Journal of Engineering Science and Technology Review 18 (2) (2025) 67- 75

Research Article

JOURNAL OF
**Engineering Science and Technology Review**

www.jestr.org

# Breaking Barriers in Satellite Image Segmentation: A U-Net Ensemble Approach

**Vikas Sharma[1,*], Akshi Kumar[2] and Kapil Sharma[3]**

*[1]Department of Computer Science and Engineering, Jaypee Institute of Information Technology, Noida, India*
*[2]Department of Computing, Goldsmiths, University of London, United Kingdom*
*[3]Department of Information Technology, Delhi Technological University, New Delhi, India*

Received 11 February 2025; Accepted 23 March 2025

***Abstract***

Satellite image analysis is revolutionizing fields like urban planning, environmental conservation, and disaster management, but achieving high-precision object detection remains a challenge. Among deep learning architectures, U-Net and its variants have proven to be highly effective for image segmentation. This paper explores three widely used U-Net variants—U-Net, Attention U-Net, and Attention Res U-Net—and evaluates their performance in identifying objects within satellite images. To push the boundaries of segmentation accuracy, we propose a novel ensemble model that strategically integrates the strengths of these three architectures. Our approach follows a structured pipeline comprising data preprocessing, patchification, training data preparation, model training, and ensemble integration. By combining multiple models, the ensemble method enhances performance, robustness, and generalization in complex satellite image segmentation. We assess the effectiveness of the proposed model using the Dubai Satellite Imaging dataset from the Mohammed Bin Rashid Space Centre (MBRSC). Experimental results show that our ensemble model significantly outperforms individual models, achieving an Intersection over Union (IoU) of 89.5% and a Dice Score of 94.46%, demonstrating its superior ability to delineate objects accurately. Furthermore, a comparative analysis with state-of-the-art architectures, including Inception ResNet V2 U-Net and 3D U-Net, further validates its effectiveness. The results highlight the exceptional segmentation performance of our approach, making it a promising solution for high-precision satellite image analysis. In conclusion, our ensemble model sets a new standard in satellite image segmentation, surpassing existing methods and paving the way for more accurate and reliable deep-learning-based object detection.

*Keywords:* U-Net, Satellite image segmentation, Ensemble, Attention U-Net, Attention Res U-Net, Deep learning

## 1. Introduction

Satellite image segmentation is a process that involves dividing satellite images into distinct sections or sectors, facilitating recognition and analysis of specific objects or categories within the images [1, 2]. Combining segmentation with other image processing techniques, such as classification, allows for deeper insights to be derived from satellite imagery [3]. The increasing demand for up-to-date geospatial data and efficient information extraction methods emphasizes the significance of satellite image segmentation for stakeholders [4]. Segmenting satellite imagery is vital for the analysis and interpretation of satellite data, facilitating the identification and examination of specific features like buildings, roads, or bodies of water [5]. This information has diverse applications, including mapping, monitoring changes in land use and cover, and supporting disaster management efforts [6]. Several approaches to satellite image segmentation exist, including:

i. Hard segmentation: In hard segmentation, each pixel in the image is categorized into a single class or region, with distinct lines separating each region.
ii. Soft segmentation: Soft segmentation allows for some degree of overlap between distinct regions by giving each pixel in the image a probability or likelihood of belonging to each class or region.
iii. Object-based segmentation: Unlike class- or region-based segmentation, object-based segmentation divides the image into discrete objects or features.
iv. Unsupervised segmentation: In unsupervised segmentation, the algorithm must identify patterns and connections in the image on its own without the aid of training data or any prior knowledge.
v. Supervised segmentation: With supervised segmentation, the algorithm is taught how to categorize various sections of the image using a collection of labeled training data.
vi. Semi-supervised segmentation: In semi-supervised segmentation, the system must classify the remaining portions of the image using a small quantity of labeled training data. When it is not practical to label the full image but correct classification is still required, this can be helpful.

Gaps in Existing Research

a) Limited Comparative Analysis – Previous studies have explored U-Net variants, but a detailed comparison of their effectiveness in satellite image segmentation is lacking.
b) Lack of Robust Ensemble Methods – Most approaches use a single architecture, whereas an ensemble approach can improve segmentation accuracy and model robustness.
c) Challenges with Data Availability – Many deep learning models require large annotated datasets, but their performance on limited satellite image data remains underexplored.

*E-mail address: vikassharma12387@gmail.com

doi:10.25103/jestr.182.09

d) Scalability and Computational Efficiency Issues – Existing models demand high computational power, making them difficult to deploy in real-time applications.
e) Insufficient Benchmarking on Real-World Datasets – Several studies use small or synthetic datasets, but evaluating large-scale datasets (e.g., MBRSC's Dubai Satellite Imaging dataset) can provide more realistic insights.

Motivation for the Study

i. Enhancing Accuracy in Satellite Image Segmentation – A more precise model can improve applications like disaster management, urban planning, and environmental monitoring.
ii. Bridging Performance Gaps in U-Net Variants – While individual U-Net models have strengths, their limitations motivate an ensemble approach to achieve better feature extraction and segmentation.
iii. Improving Generalization Across Diverse Terrains – Satellite images contain varied features like buildings, roads, water bodies, and vegetation, requiring a model that performs consistently well across different landscapes.
iv. Ensuring Practical Deployment with Computational Efficiency—The proposed ensemble model balances accuracy with efficiency, making it feasible for real-time satellite image analysis.

This paper focuses on hard segmentation, which assigns each pixel to distinct classes or regions, clearly delineating their boundaries. Convolutional neural network (CNN) technology has emerged as a powerful tool in remote sensing applications [7]. CNNs automatically learn multilayer representations, enabling the mapping of input data to vectors (regression) or the classification into multiple or binary labels [8]. CNN's "representation learning" capabilities have revolutionized traditional feature design in classification and detection tasks, simplifying the process, including building extraction [9]. To achieve dense prediction, the original CNN architecture was extended in 2015 to create fully convolutional networks (FCNs), which up-sample low-resolution features to the input size using transposed convolutions and downscale feature maps using convolutions [10]. Since then, various FCN structures, including U-Net, have been proposed [11]. While U-Nets have shown significant progress in terrestrial segmentation, they face limitations when data availability is scarce. To address the challenges of aerial segmentation with limited data, this research introduces a novel deep-learning framework that combines three U-Net variants (U-Net, Attention-U-Net, and Attention-Res U-net) into an ensemble model. These models were selected for their improved performance and faster training compared to other complex models.

