





## Investigating the performance of super-resolved remote sensing images on coastline segmentation with deep learning-based methods

Ilhan Pala <sup>1</sup>, Ugur Alganci <sup>2\*</sup>

<sup>1</sup> Istanbul Technical University, Communication Systems, Satellite Communication and Remote Sensing Program, Istanbul, Türkiye

<sup>2</sup> Istanbul Technical University, Civil Engineering Faculty, Geomatics Engineering Department, Istanbul, Türkiye

Cite this study:

Pala, I., & Alganci, U. (2025). Investigating the performance of super-resolved remote sensing images on coastline segmentation with deep learning-based methods. *International Journal of Engineering and Geosciences*, 10 (1), 93-106.

<https://doi.org/10.26833/ijeg.1522143>

### Keywords

Remote Sensing  
Super Resolution  
Image Processing  
Coastline Segmentation  
Deep Learning

### Research Article

Received:25.07.2024

Revised: 04.09.2024

Accepted:11.10.2024

Published:01.02.2025



### Abstract

The use of satellite imagery in critical areas, such as environmental monitoring and natural disaster management, is becoming increasingly important. Applications like monitoring coastal areas, detecting coastal erosion, and tracking land use changes demand high accuracy and detailed analysis. Traditional methods for coastline segmentation are often limited by the low resolution (LR) and high complexity of satellite imagery. To address this challenge, Super Resolution (SR) algorithms are employed to enhance the resolution of satellite images, which is particularly beneficial when examining areas with intricate structures, such as coastlines. In this context, the integration of SR and segmentation techniques presents an innovative approach to achieving greater accuracy and efficiency in satellite image analysis. In this study, the resolution of satellite images was enhanced using the Super Resolution Generative Adversarial Networks (SRGAN) model. Thanks to the flexible architecture of the SRGAN model, it was successfully adapted to work with satellite images, yielding satisfactory results. Coastal segmentation was performed using low-resolution, super-resolved, and high-resolution Gokturk-1 (GT-1) satellite images, employing U-net, LinkNet, and DeepLabV3+ segmentation models for comparison. The results indicated that increment in image resolution significantly affects segmentation success. Additionally, better performance in coastline segmentation was achieved with U-net and LinkNet models. Although the DeepLabV3+ model is effective for segmentation, it tends to capture less detail compared to the other two models. Overall, the combination of SRGAN and the LinkNet segmentation model produced results that were closest to reality.

## 1. Introduction

Remote sensing today finds a wide range of applications, including cartography, agricultural land analysis, urban planning, natural disaster investigation, geological studies, and atmospheric and climate research [1-4]. In these fields, the spatial resolution of images is a critical factor influencing the accuracy of the results. Although advancements in satellite technology have improved the spatial resolution of satellite images, obtaining high-resolution images can be costly [5]. However, for relatively low-resolution, open-access satellite images such as those from Landsat-8 and Sentinel-2, deep learning algorithms can be utilized to enhance spatial resolution. This approach facilitates more accurate and consistent analysis.

Numerous studies in the literature address super resolution (SR) of satellite images [6-9]. Goodfellow et al.

[6] introduced the Generative Adversarial Networks (GAN) model. In this model, two different networks are presented. The generative network attempts to produce data that is similar to real data, while the discriminative network separates real and generated data. These two networks work in an adversarial way to make better predictions. Thus, data similar to real data is obtained. Ledig and et al. [7] proposed the SRGAN model and this model aimed to preserve details at high scales in SR tasks by using GAN model. It was confirmed that this model provides more photo-realistic outputs compared to state-of-the-art reference methods. Romero et al. [8] proposed RS-ESRGAN model and aimed to enhance the resolution of Sentinel-2 and Landsat-8 satellite images. RS-ESRGAN was based on the SRGAN model. Modifications were applied to the SRGAN by removing up sampling blocks. In addition, this study evaluated different SR models and compared them with proposed

one. Results were evaluated using PSNR, SSIM, ERGAS, SAM, and CC metrics. In another study, Xiong et al. [9] proposed ISRGAN model which also relied on SRGAN model. This study examined the proposed model's ability to generalize across different types of sensors and contexts. The findings demonstrate a notable improvement in land cover classification accuracy after SR process.

