications for Mapping Geomorphic Landforms

Muhammad Ali[1*] and Hussain Asghar[2]
[1]Department of Geology and Geography, West Virginia University, Morgantown, WV 26505, United States
[2]Institute of Geology University of Azad Jammu and Kashmir, Muzaffarabad AJ&K, Pakistan
geomuhammad00@gmail.com

Available online at: www.isca.in
Received 26th January 2026, revised 12th March 2026, accepted 24th April 2026

Abstract

The importance of accurate mapping of geomorphic and archeological features has arisen across the globe after the great success of deep learning techniques. Deep learning models are novel approaches utilized for the mapping of geomorphic landforms, and archeological features to improve accuracy, reduce time, provide automation, and produce high-resolution maps. Convolutional Neural Networks (CNNs) have proven to be one of the most promising techniques applied effectively for the mapping of geomorphic landforms. Spatial and semantic parameters of the CNNs and U-Net are reviewed in this work, and their implications for the accurate mapping of geomorphic landforms. Additionally, very recently, a series of studies have established that Convolutional Neural Networks (CNN) are very effective in pixel-level classification and aggregated landforms. This review presents a few examples where CNN has shown accurate results. CNN has measured an ice-wedge polygon in Alaska with a 91% accuracy rate, and the CNN used as an object-based image analysis in the northern Netherlands to map low-relief features. Furthermore, this study presents a comprehensive deep insight review into CNN and its utilization for accurate mapping of complex and transitional landforms since pixel and object-based techniques do not perform well in classifying complex and transitional landforms.

Keywords: Deep Learning; Geomorphic features; U-Net; Landform Mapping; CNN.

Introduction

Geomorphological maps, which have been created all around the world, contain details on the relief, landforms, and origins of the earth's surface[1-3]. They are used for a variety of regional spatial developments, including archaeological prospecting[4], and natural hazard mapping[5]. One of the most significant methods in geomorphological mapping and comprehending the Earth's surface process is landform classification[6,7]. Based on the advancement of image segmentation techniques, Machine Learning (ML) approaches have emerged as one of the most widely used techniques for the recognition and segmentation of landforms in the field of geomorphological study[8]. Information can be extracted from numerous and large-scale data sources using machine learning techniques[9]. ML techniques can leverage and combine the surface spectra collected from imagery and the morphological features produced from digital elevation models (DEMs) for data training. These methods apply to geomorphological research.

Deep learning (DL) techniques have recently attracted a lot of attention among ML techniques since they are capable of active learning[10,11]. Due to its strong capacity to recognize and extract information, DL algorithms have been frequently utilized to recognize terrain characteristics, or landform recognition[12]. Deep learning, particularly convolutional neural networks (CNNs), is a potential machine-learning technique for geomorphological mapping[13].

CNNs can be used for a variety of tasks, such as speech recognition[13], text analysis[14], image classification[15] and semantic image segmentation[16]. Due to the combination of spatial data generalization and segmentation, CNNs can be seen as a combination of object-based and pixel-based mapping techniques. This promising method is increasingly applied for mapping landforms and geomorphology[17-22].

Since traditional landform classification methods (pixel and object-based) do not perform well in classifying complex and transitional landforms, a typical structure of the deep convolutional network, U-Net is given special consideration in the current investigation for the mapping of geomorphic landforms[20]. Thus, one of the most crucial steps in geomorphological mapping and a deeper understanding of the processes that shape the Earth's surface is landform classification[6,7].

The main goal of the current study is to provide a thorough analysis of U-Net and CNNs in terms of semantic segmentation and pixel-level classification. This study also includes a landform classification using CNNs and U-Net.

Convolutional neural network (CNN)

CNNs are built on the concept of a layer of convolving windows that move along a data array to detect features (e.g., edges) in the data by utilizing different filters[23] (Figure-1). A CNN has some three-dimensional hidden layers, with each layer learning to detect different features of the input images[24]. Each layer can perform one of two types of operations: convolution or pooling. Convolution, activation, and pooling layers are typical neural layers found in CNNs[24,25]. CNNs are deep learning models that can learn to recognize patterns or objects by applying a series of mathematical operations to the input data[26]. Convolutions, max pooling operators, batch normalizations, and rectified linear unit activation functions (ReLU) are among the mathematical operations. By using a mathematical process with two inputs, such as an image matrix and a filter or kernel, the convolution layer learns picture features from small squares of input data while maintaining correlations between individual pixels. Convolution processes the input images through a set of convolutional filters (e.g., a 3×3 size filter), each of which detects and enhances specific features in the images, a convolutional layer's units are organized in feature maps.

Through a set of weights, each unit of the feature map is linked to local patches in the feature maps of the previous layer. A non-linear transfer function is then applied to this locally weighted sum[23]. Pooling operators are techniques for reducing the size of input images by aggregating them locally. Whereas, Batch normalizations are functions that normalize the output of convolutions to make the model faster and more stable[27]. After the convolutional layer, the activation layer is composed of a nonlinear activation function. The rectified linear unit (ReLU) function is the most popular and efficient activation function[28] intuitively, its gradients are always 0 s and 1 s. When photos are too huge, the pooling layer minimizes the number of parameters, while keeping the crucial data.

A fully linked basic multilayer perceptron engages in discriminative learning to recognize an object class in a deep neural network. According to the sorts of data and images, numerous CNN structures have been created, including ZFNet, VGGNet, GoogleNet, ResNet, LeNet-5, and AlexNet[29].

U-Net: U-Net is the typical structure of the deep convolutional network[30] (Figure-2). The network as a whole is shaped like a "U," thus the name U-Net[31]. The U-Net can be divided into two paths: contracting and expanding paths, the contracting path for extracting the appropriate features, and the expanding path for precise[16]. The contracting path's function is to extract useful information from input data and to compress images. The architecture of the expanding path differentiates U-Net from conventional CNN.

The fully connected layers are replaced by up-convolution for up sampling[20]. The contracting path consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for down sampling. The number of feature channels is doubled at each down-sampling step.