A U-Net ensemble approach can significantly enhance satellite image segmentation by leveraging multiple models trained on diverse spectral bands, improving accuracy and robustness. When integrated into a Digital Twin (DT) framework [12,13] this approach enables real-time monitoring and predictive analytics for applications like disaster management, urban planning, and precision agriculture [14]. The DT continuously updates itself with segmented satellite images, allowing dynamic environmental tracking, early anomaly detection, and automated decision-making. However, optimized architectures and cloud-based solutions must address challenges like computational complexity and real-time processing constraints. This fusion of AI-driven segmentation and DT analytics paves the way for smarter, more responsive geospatial systems.

The research utilizes Kaggle's semantic segmentation of the aerial imagery dataset, which provides pixel-by-pixel annotations of aerial images of Dubai collected by MBRSC satellites and classified into six categories: buildings, land, roads, vegetation, water, and unlabeled areas. This paper makes the following key contributions:

i. The paper presents a thorough review of research papers that concentrate on U-Net, Attention UNet, and Attention Res U-Net in the domain of aerial imaging.
ii. By integrating the strengths of U-Net, Attention U-Net, and Attention Res U-Net, the proposed ensemble model offers a promising approach for improving image segmentation accuracy and simplifying the model architecture.
iii. The effectiveness of the ensemble model is assessed through experimental results, which strongly support its efficacy for image segmentation tasks. By leveraging the ensemble approach, both segmentation accuracy and efficiency are significantly improved.

## 2. Literature Exploration

In this section, the work related to the field of image segmentation is reported in the literature. A few high-quality studies on segmentation techniques have been reported during the past few decades [15,16], provided several important image segmentation techniques and examples. In [17], the researchers categorize image segmentation techniques and include pertinent applications. Additionally, [18] assessed the application of four algorithms from each main category, underlining the significance of the values chosen for segmentation parameters, which have a direct bearing on the segmentation outcome [19] displayed some creative theoretical and experimental findings from work on image segmentation and categorization. The authors in [20] reviewed image segmentation techniques to categorize the most recent research findings on optical remote sensing pictures for readers. Also, the authors in [21,22] have done a lot of research. In [23] authors describe a precise neural network technique in this study to identify roads in satellite pictures. This model is based on convolutional neural networks and also utilizes U-Net. According to experimental findings, pretraining on ImageNet typically enhances several models' segmentation abilities [24]. In [25] the author recommends separating the mosaic floor tiles of the S. Stephen church in Umm Rasas using an automated process based on deep learning. This technique enables the automatic and reliable extraction of the description of the main non-homogeneous tesserae of the mosaic. This review primarily covers U-Net architectures, such as Res U-Net and GC U-Net, used in various satellite image analysis applications, including road mapping, building detection, and small target identification. The main findings from a few of the relevant papers are represented in Table 1.

**Table 1.** Summary of Literature of U-Net in terms of Techniques, Dataset and Analysis

| Ref No. | Paper Method | Dataset | Remarks/Analysis |
|---|---|---|---|
| [23] | Modified U-Net | Massachusetts roads dataset | The proposed model utilized a U-Net type model and atrous Convolution Architecture from a 6-channel image to transfer the road structure to the output. |
| [24] | E-Net, Link Net, U-Net(with various backbones are used) | Open data available from OpenStreetMap (OSM) | This technique employs graph theory, and image-Net pretraining and enhances a number of models' segmentation ability. |
| [25] | DCNN U-Net | Tesserae dataset | It focuses on automatic and reliable acquisition of the description of tesserae of mosaic. |
| [26] | Hybrid First and Second Order Attention Network (HFSA) | Massachusetts buildings dataset, Inria building dataset, Hunan University building dataset, Wuhan University | In this, global mean and inner product among several channels are considered. It used first-order and second-order feature statistics. |
| [27] | U-Net | New York, Boston region satellite images | The presented method is used for identifying and extracting building footprints. Post-processing algorithms are also applied. |
| [28] | Res U-Net | ISPRS 2D Potsdam dataset | It uses residual connections and atrous Convolutions. This also uses pyramid scene parsing and multi-task inference. |
| [29] | EOSRes U-Net | Deep Globe Road Extraction Challenge dataset | This is used to extract structures with different sizes, forms, and textures.<br>It utilized CNN architecture. |
| [30] | Dilated-Res U-net | Building datasets for Delhi, Hyderabad, and Bengaluru (Indian cities), using Sentinel-2 satellite images and Planet OSM | It extracts building extraction,<br>It also focuses on image enhancement and extensive experimentation. |
| [31] | Deep U-Net | Sea-Land Segmentation dataset | It integrates the Global Context Block to capture long-range dependencies, utilizes SPConv to emphasize intrinsic information, and incorporates Adaptively Spatial Feature Fusion to handle features from multiple levels. It employs Up Block and Down Block. |
| [32] | Res U-Net | South Africa field extraction dataset | In this, the composite image is used to achieve high thematic and geometric accuracy.<br>Here, Accuracies increased by averaging predictions from several dates. |
| [33] | Siamese U-Net | Massachusetts buildings dataset, ISPRS dataset | This proposes a Siamese fully convolutional network model. This also produces better segmentation accuracy. |
| [34] | GC U-Net | Landsat8 dataset | This uses SPConv to focus on intrinsic information and Global Context Block to obtain long-range dependencies. This method employs adaptively Spatial Feature Fusion to process features from various levels. |
| [35] | U-Net | ATLANTIS dataset | This employs a security component - ECDSA. |
| [36] | Modified U-Net | Spacenet dataset | This method added two encoders in U-Net. |
| [37] | SR-FCN | WHU Building dataset | In this, a scale-robust CNN structure is employed, which also combines data augmentation and relative radiometric calibration. |
| [38] | UNet with Resnet 18 backbone | Airbus Ships Detection dataset | This method employs a classifier based on Xception and uses U-Net with Resnet 18 for segmentation. |