Segmentation models have an important role in image processing and computer vision. These models provide the ability to identify and analyze specific objects, regions, or structures by using deep learning algorithms. Segmentation is utilized in various application areas such as medical imaging, autonomous driving, security systems, and remote sensing [10-13]. Several deep learning (DL) based architectures were developed for segmentation. Ronneberger et al. [10] offered U-Net model. The U-net architecture is capable of end-to-end training with a small number of images using efficient data augmentation strategies. The model has a symmetrical structure between the encoder and decoder paths, thus enabling it to learn both low-level details and contextual information. Chaurasia et al. [11] developed LinkNet model, which is an important innovation for efficient semantic segmentation. LinkNet architecture has direct connections between the encoder and decoder that enables recovery of the lost spatial information. This approach provides accurate segmentation results while significantly reducing the network's processing time. Chen et al. [12] developed the DeepLabV3+ segmentation model. DeepLabv3+ is a sophisticated model that uses Spatial Pyramid Pooling (SPP) and encoder-decoder structures to improve object segmentation in images. This model uses SPP to capture extensive contextual information. Its encoder-decoder architecture generates reliable segmentation results that include rich contextual information and fine details. DeepLab's most recent version [13] is DeepLabV3+.

Remote sensing technologies have advanced significantly in recent years and serve an important role in a variety of industries, including geographic information systems (GIS), environmental monitoring, agriculture, forestry, and disaster management. These advances have also increased the need for efficient algorithms such as powerful segmentation models that can handle large data sets for effective processing and analysis of remote sensing data. As a result, segmentation models have become vital tools for accurately classifying and analyzing remote sensing data. Alsabhan et al. [14], conducted a study with the U-Net model to extract buildings from satellite images. The model was trained using the Massachusetts building dataset. Various U-Net model encoders were tested and compared across experiments. The results revealed the potential of the segmentation model. In another research, Sariturk et al. [15], studied on building segmentation from high-resolution images with four U-net based models. The results were evaluated with Intersection over Union (IoU) and F1 score and indicated acceptable performance of these models. Zhang et al. [16], proposed a deep learning method named Deep Residual U-Net (Deep ResUNet) for semantic segmentation, specifically for

extracting roads. Their proposed network aimed to combine the strengths of residual learning and U-Net. The study revealed the success of U-net and Deep ResUNet models. Han et al. [17], studied on building extraction from remote sensing images and proposed an improved DeepLabV3+ network. The method was compared with U-Net, SegNet, PSPNet, and DeepLabv3+. Tests were performed on the WHU and Massachusetts building datasets. The improved DeepLabv3+ model achieved better results, followed by the DeepLabV3+ model according to IoU and F1 score metrics. Gupta et al. [18], have utilized U-Net and LinkNet models for the segmentation of satellite images to assess post-disaster damage and analyze road conditions. The U-Net model demonstrated high accuracy in segmenting buildings and roads, while the LinkNet model was similarly effective. Experimental results revealed that the U-Net (VGG16) model exhibited superior performance. The results were evaluated with precision, recall and F1 score. Arıkan et al. [19], used the U-Net model to evaluate the performance of Göktürk -1 (GT-1) satellite images for building segmentation. According to results, high Jaccard coefficient and high dice similarity coefficient was obtained.