Every step in the expanding path consists of an up-sampling of the feature map, followed by a 2x2 convolution ("up convolution") that reduces the number of feature channels by half, a concatenation with the corresponding cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. Cropping is required due to the loss of border pixels in each convolution. A 1x1 convolution is used at the final layer to map each 64-component feature vector to the desired number of classes. The network has 23 convolution layers in total[16].

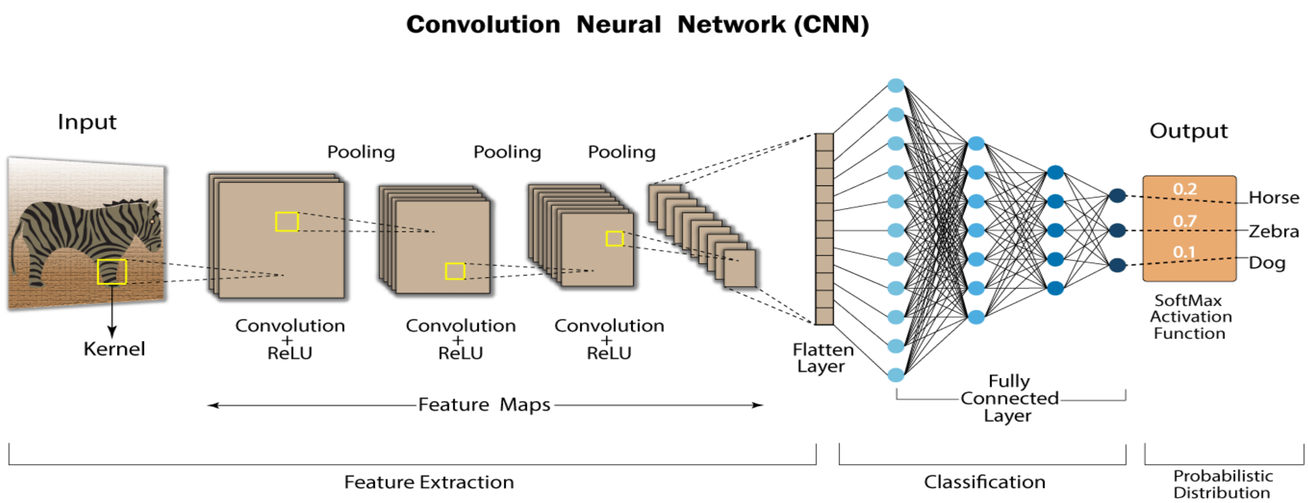

Figure-1: Schematic representation of a convolutional neural network (CNN) architecture, illustrating convolutional, pooling, and fully connected layers[32].

Mask R-CNN: The state-of-the-art in image segmentation is the Convolutional Neural Network (CNN) Mask R-CNN[33] (Figure-3). In terms of structure, the Mask R-CNN mainly consists of (1) backbone architecture Residual Learning network (ResNet)[34] for feature extraction; (2) Feature Pyramid Network (FPN)[35] for improving representation of objects at multiple scales; (3) Region Proposal Network (RPN) for generating RoI; (4) RoI Classifier for class prediction of each RoI; (5) Bounding Box Regress or (BBR) for refining RoI; (6) FCN[36] with RoIAlign[33] and bilinear interpolation for predicting pixel-accurate mask.

Pixel-Level Classification Using CNN and U-Net

A CNN architecture designed for semantic segmentation, also known as pixel-by-pixel classification, aims to get a semantic map in which each pixel in the image is assigned to a class. In the U-Net model, the loss in spatial information is compensated by up-sampling deeper layers in the model and combining them with shallower ones where the spatial information is richer. With the help of this architecture, the model can get a larger receptive field, while still maintaining spatial detail. The combination of a larger receptive field and preservation of spatial information allows for semantic segmentation. Moreover, this combination enables associating each pixel in the input images with a certain class, instead of giving one class to the entire image. Semantic segmentation CNNs, such as U-Net, can thus recognize objects based on local information while considering the larger context[16]. This makes them suitable for geomorphological mapping.

Semantic and Instance segmentation methods

Image Segmentation: Image segmentation partitions a digital image into multiple segments, known as image regions or image objects (sets of pixels) (Figure-4). The goal is to transform the representation of the image into a simpler, more understandable, and more analyzable form[37]. Image segmentation is typically used to identify objects and boundaries (such as lines and curves) within image data. Specifically, it involves assigning a label to every pixel in an image such that pixels with the same label possess certain characteristics. This process isolates objects of interest for further analysis, including classification and object recognition[38]. Numerous techniques exist for segmenting images. Based on the characteristics of the object to be classified, these methods can be divided into two primary categories: semantic segmentation and instance segmentation.