| | | | |
|---|---|---|---|
| [39] | Attention U-Net | SentinelHub dataset | Presented Model with fewer parameters, Less training time, and Improvement in complex datasets. |
| [40] | K U-Net | 2018 Data Science Bowl dataset | Convolution Long Short-Term Memory (CLSTM) is used., Pretrained weights used. |
| [41] | Inception U-Net | Cloud Detection dataset | Convolution layers and Inception layers were used., Factorized and unfactorized convolutions were used. |
| [42] | Solar-Net | Deepsolar dataset | It is made to recognize solar farms. In this, the Location and size of farms can be determined. |
| [43] | Modified encoder-decoder convolution network | Buildings dataset, INRIA Aerial Image Labeling dataset, Vaihingen ISPRS 2-D Semantic Labeling dataset | Here, Batch normalization layers are updated. Boosts the model performance by refining it on a few samples. |
| [44] | Deep Res U-Net | Jiangsu LINYANG Power Station dataset | Uses residual units in the encoder and decoder path. Employed modified ResNet-34 to get good features. |
| [45] | W-Net | Water Body dataset | The proposed Model employs an inception block in the encoder and decoder paths. Also, asymmetric convolutions and Refinement modules are used. |
| [46] | Wetland Net | Yellow River Estuary wetland dataset | In this, the author utilizes Depth-wise separable convolutions. The presented method extracts the object's boundary contour features using deconvolution. |
| [47] | Link-Net and D-link-Net | UAV images | The presented model was developed by coupling adversarial networks and multiscale context aggregation. |
| [48] | U-Net | Massachusetts Road dataset | This model is used for edge detection, detection in blurry images, and detection in low-resolution images. It studies the performance of ML algorithms in urban environments. |
| [49] | M U-Net | Blood sample images | The presented framework identifies the early detection of blood cancer and its types. |
| [50] | Dual-Path Morph-U-Net | Massachusetts Roads and Buildings dataset | The suggested model develops a structural element that can capture different footprint sizes. Also, uses morphological networks. |
| [51] | U-Net | Airbus ship detection challenge dataset | It offers a cutting-edge framework for removing highways and structures from remote sensing (RS) images by utilizing morphological networks. Also, analyze the performance of models in various backgrounds. |
| [52] | Modified U-Net | Satellite image datasets | The proposed model is applied to analyze flood-affected areas. Here, Segmentation using the Mean shift clustering algorithm. |
| [53] | Broad U-Net | Precipitation maps dataset, Cloud cover dataset | This is built as an extension to the U-Net model and learns more complicated patterns. The model also requires fewer parameters. |
| [54] | CAA-U-Net | Landsat-8, high-resolution cloud (HRC) cover validation dataset | This method has an attention mechanism. It uses an asymmetric encoder and decoder blocks and an Attention gate which is modified. |
| [55] | Deep U-Net | Satellite image dataset | This utilizes FAAGKFCM and SLIC preprocessing techniques. It also analyses Mapping, usability, and change identification. |

| | | | |
|---|---|---|---|
| **Proposed Model** | **Ensemble Model (U-Net, Attention U-Net, and Attention Res-U-Net)** | **MBRSC's Dubai Satellite Imaging dataset.** | **It utilizes the concept of three variants of U-Net architecture, U-Net, Attention U-Net, and Attention Res U-Net.** |

## 3. Related Terminology

These three models—U-Net, Attention U-Net, and Attention Res U-Net have all been employed in the research.

### 3.1 U-Net

A deep learning architecture for segmenting images is called U-Net [5]. An encoder network and a decoder network make up U-Net, and they are connected via several skip links. The decoder network uses the features that the encoder network has extracted from the input picture to rebuild the image and produce the segmentation mask [27]. Through the use of skip connections, the decoder network can blend low-level characteristics from the encoder network with high-level features it has learned, enhancing the model's overall accuracy [51].

When trained on a large dataset, U-Net has been demonstrated to attain extremely high accuracy on several image segmentation tasks. U-Net is an excellent option for applications when data is limited because it has been found to perform well even when trained on short datasets [55]. U-Net generates a segmentation map that is simple to understand and visualize, making it reasonably simple to interpret. The softmax, which is usually used to train the U-Net, is defined as:

$$p_k(x) = \exp\big(a_k(x)\big) / \left(\sum_{k'=1}^{k} \exp\left(a_{k'}(x)\right)\right) \quad (1)$$

where $a_k(x)$ represents the feature channel $k$ activation at the pixel position $x$ with Z2. $k$ Is the total number of classes, while $p_k(x)$ Is the maximum function's approximation. Figure 1 shows the architecture of U-Net.

### 3.2 Attention U-Net

Attention U-Net is a Convolutional neural network architecture used to segment images. As an upgrade over the U-Net architecture, which was also created for image segmentation, it was built [28]. Instead of analyzing the entire image equally, attention mechanisms allow the model to concentrate on particular areas of the input image while making predictions. When the model needs to recognize and classify several objects or regions inside the image [32], such as during image segmentation tasks, this can be quite helpful. Similar to the U-Net architecture, the Attention U-Net architecture comprises an encoder network and a decoder network. The input image is processed by the encoder network, which creates a feature map [39]. This feature map is then sent via the decoder network to create a segmentation map. The decoder network has attention methods that let the model concentrate on particular regions of the feature map while it creates the segmentation map. Attention U-Net can analyze the full picture in a single pass, it can handle high-resolution photos better than U-Net. The attention gates, which are included in the typical U-Net architecture, are used to draw attention to important details that are sent over skip connections. To filter through noisy and irrelevant answers in skip connections, information acquired from a coarse scale is used in gating. To ensure that only pertinent activations are combined, this is done just before the concatenation procedure. The activations of neurons both during the forward pass and the backward pass are additionally filtered by Attention Gates. Below is a formulation of the update rule for the convolution parameters in layer $l-1$ where the Attention coefficients, $\alpha i \in [0, 1]$ Input features $(xl)$, trainable kernel parameters $(\Phi^l)$:

$$\frac{\partial\left(x_i^l\right)}{\partial\left(\Phi^{l-1}\right)} = \frac{\partial\left(\alpha_i^l f\left(x_i^{l-1};\Phi^{l-1}\right)\right)}{\partial\left(\Phi^{l-1}\right)} = \alpha_i^l \frac{\partial\left(f\left(x_i^{l-1};\Phi^{l-1}\right)\right)}{\partial\left(\Phi^{l-1}\right)} + \frac{\partial\left(\alpha_i^l\right)}{\partial\left(\Phi^{l-1}\right)} x_i^l \quad (2)$$

### 3.3 Attention Res U-Net

The attention mechanisms of the Attention U-Net are combined with the residual connections of a Residual U-Net in the Attention Res U-Net variation of the Attention U-Net architecture (Res U-Net). Residual U-Net (Res U-Net) is a convolutional neural network that was created as an enhancement to the UNet design for picture segmentation [24]. The U-Net architecture is given residual connections, which can enhance the model's capacity for learning and lower the likelihood of overfitting. When producing predictions, the model can concentrate on particular areas of the input image thanks to the attention processes, and the residual connections can help the model learn and become broader. These three models have been used in a range of domains, including medical image analysis, computer vision, and satellite image segmentation, and they have generally demonstrated good results in a variety of image segmentation tasks [42].

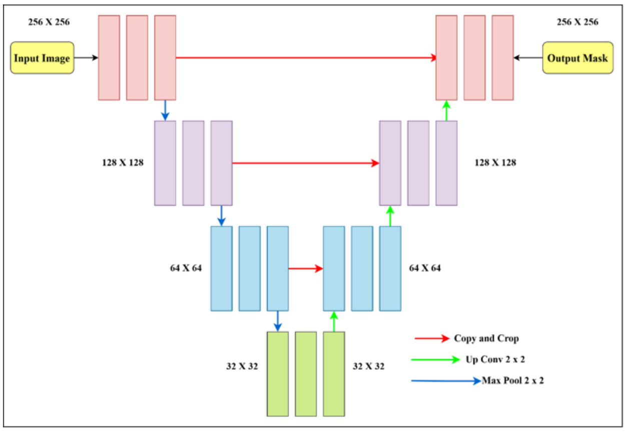



**Fig. 1.** U-Net, which developed from the conventional convolutional neural network, was created and used for the first time in 2015 to process images used in biomedicine [11].

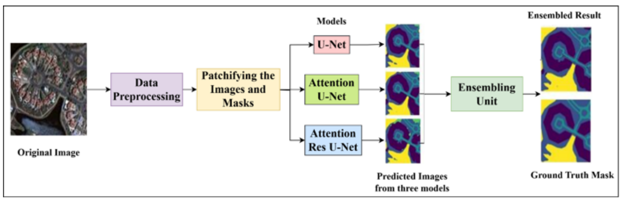



**Fig. 2.** Proposed Model of Image Segmentation

## 4. Proposed Ensemble Model

This section presents the new ensemble model of three U-Net variants—U-Net, Attention U-Net, and Attention Res U-

Net—graphically represented in Figure 2. The proposed ensemble model incorporates six distinct parts: data preprocessing, patching the images and masks, preparing training data, training the models, testing, and ensemble of the results, and evaluating the proposed model.

**Step 1: Data Preprocessing –** Firstly, the image dataset is loaded. Then, preprocessing is applied to the image dataset to get it on the same scale.
**Step 2: Patching the Images and Masks** – The dimensions of the image dataset are large, the patching process and the mask are applied to get a minimal dimension. The preprocessed dataset is passed as input and produces patched images of minimal dimensions. The minimal dimension makes faster the training process and reduces the training time.
**Step 3: Preparation and Training the Model:** The patched images are further broken up into training, testing, and validation splits with a ratio of 70:20:10. The patched images are passed as input to three U-Net models – U-Net, Attention U-Net, and Attention Res U-Net. All three U-Net models are trained and validated independently on the prepared dataset.
**Step 4: Ensemble Unit-** The proposed model is tested on each U-Net, Attention U-Net, and Attention Res U-Net. The predictions from each model are taken and passed to the ensemble unit. In furtherance, the optimal results of the proposed model are obtained. The process of the ensemble is explained through procedure 4.1

### 4.1 Ensemble Procedure
The ensemble procedure used in the proposed model is mentioned below:

---

$$for\ rim(i)\ \in\ \{0,1,2,\ldots\ldots.., nr\ -\ 1\}$$
$$for\ cim\ (j)\ \in\ \{0,1,2,\ldots\ldots.., ncl\ -\ 1\}\ do$$
$$if\ (gt[i][j]\ ==\ m1p[i][j])\ do$$
$$eo[i][j]\ =\ m1p\ [i][j]$$
$$elif\ (gt[i][j]\ ==\ m2p[i][j])\ do$$
$$eo[i][j]\ =\ m2p[i][j]$$
$$elif\ (gt[i][j]\ ==\ m3p[i][j])\ do$$
$$eo[i][j]\ =\ m3p[i][j]$$
$$final\_pred\ =\ eo$$

---

where *nr*= no. of rows, *p* prediction, *m1p* – model 1 prediction, $e_O$- ensemble output, ground _truth, *rim*= rows in image, *cim*= columns in image, *nc*= no. of columns

**Step 6: Model Evaluation:** The predicted patches are joined together to generate the original image. In furtherance, the effectiveness of the proposed model is determined through IoU and Dice Score.

By integrating the strengths of U-Net, Attention U-Net, and Attention Res U-Net, the proposed ensemble model offers a promising approach for improving image segmentation accuracy and simplifying the model architecture.