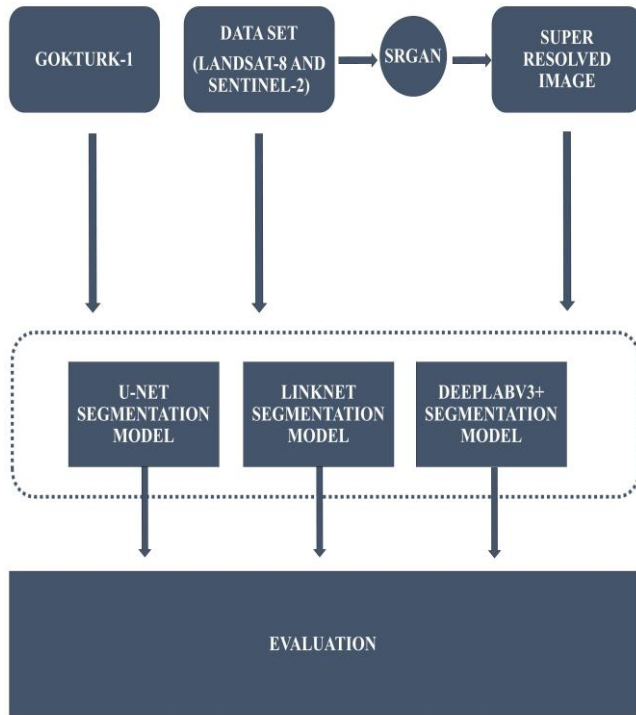
Sea-land segmentation using satellite imagery is an important study field with environmental and socioeconomic implications. Determining the sea - land borders play a crucial role in natural disaster management by enabling rapid identification and mapping of affected areas following events such as tsunamis, floods or coastal erosion. Thus, emergency responders and decision makers can allocate resources efficiently. In addition, sea -land segmentation provides critical data for the sustainable management of coastal ecosystems, detection of marine pollution and protection of coastal habitats in environmental monitoring and protection studies. It also helps to clearly define maritime and land boundaries in planning and managing maritime activities, ensuring safe and efficient operations in areas such as ports, fishing grounds and waterways. Yang et al. [20], studied on DeepLabV3+, SegNet [22], PSPNet [23], FC-DenseNet [24], RefineNet [25] and U-Net models for sea-land segmentation of Landsat-8 OLI satellite images, where the FC-DenseNet model achieved the highest accuracy and the DeepLabV3+ model acquired the best performance in training time efficiency. In another study, Panuntun et al. [21], aimed to identify the optimal semantic segmentation architecture for land cover mapping using remote sensing images. U-net, LinkNet, FPN [26], and PSPNet [23] semantic segmentation models were applied to multispectral, hyperspectral, and high-resolution aerial image datasets to classify vegetation, water, soil and impervious surfaces, etc. According to results, the LinkNet model achieved the highest accuracy. Multispectral images exhibited the highest performance based on IoU and F1 scores, with 0.993 and 0.997, respectively. The results highlight the efficiency and broad applicability of LinkNet and multispectral images for land cover classification.

This research aimed to perform a comparative and quantitative analysis on the effects of super resolved

mid-spatial resolution satellite images on sea-land segmentation problem by use of three DL based segmentation architectures and using Gökürk -1 very high-resolution satellite image as a reference.

## 2. Method

In this study, we aimed to enhance the resolution of Landsat-8 and Sentinel-2 satellite images using the SRGAN model and examine its effects on segmentation. Additionally, we compared the performance of segmentation models, including those using high-resolution Gokturk-1 satellite images. A flowchart of the process is provided in Figure 1.



**Figure 1.** Study Flowchart

### 2.1. Super Resolution

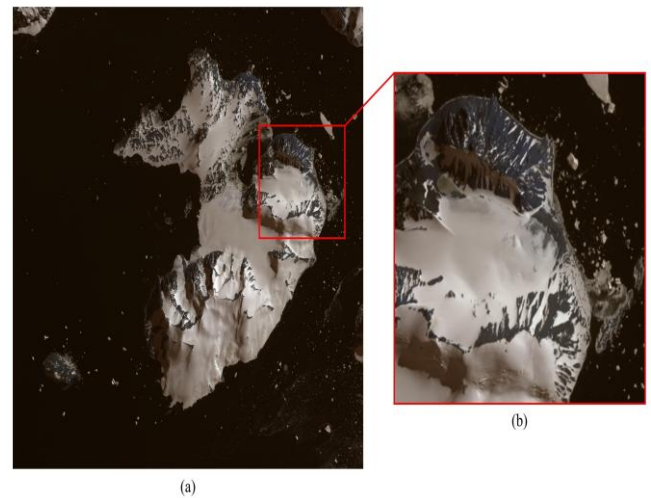
The SRGAN model was chosen for Super Resolution due to its ability to generate detailed, high-resolution images from low-resolution inputs. Its perceptual loss function enhances detail capture. The dataset, configured with 3 bands, was down sampled to create low-resolution images and enriched with data augmentation. Histogram equalization was applied to the SR images to maintain consistent histogram distribution.