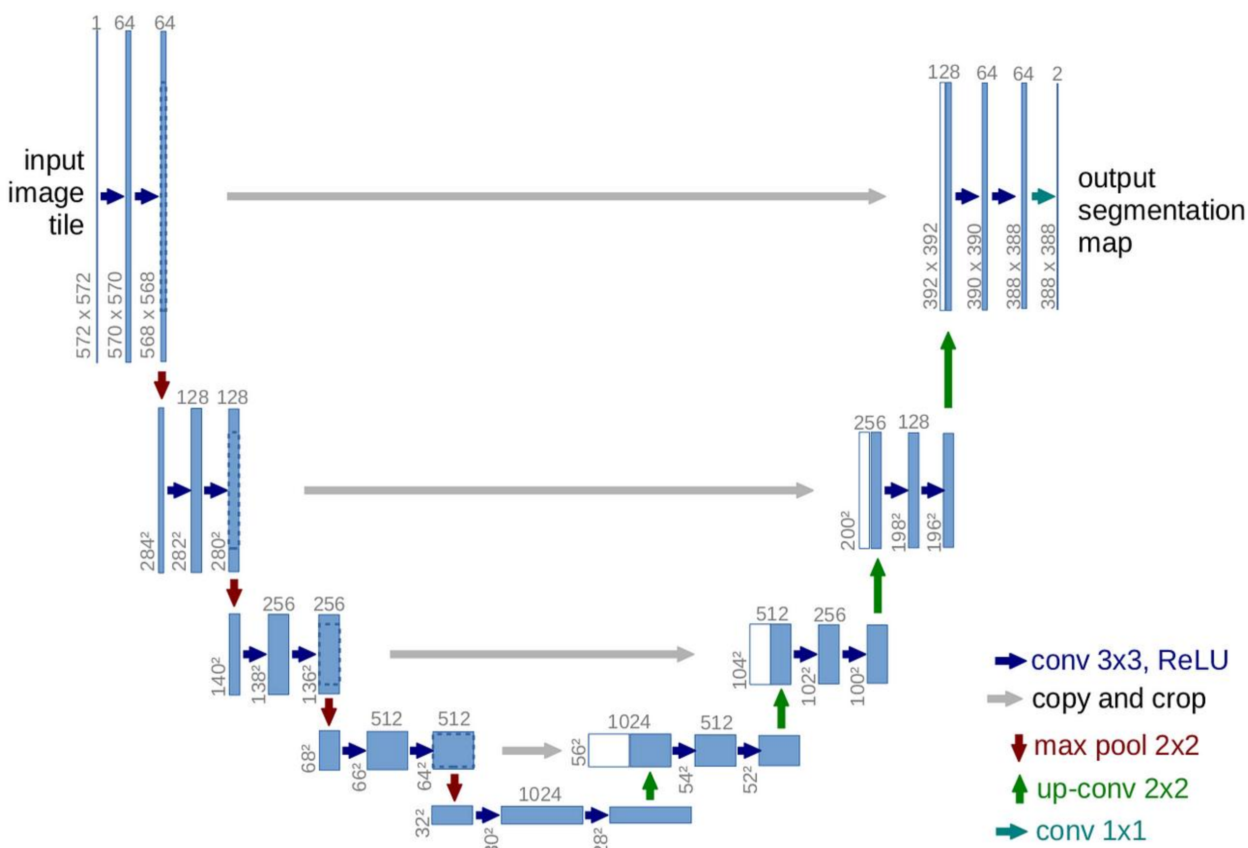

Figure-2: The U-net architecture, illustrated here for a resolution of 32x32 pixels at the lowest level, features several components. Each blue box signifies a multi-channel feature map, with the number of channels specified above the box. The dimensions in the x-y plane are indicated at the lower left edge of each box. White boxes depict copied feature maps, and the various operations are represented by arrows[16].

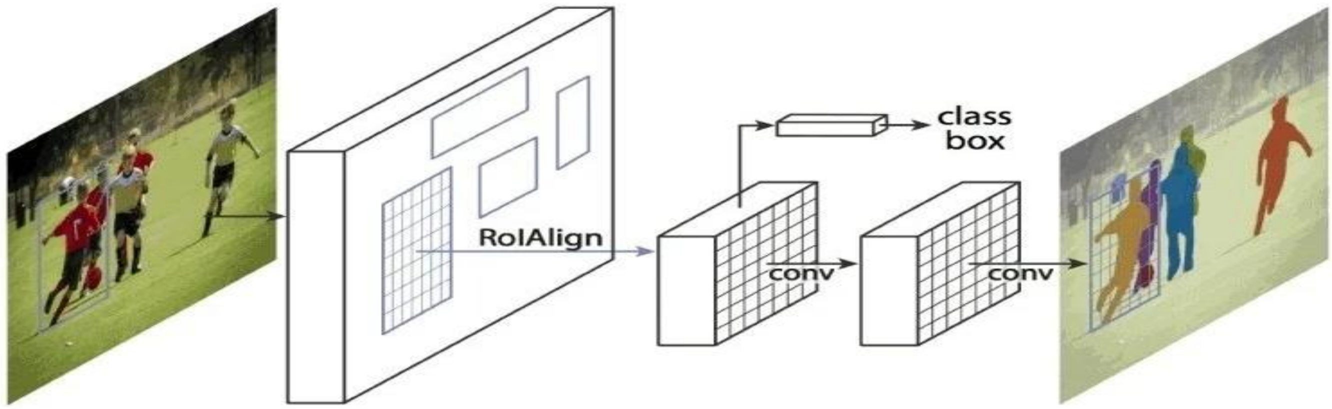

Figure-3: Conceptual framework of the Mask R-CNN architecture for instance segmentation, including feature extraction, region proposal, classification, and mask prediction[33].

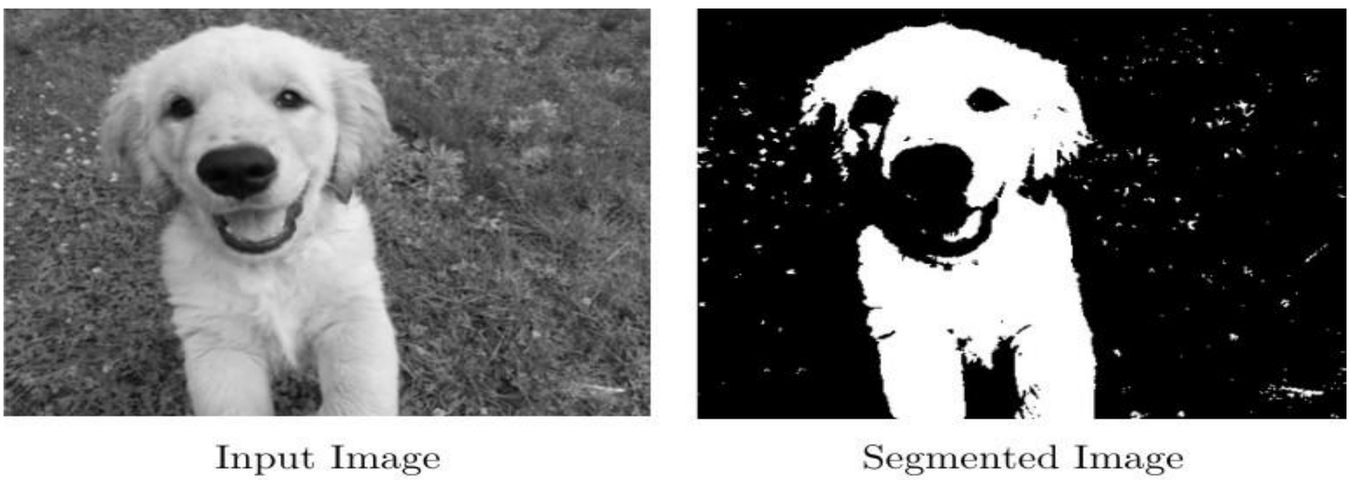

Figure-4: An example of Image Segmentation[39].