## 5. Experimental Results and Comparative Analysis

In this paper, a comprehensive series of experiments was carried out to assess the performance of an ensemble model consisting of three variants of the U-Net architecture: U-Net, Attention U-Net, and Attention Res U-Net. The competence of the proposed model was determined on the MBRSC's Dubai Satellite Imaging dataset, obtained from Kaggle. Due to the large size of the dataset images, such as one example with dimensions of 797 x 644, they were resized into 256 x 256 patches. The satellite images dataset was divided into six larger tiles: Building, Land, Road, Vegetation, Water, and Unlabeled. The details of the specific parameters used in training are provided in Table 2.

**Table 2.** Parameters for all the three models

| Epochs | 100 |
|---|---|
| **Learning Rate** | 0.0001 |
| **Optimizer** | Adam Optimizer |
| **Loss Function** | Dice Loss |

To evaluate the effectiveness of the models, Dice Score and IoU Score are evaluated. The IoU and Dice Score are evaluated through equation 3 and equation 4.

**IoU Score:** The projected segmentation map and the ground truth labels overlap are measured by IoU. IoU is between 0 (no overlap) and 1. (perfect overlap). IoU score is formulated in equation 2.

$$IoU\ score\ =\ (P\ \cap\ Q)\ /\ (P\ \cup\ Q) \tag{3}$$

IoU, which accounts for both false positive and false negative errors, is frequently used as a gauge of the accuracy of image segmentation models. An improved match between the anticipated and ground truth labels is shown by a higher IoU score.
**Dice Score:** The overlap between the ground truth labels and the expected segmentation map is also measured by the dice score. The range of the dice score is 0 (no overlap) to 1. (Perfect overlap). Its formula is in equation 4 as follows

$$Dice\ score\ =\ 2\ *\ (P\ \cap\ Q)\ /\ (|P|\ +\ |Q|) \tag{4}$$

A higher Dice Score denotes a better match between the predicted and ground truth labels, thus it is frequently used, like IoU, as an accuracy metric for image segmentation algorithms. Dice Score does not account for the size of the union of the two sets of pixels, hence it is more susceptible to false negatives than IoU. The evaluated results are depicted in Table 3.

As shown in Table 3, the ensemble approach outperformed all three individual models in terms of both IoU Score and Dice Score, while also utilizing less data. This research methodology demonstrated superior performance compared to the others, who employed Inception ResNet V2 as an encoder in a UNet model to increase mathematical and structural complexity. The Dubai Satellite Imaging dataset from MBRSC was employed for model training. Shift, scale, and rotate transformations were applied to the satellite images and segmentation masks to achieve optimal performance, which were then trained on the augmented dataset.

**Table 3.** Results of the models

| Models | IoU Score | Dice Score |
|---|---|---|
| Attention Res-UNet | 78.42% | 63.49% |
| Attention UNet | 79.99% | 88.69% |
| UNet | 80.28% | 88.93% |
| **Ensembled Approach** | **89.51%** | **94.46%** |

Table 4 illustrates the Training and Validation IoU curves and Training and Validation Loss curves for the Attention Res-UNet, Attention UNet, and UNet models. As depicted, all three models underwent 100 training iterations (epochs). Generally, as the number of epochs increased, the IoU Score tended to increase, while the loss decreased.

**Table 4.** Training and Validation IoU and Loss Curves

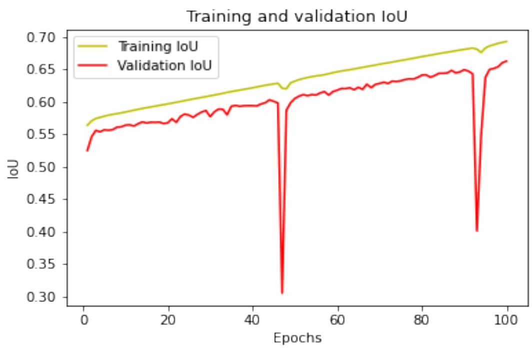
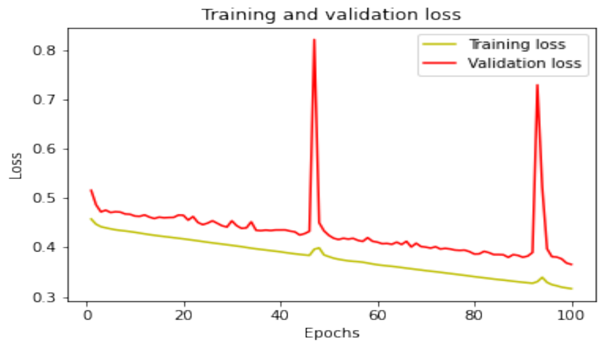
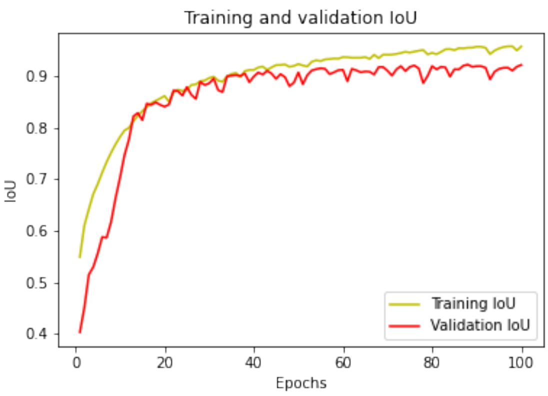
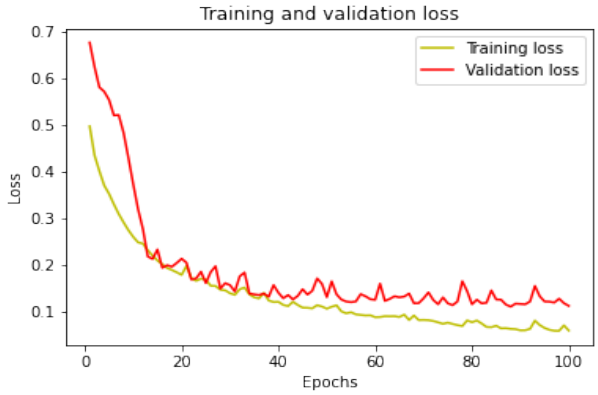
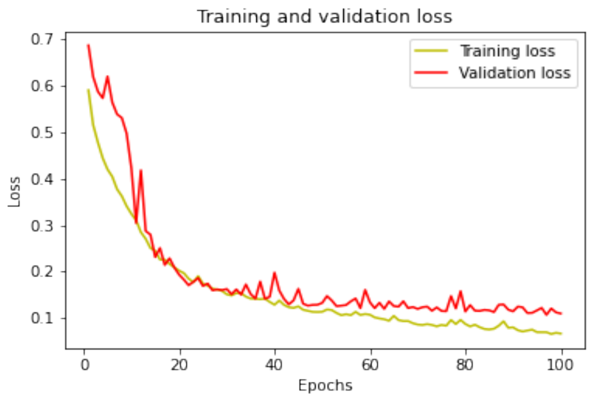
| Models | Training and Validation IoU | Training and Validation Loss |
|---|---|---|
| Attention Res-UNet |  |  |
| Attention UNet | 77  |  |
| UNet |  |  |