#### 2.1.1. Dataset

This study used Landsat 8 and Sentinel 2 satellites for their free access and wide coverage. Landsat 8, operated by the USGS, features the Operational Land Imager (OLI) with a 30-meter spatial resolution. Sentinel 2, managed by the ESA, uses the Multi-Spectral Instrument (MSI), providing images with 10-meter resolution. Additionally, GT-1 satellite imagery from the same region, offering 2-meter MS and 0.5-meter PAN resolution, was used for segmentation comparison.

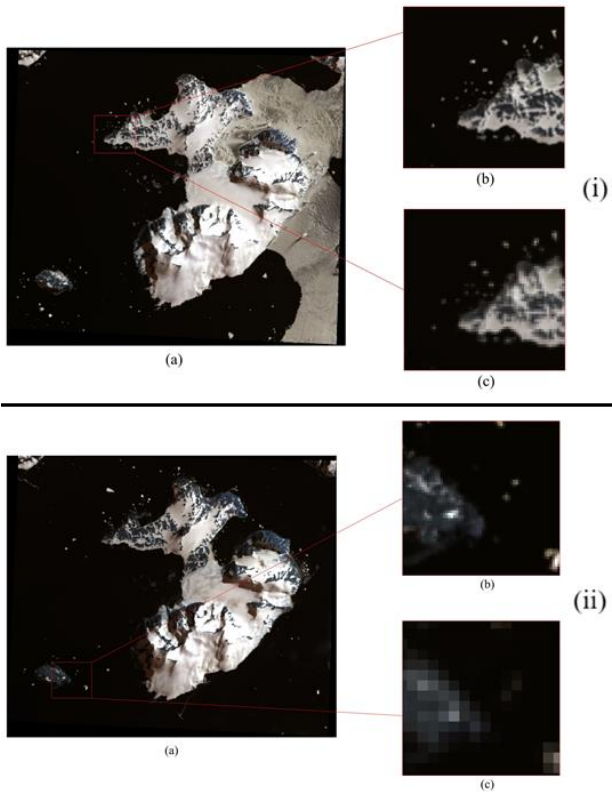
### 2.1.2. Preprocess

To train a deep learning model, a large amount of training data is typically required. However, collecting this data can be difficult, time-consuming, and costly. Therefore, various methods exist to artificially augment the existing dataset. Some of these methods include cropping images from different sections, rotating, scaling, and recreating them by adding noise. By using one or several of these methods together, the amount of image in the dataset can be increased. For more effective training, the images were cropped into  $200 \times 200$  patches with overlapping sections (achieved by shifting). An example of this cropping pattern is shown in Figure 2. To augment the dataset, each image was rotated by  $90^\circ$ .



**Figure 2.** Example of remote sensing image (a) and cropped section image (b)

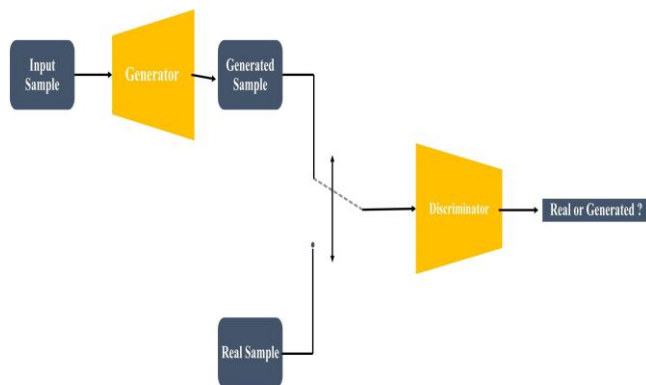
Super Resolution models need datasets with both high- and low-resolution images. Low-resolution images are created by down sampling high-resolution ones, which reduces detail and sharpness. This process can be done at various scales, allowing the model to train for different resolutions. While the overall structure is retained, much fine detail is lost, highlighting the challenges of low-resolution imagery. Figure 3(i) provides a visual of this process for scale factor of 2. As the scale factor increases, the resolution of the image obtained through down sampling decreases significantly. This makes the enhancement of the image more challenging. As seen in Figure 3 (ii), detail information is lost for a scale factor of 4.