Semantic segmentation method: Semantic segmentation is an advanced task that equips computer vision systems with the capability to analyze an image and identify objects or regions by assigning a label to each pixel within the specified object or region[40] (Figure-5). This process entails assigning a label to each pixel within the image, wherein multiple objects belonging to the same category are identified as a single collective entity[41].

Instance segmentation method: Instance segmentation involves the precise delineation of object boundaries at a granular level (Figure- 5). Instance detection, on the other hand, refers to the process of repeatedly identifying the same instance of an object across different contexts or frames[40]. The process involves assigning a unique label to each pixel in the image, wherein multiple objects belonging to the same class are recognized as distinct individual entities (or instances)[41]. To further elucidate the distinctions between semantic segmentation and instance segmentation, an additional example incorporating object detection, semantic segmentation, and instance segmentation is presented in Figure-6.

Deep Learning for Terrace Mapping: The utilization of deep learning techniques, particularly the semantic segmentation approach, has significantly enhanced the ability to map terraces from remote sensing images and digital elevation models (DEMs). As demonstrated in recent studies, U-net architectures have proven to be exceptionally effective in delineating terraces with high precision, critical for accurate geographical analyses and agricultural planning[31].

The application of these techniques, illustrated in the provided image (Figure-7), showcases their capability in interpreting complex spatial relationships and textures of terraces, automating feature extraction that traditionally required extensive manual effort. This integration of deep learning into geomorphological studies marks a significant advancement towards dynamic and precise environmental modeling, especially vital in regions such as the Loess Plateau of China, where terraces are integral for soil and water conservation[31].

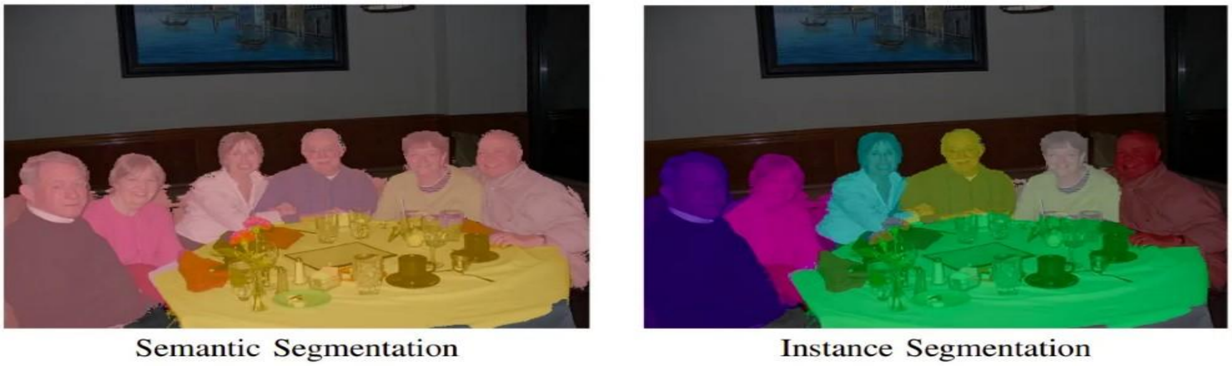

Figure-5: An example of semantic segmentation and instance segmentation from left to right, respectively[32].

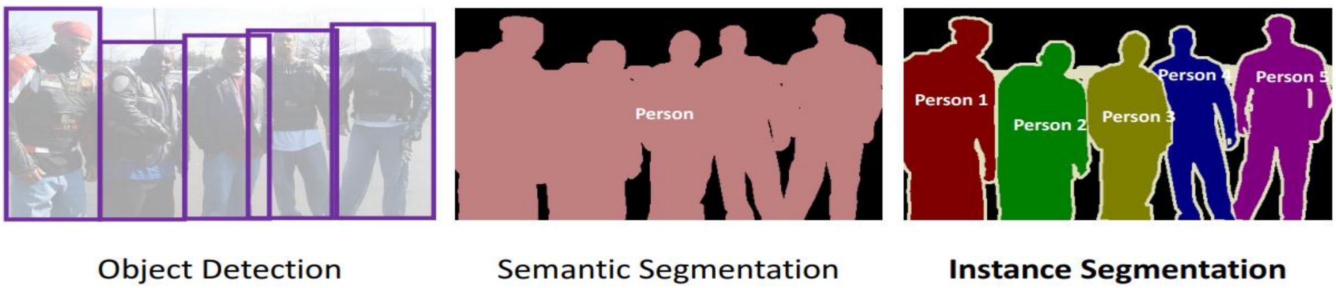

Figure-6: Illustration comparing object detection, semantic segmentation, and instance segmentation approaches in computer vision[42].