Table 5 presents a comparative analysis between the existing method and the proposed ensemble scheme, aiming to demonstrate the competence of the proposed approach.

**Table 5.** Comparative analysis of the research

| **Methods** | **Dice Score** |
|---|---|
| Inception ResNet V2 U-Net [62] | 82% |
| Ensembled (3D U-Net, Res 3-D U-Net) [63] | 91.50% |
| **Ensembled Approach** | **94.46%** |

The ensemble technique employed in this research achieved superior results compared to the previous approach, as it attained a higher Dice Score. This indicates a stronger correlation between the predicted and ground truth labels. Furthermore, the models utilized in this research were less intricate in comparison to the Inception ResNet V2 U-Net and 3D U-Net models utilized previously.

## 6. Conclusion and Future work

This paper comprehensively investigates U-Net architectures, including Attention U-Net and Attention Res U-Net, for satellite image segmentation. Building on these architectures, we propose a novel ensemble model that leverages the strengths of three U-Net variants to enhance segmentation accuracy. The effectiveness of the proposed model is evaluated using the Dubai Satellite Imaging dataset from the Mohammed Bin Rashid Space Centre (MBRSC). Experimental results demonstrate that our ensemble model achieves an Intersection over Union (IoU) of 89.51% and a Dice Score of 94.46%, underscoring its superior segmentation capability. A comparative analysis with Inception ResNet V2 U-Net and 3D U-Net further validates its efficiency, demonstrating its ability to outperform existing state-of-the-art models. Notably, the high Dice Score highlights a strong correlation between predicted and ground truth labels, ensuring precise segmentation. Additionally, the ensemble model simplifies the overall architecture by integrating multiple U-Net variants while surpassing more complex models, emphasizing both its practicality and effectiveness for satellite image analysis. Future work could explore enhanced ensemble strategies by incorporating additional U-Net variations, such as U-Net++ and R2 U-Net, to further improve performance and adaptability in diverse image segmentation tasks.



## References


[1] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523-3542, Jul. 2021, doi: 10.1109/TPAMI.2021.3059968.

[2] G. Pai and S. K. M, "Semi-Dense U-Net: A Novel U-Net Architecture for Face Detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, Jul. 2023, doi: 10.14569/IJACSA.2023.0140643.

[3] I. Kotaridis and M. Lazaridou, "Remote sensing image segmentation advances: A meta-analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 309–322, Mar. 2021, doi: 10.1016/j.isprsjprs.2021.01.020.

[4] S. Lang, "Object-based image analysis for remote sensing applications: modeling reality – dealing with complexity," in *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, T. Blaschke, S. Lang, and G. J. Hay, Eds., Berlin, Heidelberg: Springer, 2008, pp. 3–27. doi: 10.1007/978-3-540-77058-9_1.

[5] H. T. Ngoc, H. V. N. Nguyen, K. H. Nguyen, and L.-D. Quach, "Lane Road Segmentation Based on Improved UNet Architecture for Autonomous Driving," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 7, Jul. 2023, doi: 10.14569/IJACSA.2023.0140724.

[6] S. Kapoor, I. Zeya, C. Singhal, and S. J. Nanda, "A Grey Wolf Optimizer Based Automatic Clustering Algorithm for Satellite Image Segmentation," *Procedia Comput. Sci.*, vol. 115, pp. 415–422, Sep. 2017, doi: 10.1016/j.procs.2017.09.100.

[7] A. Abdollahi, B. Pradhan, and A. Alamri, "RoadVecNet: a new approach for simultaneous road network segmentation and vectorization from aerial and google earth imagery in a complex urban set-up," *GIScience Remote Sens.*, vol. 58, no. 7, pp. 1151–1174, Oct. 2021, doi: 10.1080/15481603.2021.1972713.

[8] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017, doi: 10.1109/TGRS.2016.2612821.

[9] J. Yuan, "Learning Building Extraction in Aerial Scenes with Convolutional Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 11, pp. 2793–2798, Nov. 2018, doi: 10.1109/TPAMI.2017.2750680.

[10] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science, vol. 9351. , Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[12] V. Sharma, A. Kumar, and K. Sharma, "Digital twin: securing IoT networks using integrated ECC with blockchain for healthcare ecosystem," *Knowl. Inf. Syst.*, vol. 67, no. 3, pp. 2395–2426, Mar. 2025, doi: 10.1007/s10115-024-02273-6.

[13] V. Sharma, A. Kumar, and K. Sharma, "Digital twin application in women's health: Cervical cancer diagnosis with CervixNet," *Cogn. Syst. Res.*, vol. 87, Sep. 2024, Art. no. 101264 doi: 10.1016/j.cogsys.2024.101264.