**Figure 3.** Satellite Image (a), Image patch (b), down sampled image (c); for (i) 2x and (ii) 4x scale factor.

### 2.1.1.3. SRGAN

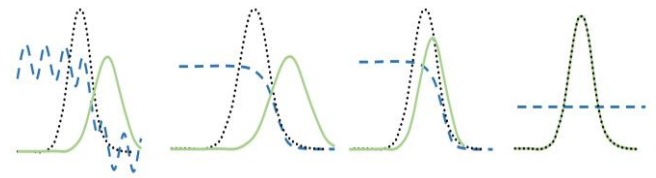
Super Resolution Generative Adversarial Networks (SRGAN) [7] model enables the generation of high-resolution images from low-resolution images. The SRGAN model is based on the architecture of Generative Adversarial Networks (GAN), a type of deep neural network. The GAN model comprises a generator network (model G) and a discriminator network (model D). The generator part facilitates the generation of artificial images through predictions. The discriminator part identifies the differences between the generated images and the original images. The general GAN architecture is shown in Figure 4.



**Figure 4.** General GAN Architecture

In Figure 5 the black dotted lines represent the data distribution to be produced, the blue dashed lines D represents the discriminant distribution, and the green solid lines G represents the generator distribution. The

discriminator is trained simultaneously with the generator. This allows the generator progressively creates more realistic data, and the discriminator enhances its ability to distinguish between genuine and generated data. As training progresses, the data generated by the generator converges towards the input data, becoming more accurate and realistic.



**Figure 5.** Black dotted lines: data distribution to be produced, blue dashed lines: D discriminant distribution, green solid lines: G generator distribution

The generator produces new data based on a specific probability distribution. The generator part utilizes a deep ResNet [27] structure with 16 blocks. Within these blocks, two convolutional layers with 3×3 kernels and 64 feature maps are used. These are followed by a batch normalization layer activated by ReLU [28]. In this layer, parametric values are used instead of fixed values. The purpose of the generative network is to generate new data that cannot be distinguished by the discriminator. At the end of the training process, the generator network improves and reaches a point where the generated images become indistinguishable from real images.

The discriminator part is fed with data from the real high-resolution dataset and the data generated by the generator, and it determines which images are real and which are generated. The discriminator network includes layers with 3×3 filter kernels, with the feature map size increasing from 64 to 512 kernels, adopting a structure similar to the Vgg-19 [29] network. Leaky ReLU is used for the activation function. Finally, the probability of the real high-resolution and the predicted image being the same is determined. During training, losses are continuously calculated, and the training process continues by minimizing these losses.

### 2.1.1.4. Refactoring

In models, pixel values need to be processed and multidimensional arrays created for image enhancement. Generally, models are designed to process these values at an 8-bit depth. Satellite images, such as those from Sentinel-2 and Landsat-8, are obtained with a 32-bit depth, providing rich radiometric information. This difference in data type can be addressed by converting 32-bit images to 8-bit; however, this conversion would result in the loss of color information in pixel values. In scenarios such as satellite image studies, where feature loss must be minimized, it is essential to train and test image data without changing the bit format. Therefore, instead of performing this conversion, the model has been adjusted to work with 32-bit images.

### 2.1.5. Hyperparameters

For training and evaluation, images are cropped and resized to  $200 \times 200$  pixels. The scale is set at 2x and 4x, with down sampling used to create low-resolution images, allowing for high- and low-resolution image pairs. All training runs use the same hyperparameters on an NVIDIA GeForce RTX 3060.

The generator network follows a ResNet architecture, while the discriminator uses a VGG-based design. The difference between the enhanced and original images is calculated with MSE loss, which measures the mean squared error. Additionally, Total Variation Loss (TVLoss) is used to calculate image smoothness. Losses are computed using Adversarial Loss (weight: 1e-3), Perceptual Loss (weight: 6e-3), and TVLoss (weight: 2e-8), with the goal of minimizing these losses throughout training.

### 2.1.6. Postprocess

After obtaining high-resolution images with the SR model, differences in histogram distribution may occur. Matching the histogram distributions of the input images with those of the super-resolved images can improve the results by ensuring the alignment of pixel intensity distributions. This approach is crucial for accurate interpretation and analysis.