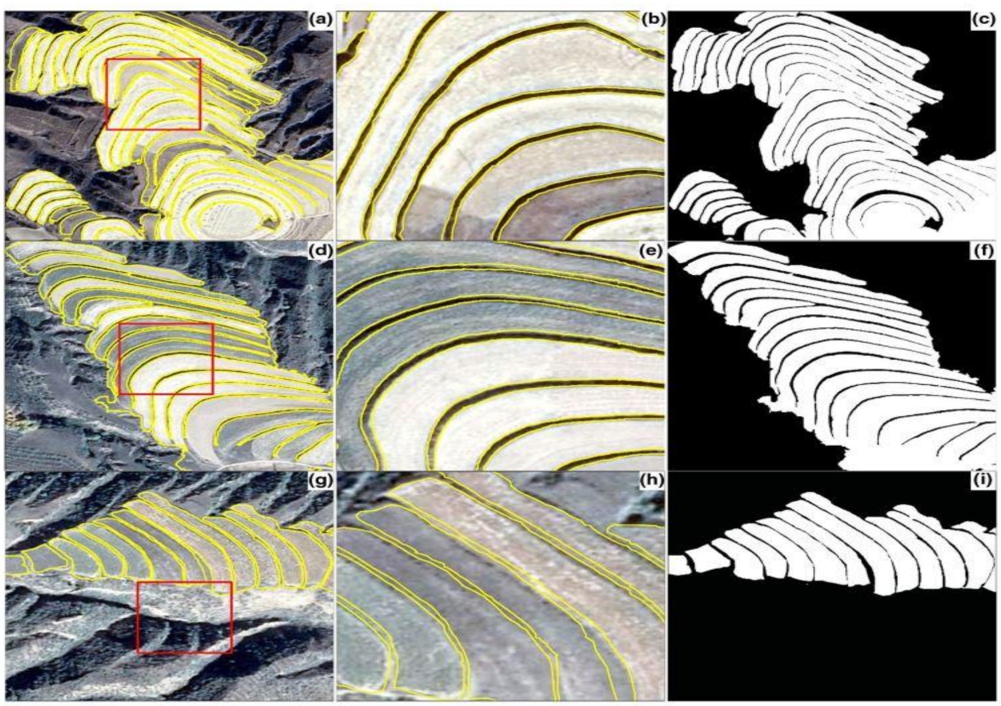

Figure-7: Deep Learning-Based Terrace Extraction Outcomes**:** This figure displays the outcomes of terrace edge and field extractions across three distinct sample areas using a deep learning approach. It encompasses: (a) delineation of terrace edges in Sample Area A; (b) detailed view of edge extractions in Sample Area A; (c) terrace field extraction results in Sample Area A; (d) delineation of terrace edges in Sample Area B; (e) detailed view of edge extractions in Sample Area B; (f) terrace field extraction results in Sample Area B; (g) delineation of terrace edges in Sample Area C; (h) detailed view of edge extractions in Sample Area C; and (i) terrace field extraction results in Sample Area C[31].

CNNs in Semantic and Instance Segmentation

U-Net, a particular CNN architecture, is frequently employed for semantic segmentation tasks. It is designed to classify each pixel of an image into a single semantic class, enabling pixel-level classification. U-Net combines the capabilities of CNNs with an encoder-decoder structure and skip connections, allowing for the preservation of spatial information, while acquiring a more extensive receptive field. The encoder part captures contextual information, while the decoder part recovers spatial details.

This combination makes U-Net an effective model for semantic segmentation because it is capable of object recognition based on local information while taking the larger context into account. U-Net has been effectively used in many different fields, including geomorphological mapping, where it enables accurate pixel-level classification of landforms[9,17,18,20-22].

Mask R-CNN is a Convolutional Neural Network (CNN) that is state-of-the-art in image segmentation[33]. The Mask R-CNN method is an extended method for object instance segmentation that is based on the Faster R-CNN (a fast and effective algorithm for object detection) by including a function for predicting masks for distinct objects[33,43]. The Mask R-CNN algorithm consists of two stages: (1) After scanning an image, the Mask R-CNN generates proposals (i.e., candidate object bounding boxes). (2) For each region of interest, the Mask R-CNN predicts the class, bounding box, and binary mask (RoI)[33].

Convolutional Operations in Segmentation

Convolutions are matrix calculations with moving-window filters (kernels), typically of 3 ×3 cells that are used to summarize local spatial information in the input images. This operation is similar to moving window calculations for determining slope or topographic properties from a DEM. Pooling operators are methods that reduce the extent of the input images by locally aggregating them, by passing on the maximum value in the aggregation of 2 × 2 cells. This pooling clearly state that pooling reduces the spatial dimensions of the feature maps; or receptive field, over which the convolutions can summarize spatial information, but this comes with a loss of spatial information due to aggregation.

Batch normalizations are functions that normalize the output of the convolutions, to make the model faster and more stable. A ReLU is a function that is designed to introduce non-linearity in the calculations of the CNN, by setting the negative outputs of a convolutional calculation to zero, while the positive values remain the same. CNNs need this nonlinearity to be able to fit non-linear functions. Applying multiple linear operations, such as convolutions, one after the other would only allow the CNN to learn a linear function of the input, which would not be suitable to solve most tasks[27]. The spatial dimensions of the input images are reduced due to the pooling operators. This means that spatial information gets lost. Although this may be a desirable behavior for image classification, where the image is assigned to one single class, we would like to preserve the spatial information contained in the input images to classify each pixel into its corresponding geomorphological unit[16].

In pixel-level segmentation tasks, convolutional techniques are essential for obtaining spatial context information. Fully Convolutional Networks (FCNs) are used in the mask-prediction branch of Mask R-CNN to generate dense pixel-wise predictions. These processes provide the network the ability to understand the layout of objects and accurately segment individual instances[33].

Case Study Review: Deep Learning Applications in Geomorphic Mapping of Utrechtse Heuvelrug, Netherlands

The Utrechtse Heuvelrug, located in the central push-moraine district of the Netherlands (Figure-8), presents a unique study area characterized by its distinct relief and the coexistence of large regional and small local landforms. These features are distinguishable both on digital elevation models (DEMs) and in the field, making it an ideal location for applying Convolutional Neural Networks (CNNs) for automated geomorphological mapping. The region's well-documented geomorphological history and subsurface geology provide a solid foundation for deep learning models to thrive, as highlighted in the comprehensive work of Van der Meij et al.[27] who build upon the findings of Jongmans et al[44], Stouthamer et al.[45] and Van den Berg and Beets[46].

In their research, Van der Meij and colleagues[27] utilized deep learning techniques, particularly CNNs, to classify and map the diverse geomorphological units within the area. These ranged from Saalian-era push moraines to Weichselian incised valleys and covers and deposits[46]. Leveraging high-resolution DEMs and extensive historical coring data, the models adeptly differentiated natural geomorphological units from anthropogenic structures, which were selectively included based on their landscape impact[47]. This case study underscores the potential of deep learning to enhance the accuracy and efficiency of geomorphological mapping, with significant implications for environmental management and archaeological research. Van der Meij's work addresses challenges such as model validation and the handling of mixed spectral signals from varied terrain types, offering insights that are detailed in their publication's supplementary materials, including Figure-8.