[14] V. Sharma, K. Sharma, and A. Kumar, "From Theory to Practice: A Systematic Review of Digital Twin Implementations Across Industry 4.0," in *2023 14th Int. Conf. Comput. Communic. Network. Technolog. (ICCCNT)*, Delhi, India: IEEE, Jul. 2023, pp. 1–7. doi: 10.1109/ICCCNT56998.2023.10308052.

[15] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Comput. Vis. Graph. Image Process.*, vol. 29, no. 1, pp. 100–132, Jan. 1985, doi: 10.1016/S0734-189X(85)90153-7.

[16] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, Sep. 1993, doi: 10.1016/0031-3203(93)90135-J.

[17] J. Schiewe, "Segmentation of High-Resolution Remotely Sensed Data - Concepts, Applications And Problems", *in Sympos. Geospatial Theo., Process. Applicat.,* Ottawa, Aug. 2002, pp. 1-6.

[18] A. P. Carleer, O. Debeir, and E. Wolff, "Assessment of Very High Spatial Resolution Satellite Image Segmentations," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 11, pp. 1285–1294, Nov. 2005, doi: 10.14358/PERS.71.11.1285.

[19] B. U. Shankar, "Novel Classification and Segmentation Techniques with Application to Remotely Sensed Images," in *Transactions on Rough Sets VII: Commemorating the Life and Work of Zdzisław Pawlak, Part II*, J. F. Peters, A. Skowron, V. W. Marek, E. Orłowska, R. Słowiński, and W. Ziarko, Eds., Berlin, Heidelberg: Springer, 2007, pp. 295–380. doi: 10.1007/978-3-540-71663-1_19.

[20] V. Dey, Y. Zhang, and M. Zhong, "A Review on Image Segmentation Techniques with Remote Sensing Perspective," *in ISPRS TC VII Sympos. – 100 Years ISPRS*, Vienna, Austria, Jul. 2010, pp. 31–42. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=72ea0209ea48769cbda2863c4295cafdaeac41d1

[21] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 277–293, Aug. 2017, doi: 10.1016/j.isprsjprs.2017.06.001.

[22] G. Chen, Q. Weng, G. J. Hay, and Y. He, "Geographic object-based image analysis (GEOBIA): emerging trends and future opportunities," *GIScience Remote Sens.*, vol. 55, no. 2, pp. 159–182, Mar. 2018, doi: 10.1080/15481603.2018.1426092.

[23] A. Constantin, J.-J. Ding, and Y.-C. Lee, "Accurate Road Detection from Satellite Images Using Modified U-net," in *2018 IEEE Asia Pacific Conf. Circ. Sys. (APCCAS)*, Chengdu: IEEE, Oct. 2018, pp. 423–426. doi: 10.1109/APCCAS.2018.8605652.

[24] A. Gupta, S. Watson, and H. Yin, "Deep learning-based aerial image segmentation with open data for disaster impact assessment," *Neurocomputing*, vol. 439, pp. 22–33, Jun. 2021, doi: 10.1016/j.neucom.2020.02.139.

[25] A. Felicetti *et al.*, "Automatic Mosaic Digitalization: a Deep Learning approach to tessera segmentation," in *2018 Metrol. Archaeol. Cultur. Herit. (MetroArchaeo)*, Cassino FR, Italy: IEEE, Oct. 2018, pp. 132–136. doi: 10.1109/MetroArchaeo43810.2018.13606.

[26] N. He, L. Fang, and A. Plaza, "Hybrid first and second order attention Unet for building segmentation in remote sensing

images," *Sci. China Inf. Sci.*, vol. 63, no. 4, Apr. 2020, Art. no. 140305, doi: 10.1007/s11432-019-2791-7.

[27] C. Chawda, J. Aghav, and S. Udar, "Extracting Building Footprints from Satellite Images using Convolutional Neural Networks," in *2018 Int. Conf. Adv. Comput., Communicat. Informat. (ICACCI)*, Sep. 2018, pp. 572–577. doi: 10.1109/ICACCI.2018.8554893.

[28] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, Apr. 2020, doi: 10.1016/j.isprsjprs.2020.01.013.

[29] O. Filin, A. Zapara, and S. Panchenko, "Road Detection with EOSResUNet and Post Vectorizing Algorithm," in *IEEE Conf. Comp. Vision Patt. Recogn. Workshops*, Salt Lake City, UT, USA, Sep. 2018, pp. 211–215.

[30] M. Dixit, K. Chaurasia, and V. Kumar Mishra, "Dilated-ResUnet: A novel deep learning architecture for building extraction from medium resolution multi-spectral satellite imagery," *Expert Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115530, doi: 10.1016/j.eswa.2021.115530.

[31] R. Li *et al.*, "DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 11, pp. 3954–3962, Nov. 2018, doi: 10.1109/JSTARS.2018.2833382.

[32] F. Waldner and F. I. Diakogiannis, "Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network," *Remote Sens. Environ.*, vol. 245, Aug. 2020, Art. no. 111741, doi: 10.1016/j.rse.2020.111741.

[33] S. Ji, S. Wei, and M. Lu, "Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019, doi: 10.1109/TGRS.2018.2858817.

[34] Y. Dong, K. Yu, and W. Hu, "GC-UNet: An Improved UNet Model for Mangrove Segmentation Using Landsat8," in *The 2021 3rd Int. Conf. Big Data Engin.*, Shanghai China: ACM, May 2021, pp. 58–63. doi: 10.1145/3468920.3468928.

[35] A. Ch, R. Ch, S. Gadamsetty, C. Iwendi, T. R. Gadekallu, and I. B. Dhaou, "ECDSA-Based Water Bodies Prediction from Satellite Images with UNet," *Water*, vol. 14, no. 14, Jul. 2022, Art. no. 2234, doi: 10.3390/w14142234.

[36] V. Khryashchev, R. Larionov, A. Ostrovskaya, and A. Semenov, "Modification of U-Net neural network in the task of multichannel satellite images segmentation," in *2019 IEEE East-West Des. Test Sympos. (EWDTS)*, Sep. 2019, pp. 1–4. doi: 10.1109/EWDTS.2019.8884452.