Figure-8 illustrates the geomorphological mapping of the Utrechtse Heuvelrug area. Map A displays the training region utilized to train the Convolutional Neural Networks (CNNs). Map B details the test area where both manual and computational mapping approaches were developed and assessed. The maps presented are based on the 2019 geomorphological survey and are adapted from van der Meij et al.[27].

Diffusion Problem: Ambiguity in Aggregated Landforms Boundaries

Comparative Analysis of Observed and Predicted Geomorphic Units: Figure-9 provides a visual comparison between the observed (left) and predicted (right) mapping of geomorphic units in the study area, specifically addressing the diffusion problem associated with ambiguous boundaries in aggregated landforms. The observed map, serving as the ground truth, is derived from detailed field surveys and existing geomorphological data, as extensively documented by van der Meij et al.[27].

The predicted map, generated using advanced deep learning techniques, illustrates the model's ability to replicate complex natural patterns. However, it also underscores the challenges in precisely delineating distinct boundaries between various geomorphic units. Displayed landforms include push moraines (1a), other ridges (1b), sloping areas (1c), complex landforms (1d), plains (1e), non-valley-shaped depressions (1f), and valleys (1g). The vibrant colors used in both maps aid in distinguishing between different geomorphic units, highlighting the accuracy and areas requiring improvement in the computational model's performance.

Figure-9 presents a side-by-side comparison of ground-truth (observed, left) and computationally predicted (right) maps across various geomorphic units. The maps highlight differences and accuracies in the representation of aggregated landforms. The legend specifies the types of landforms depicted: 1a represents push moraines, 1b other ridges, 1c sloping areas, 1d complex landforms, 1e plains, 1f non-valley-shaped depressions, and 1g valleys. Both figures are adapted from van der Meij et al.[27].

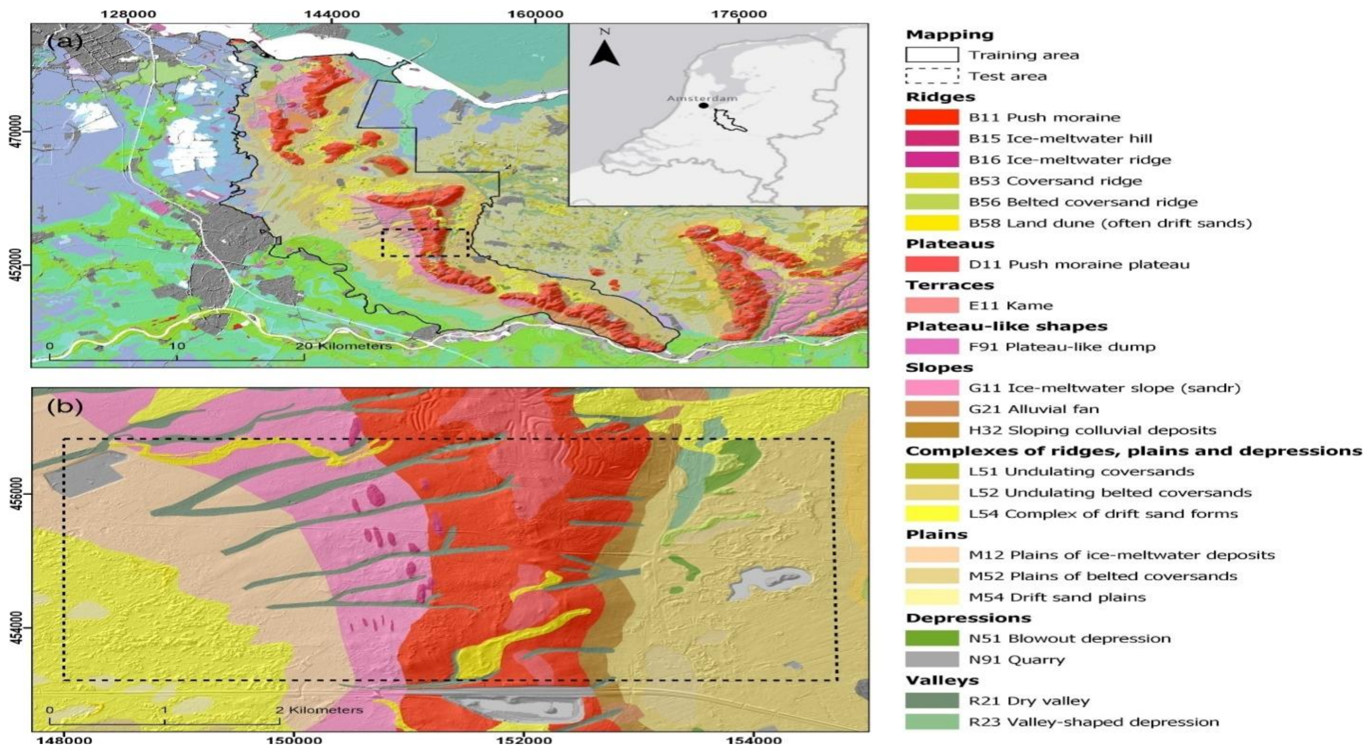

Figure-8: Geomorphological Map of Utrechtse Heuvelrug Study Area.

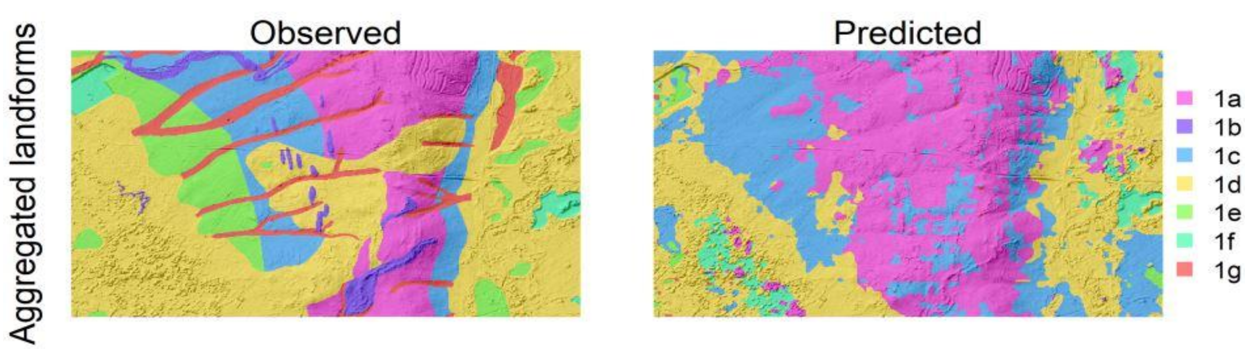

Figure-9: Comparative Overview of Ground-Truth and Computationally Predicted Maps.

Applications of CNN in geomorphic mapping and Research gap

This study investigates the role of CNN in the development of geomorphic maps. For instance, CNN has a 91% accuracy rate when indicating ice-wedge polygons in Alaska[17]. Palafox et al.[21] identification of 94% of the geological characteristics on Mars was accurate with the F1 values of 0.68 and 0.79, Verschoof-Van der Vaart and Lambers[22] found archaeological landforms throughout the Netherlands. Seijmonsbergen et al.[48] used object-based image analysis to map low-relief features in the northern Netherlands. Their results contain kappa values that are about 0.57 off the region's original geomorphological map. The present review found a research gap in deep convolutional network, U-Net method in the context of geomorphic mapping and discussed strategies to fill these gaps. For instance, geomorphological mapping by Van der Meij et al.[27] using deep convolutional network, U-Net approach in the Netherlands found that the general geographical position of each class is predicted correctly for the aggregated landforms. However, this study investigates a research gap, that is, the boundaries of aggregated landforms mapped from the Utrechtse Heuvelrug in the Netherlands are diffused by using the deep convolutional network, U-Net. Additionally, this study concludes that boundaries with a clear definition could improve the accuracy and performance of landform classification.

The current review suggests three possible strategies to address the gap, precise boundary definition – i. Introducing post-processing morphological operations such as erosion and dilation can aid in eliminating noise, closing gaps, and sharpening boundaries[49]. ii. The Dilated convolutions, often referred to as atrous convolutions, enable the extension of receptive fields without expanding the computational complexity of the model. Using dilated convolutions in the U-Net architecture, The model is capable of capturing a wider spatial context and incorporating larger-scale information, which can aid in delineating boundaries of diffused aggregated landforms[50]. iii. Adopting more sophisticated methods, such as conditional random fields, could significantly enhance the outcomes of CNNs by providing sharper boundaries of the mapped objects[32].

Conclusion

This review paper documented key literature related to deep learning methods (CNNs and U-NET) and their implications for pixel-level classification and semantic segmentation. The following conclusion can be drawn from this review:

This study demonstrates that by using CNN U-Net architecture, a model can get a larger receptive field. Additionally, the combination of a larger receptive field and preservation of spatial information allows for semantic segmentation. Moreover, this combination enables associating each pixel in the input images with a certain class, instead of giving one class to the entire image.

This review shows that U-Net can recognize objects based on local information while considering the larger context. This makes it suitable for geomorphological mapping.

Ice-wedge polygons in Alaska and archaeological landforms in the Netherlands, are the other two prime examples where CNN yields higher results than competitors. Although U-Net is beneficial, it has some restrictions. This approach does not show good performance in defining diffused aggregated landforms.

References

1. Bishop, M. P., James, L. A., Shroder, J. F., & Walsh, S. J. (2012). Geospatial technologies and digital geomorphological mapping: Concepts, issues and research. *Geomorphology*, 137(1), 5–26.
2. Paron, P., & Claessens, L. (2011). Makers and users of geomorphological maps. In *Developments in earth surface processes* (Vol. 15, pp. 75-106). Elsevier.
3. Verstappen, H. T. (2011). Old and New Trends in Geomorphological and Landform Mapping. *Developments in Earth Surface Processes*, 15(1952), 13–38.
4. van Lanen, R. J., Kosian, M. C., Groenewoudt, B. J., & Jansma, E. (2015). Finding a way: Modeling landscape prerequisites for roman and early-medieval routes in the Netherlands. *Geoarchaeology*, 30(3), 200–222.
5. Chelli, A., Bordoni, M., Cappadonia, C., Pepe, G., Rotigliano, E., & Smith, M. (2021). Geomorphological tools for mapping natural hazards. *Journal of Maps*, 17(3), 1–4.
6. Bocco, G. E. R. A. R. D. O., Mendoza, M., & Velázquez, A. (2001). Remote sensing and GIS-based regional geomorphological mapping—a tool for land use planning in developing countries. *Geomorphology*, 39(3-4), 211-219.
7. Wang, D., Laffan, S. W., Liu, Y., & Wu, L. (2010). Morphometric characterisation of landform from DEMs. *International Journal of Geographical Information Science*, 24(2), 305–326.
8. Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, 265, 62–77.
9. Berral, J. L., Gavaldà, R., & Torres, J. (2013). Power-Aware Multi-DataCenter management using machine learning. *Proceedings of the International Conference on Parallel Processing*, 858–867.
10. Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
11. Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B.

A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152(April), 166–177.

12. Li, W., & Hsu, C. Y. (2020). Automated terrain feature identification from remote sensing imagery: a deep learning approach. *International Journal of Geographical Information Science*, 34(4), 637–660.

13. Deng, L., Hinton, G., & Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: An overview. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 8599-8603). IEEE.

14. Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3), 55–75.

15. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.

16. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing.

17. Abolt, C.J., & Young, M.H. (2020). High-resolution mapping of spatial heterogeneity in ice wedge polygon geomorphology near Prudhoe Bay, Alaska. *Scientific Data*, 7(1), 1–7.

18. Bhuiyan, M. A. E., Witharana, C., & Liljedahl, A. K. (2020). Use of very high spatial resolution commercial satellite imagery and deep learning to automatically map ice-wedge polygons across tundra vegetation types. *Journal of Imaging*, 6(12), 137.