[37] S. Ji, S. Wei, and M. Lu, "A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery," *Int. J. Rem. Sens.*, vol. 40, no. 9, pp. 3308–3322, May 2019, doi: 10.1080/01431161.2018.1528024.

[38] D. Hordiiuk, I. Oliinyk, V. Hnatushenko, and K. Maksymov, "Semantic Segmentation for Ships Detection from Satellite Imagery," in *2019 IEEE 39th Int. Conf. Electron. Nanotech. (ELNANO)*, Kyiv, Ukraine: IEEE, Apr. 2019, pp. 454–457. doi: 10.1109/ELNANO.2019.8783822.

[39] D. John and C. Zhang, "An attention-based U-Net for detecting deforestation within satellite sensor imagery," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 107, Mar. 2022, Art. no. 102685, doi: 10.1016/j.jag.2022.102685.

[40] S. W. Chang and S. W. Liao, "KUnet: Microscopy Image Segmentation with Deep Unet Based Convolutional Networks," in *2019 IEEE Int. Conf. Sys., Man Cybern. (SMC)*, Bari, Italy: IEEE, Oct. 2019, pp. 3561–3566. doi: 10.1109/SMC.2019.8914048.

[41] A. Kumthekar and G. R. Reddy, "An integrated deep learning framework of U-Net and inception module for cloud detection of remote sensing images," *Arab. J. Geosci.*, vol. 14, no. 18, Sep. 2021, Art. no. 1900, doi: 10.1007/s12517-021-08259-w.

[42] X. Hou, B. Wang, W. Hu, L. Yin, and H. Wu, "SolarNet: A Deep Learning Framework to Map Solar Power Plants In China From Satellite Imagery," 2019, *arXiv*. doi: 10.48550/ARXIV.1912.03685.

[43] S. Ghassemi, A. Fiandrotti, G. Francini, and E. Magli, "Learning and Adapting Robust Features for Satellite Image Segmentation on Heterogeneous Data Sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6517–6529, Sep. 2019, doi: 10.1109/TGRS.2019.2906689.

[44] H. Zhang, X. Hong, S. Zhou, and Q. Wang, "Infrared Image Segmentation for Photovoltaic Panels Based on Res-UNet," in *Pattern Recognition and Computer Vision*, Z. Lin, L. Wang, J. Yang, G. Shi, T. Tan, N. Zheng, X. Chen, and Y. Zhang, Eds., Cham: Springer International Publishing, 2019, pp. 611–622. doi: 10.1007/978-3-030-31654-9_52.

[45] R. G. Tambe, S. N. Talbar, and S. S. Chavan, "Deep multi-feature learning architecture for water body segmentation from satellite images," *J. Vis. Commun. Image Represent.*, vol. 77, May 2021, Art. no. 103141, doi: 10.1016/j.jvcir.2021.103141.

[46] B. Cui, Y. Zhang, X. Li, J. Wu, and Y. Lu, "WetlandNet: Semantic Segmentation for Remote Sensing Images of Coastal Wetlands via Improved UNet with Deconvolution," in *Gen. Evolution. Comput.*, J.-S. Pan, J. C.-W. Lin, Y. Liang, and S.-C. Chu, Eds., Singapore: Springer, 2020, pp. 281–292. doi: 10.1007/978-981-15-3308-2_32.

[47] Y. Li, B. Peng, L. He, K. Fan, and L. Tong, "Road Segmentation of Unmanned Aerial Vehicle Remote Sensing Images Using Adversarial Network with Multiscale Context Aggregation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 7, pp. 2279–2287, Jul. 2019, doi: 10.1109/JSTARS.2019.2909478.

[48] N. Y. Q. Abderrahim, S. Abderrahim, and A. Rida, "Road Segmentation using U-Net architecture," in *2020 IEEE Int. Conf. Moroccan Geomat. (Morgeo)*, May 2020, pp. 1–4. doi: 10.1109/Morgeo49228.2020.9121887.

[49] C. Corbane, L. Najman, E. Pecoul, L. Demagistri, and M. Petit, "A complete processing chain for ship detection using optical satellite imagery," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5837–5854, Dec. 2010, doi: 10.1080/01431161.2010.512310.

[50] A. Soni, R. Koner, and V. G. K. Villuri, "M-UNet: Modified U-Net Segmentation Framework with Satellite Imagery," in *Proceed. Global AI Congress 2019*, J. K. Mandal and S. Mukhopadhyay, Eds., Singapore: Springer, 2020, pp. 47–59. doi: 10.1007/978-981-15-2188-1_4.

[51] M. S. Dey, U. Chaudhuri, B. Banerjee, and A. Bhattacharya, "Dual-Path Morph-UNet for Road and Building Segmentation from Satellite Images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Sep. 2022, doi: 10.1109/LGRS.2021.3106772.

[52] S. Karki and S. Kulkarni, "Ship Detection and Segmentation using Unet," in *2021 Int. Conf. Adv. Electr., Comput., Communic. Sustain. Techn. (ICAECT)*, Feb. 2021, pp. 1–7. doi: 10.1109/ICAECT49130.2021.9392463.

[53] S. M. Jaisakthi, P. R. Dhanya, and S. Jitesh Kumar, "Detection of Flooded Regions from Satellite Images Using Modified UNET," in *Computat. Intell. Data Sci.*, V. Krishnamurthy, S. Jaganathan, K. Rajaram, and S. Shunmuganathan, Eds., Cham: Springer International Publishing, Aug. 2021, pp. 167–174. doi: 10.1007/978-3-030-92600-7_16.

[54] J. G. Fernández and S. Mehrkanoon, "Broad-UNet: Multi-scale feature learning for nowcasting tasks," *Neural Netw.*, vol. 144, pp. 419–427, Dec. 2021, doi: 10.1016/j.neunet.2021.08.036.

[55] L. Zhang, J. Sun, X. Yang, R. Jiang, and Q. Ye, "Improving Deep Learning-Based Cloud Detection for Satellite Images with Attention Mechanism," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Sep. 2022, doi: 10.1109/LGRS.2021.3133872.