19. Du, Y., Wang, W., & Wang, L. (2019). Hierarchical recurrent neural network for skeleton based action recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June*, 1110–1118.

20. Li, S., Xiong, L., Tang, G., & Strobl, J. (2020). Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. *Geomorphology*, 354, 107045.

21. Palafox, L. F., Hamilton, C. W., Scheidt, S. P., & Alvarez, A. M. (2017). Automated detection of geological landforms on Mars using Convolutional Neural Networks. *Computers & geosciences*, 101, 48-56.

22. Verschoof-Van der Vaart, W. B., & Lambers, K. (2019). Learning to look at LiDAR: The use of R-CNN in the automated detection of archaeological objects in lidar data from the Netherlands. *Journal of Computer Applications in Archaeology*, 2(1), 31–40.

23. Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning for digital soil mapping. *Soil*, 5(1), 79-89.

24. LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, 3361(March), 255–258.

25. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2323.

26. Albawi, S., Mohammed, T. A. M., & Alzawi, S. (2017). Layers of a Convolutional Neural Network. *Icet2017*, 1–6.

27. van der Meij, W. M., Meijles, E. W., Marcos, D., Harkema, T. T. L., Candel, J. H. J., & Maas, G. J. (2022). Comparing geomorphological maps made manually and by deep learning. *Earth Surface Processes and Landforms*, 47(4), 1089–1107.

28. Nair and Hinton, (2010). Rectified linear units improve restricted Boltzmann machines. *Journal of Applied Biomechanics*, 33(5), 384–387.

29. Fang, W., Ding, Y., Zhang, F., & Sheng, V. S. (2019). DOG: A new background removal for object recognition from images. *Neurocomputing*, 361, 85–91.

30. Zhang, Z., Liu, Q., & Wang, Y. (2018). Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749–753.

31. Zhao, F., Xiong, L. Y., Wang, C., Wang, H. R., Wei, H., & Tang, G. A. (2021). Terraces mapping by using deep learning approach from remote sensing images and digital elevation models. *Transactions in GIS*, 25(5), 2438–2454.

32. Arnab, A., Zheng, S., Jayasumana, S., Romera-Paredes, B., Larsson, M., Kirillov, A., Savchynskyy, B., Rother, C., Kahl, F., & Torr, P. H. S. (2018). Conditional Random Fields Meet Deep Neural Networks for Semantic Segmentation: Combining Probabilistic Graphical Models with Deep Learning for Structured Prediction. *IEEE Signal Processing Magazine*, 35(1), 37–52.

33. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

34. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem, 770–778.

35. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. *Proceedings - 30*th *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua*, 936–944.

36. Long, J.; Shelhamer, E.; Darrell, T. (2015). Fully

Convolutional Networks for Semantic Segmentation. *IEEE Access*, 7, 43369–43382.

37. Haralick, R. M., & Shapiro, L. G. (1985). Image segmentation techniques. *Computer Vision, Graphics, & Image Processing*, 29(1), 100–132.

38. Purwanto, A., Budiarti, D. H., Wibowo, M., Tanasa, I. Y., Guno, Y., Yunata, A. S., Hidayat, A., Purnamastuti, F. N., & Dirgahayu, D. (2023). Image Segmentation in Aerial Imagery: A Review. *Sinergi (Indonesia)*, 27(3), 343–360.

39. Baeldung (2026). Computer Science Tutorials and Guides Instance Semantic Segmentation CNN.

40. Dudhat, H., Talati, B. J., Rana, A. M., & Bakaraniya, P. V. (2020). Image semantic segmentation based on E-net with different patch-size convolution. In *2020 4*th *International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 596-601). IEEE.

41. Vanschoren, J. (2020). Aerial Imagery Pixel-level Segmentation. 1–30.

42. Harshall, L. (2019). Understanding Semantic Segmentation with UNET.

43. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.

44. Jongmans, A. G., Van den Berg, M. W., Sonneveld-Van Dijk, S., Peek, G. J. W. C., & van Saparoea, M. V. D. B. (2018). Landschappen van Nederland: geologie, bodem en landgebruik.In *Landschappen van Nederland*.Wageningen Academic.

45. Stouthamer, E., Cohen, K., & Hoek, W. Z. (2020). De vorming van het land. *Geologie en geomorfologie*. *December*, 448.

46. Van den Berg, M. W., & Beets, D. J. (1987). Saalian glacial deposits and morphology in The Netherlands. *Tills and Glaciotectonics. Balkema, Rotterdam*, 235-251.

47. Pierik, H. J., van Lanen, R. J., Gouw-Bouman, M. T., Groenewoudt, B. J., Wallinga, J., & Hoek, W. Z. (2018). Controls on late-Holocene drift-sand dynamics: The dominant role of human pressure in the Netherlands. *The Holocene*, 28(9), 1361-1381.

48. Seijmonsbergen, A. C., Hengl, T., & Anders, N. S. (2011). Semi-Automated Identification and Extraction of Geomorphological Features Using Digital Elevation Data. In *Developments in Earth Surface Processes* (Vol. 15). Elsevier B.V.

49. Sarzana, T., Maltese, A., Capolupo, A., & Tarantino, E. (2020). Post-processing of Pixel and Object-Based Land Cover Classifications of Very High Spatial Resolution Images. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12252 LNCS* (September), 797–812.

50. Yu, F., & Koltun, V. (2016). Multi-scale context aggregation by dilated convolutions. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*.

51. Maxwell et al., 2020. (2020). Mapping the topographic features of mining-related valley fills using mask R-CNN deep learning and digital elevation data. *Remote Sensing*, 12(3), 1–23.

52. Passalacqua, P., Tarolli, P., & Foufoula-Georgiou, E. (2010). Testing space-scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscape. *Water resources research*, *46*(11).

53. Trier, Ø. D., Cowley, D. C., & Waldeland, A. U. (2019). Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. *Archaeological Prospection*, 26(2), 165-175.

54. Heffels, M. R., & Vanschoren, J. (2020). Aerial imagery pixel-level segmentation. *arXiv preprint arXiv:2012.020 24*.