### 2.1.7. Evaluation Metrics

#### 2.1.7.1. Mean Squared Error (MSE)

MSE measures the difference between predicted and observed values, calculated as:

$$MSE = \frac{\sum(y_i - p_i)^2}{n} \quad (1)$$

where  $y_i$  is the observed value,  $p_i$  is the predicted value, and  $n$  is the number of observations. An MSE of 0 indicates perfect predictions, while lower values signify closer alignment.

#### 2.1.7.2. Peak Signal-to-Noise Ratio (PSNR)

PSNR uses MSE to compare reconstructed images to the original signal. A higher PSNR indicates greater similarity:

$$PSNR = 10 \log_{10} \left( \frac{MAX^2}{MSE} \right) \quad (2)$$

The MAX value is the maximum pixel value.

#### 2.1.7.3. Structural Similarity Index (SSIM)

SSIM assesses perceptual differences between images, focusing on brightness, contrast, and structure. Values above 0.95 indicate high similarity, while values below 0.9 suggest significant differences:

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (3)$$

$\mu_x, \mu_y$  represent the means of the two compared datasets.  $\sigma_x^2, \sigma_y^2$  are the variances, and  $\sigma_{xy}$  is the covariance.

#### 2.1.7.4. Cosine Similarity (CS)

CS quantifies similarity between two items regardless of size, treating each as a vector. Values between 0.7 and 1.0 indicate high similarity:

$$\text{cosine similarity} = \frac{A \cdot B}{\|A\| \|B\|} \quad (4)$$

#### 2.1.7.5. Correlation Coefficient (CC)

CC calculates the linear relationship between two datasets. A value of 0 indicates no correlation:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (5)$$

#### 2.1.7.6. ERGAS Value

ERGAS assesses the ratio of the square root of MSE to the mean of a reference image, primarily for remote sensing:

$$ERGAS = 100 \frac{h}{\ell} \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{RMSE_i^2}{MR_i^2}} \quad (6)$$

$MR_i$  is the mean radiance of the  $i$ -th MS band,  $h$  denotes the spatial resolution of the high-resolution band, and  $\ell$  represents the spatial resolution of the low-resolution image.

## 2.2. Coastline Segmentation

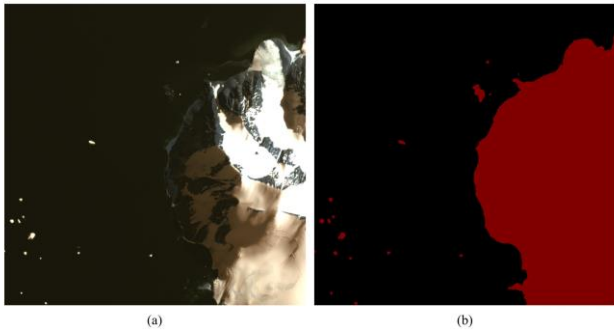
Segmentation was performed using U-net, LinkNet, and DeepLabV3+ models on original and SR Sentinel-2 and Landsat-8 images as well as GT-1 images that is used as a reference. The dataset, enlarged with data augmentation. Land masks were created from NIR band of the images. These masks were reviewed and corrected for obvious errors to finalize the labeled dataset creation.

### 2.2.1. Dataset Generation

The dataset used in the segmentation models is the same as the one used in the SR model. Landsat-8 and Sentinel-2 images were cropped to  $512 \times 512$  dimensions. Three different augmentation functions were applied to the dataset: horizontal flip, vertical flip, and random rotation.

Masks were created using the Near-Infrared (NIR) band. The reflectance values of the NIR band show distinct differences between land and water surfaces, making it effective to create masks using a threshold value determined based on the mean value of the NIR band. This approach is based on the principle that water surfaces have low reflectance values while land surfaces have high reflectance values, allowing for a simple and quick mask creation process. However, this method has limitations: the accuracy of masks can be affected by

environmental factors like snow cover or shadows. To enhance mask accuracy, threshold values may need to be tailored to specific regions or conditions. Additionally, segmented images were carefully reviewed to correct any mislabeled areas. An example image and its corresponding mask are shown in Figure 6.



**Figure 6.** Image section(a) and mask (b)

### 2.2.2. U-net

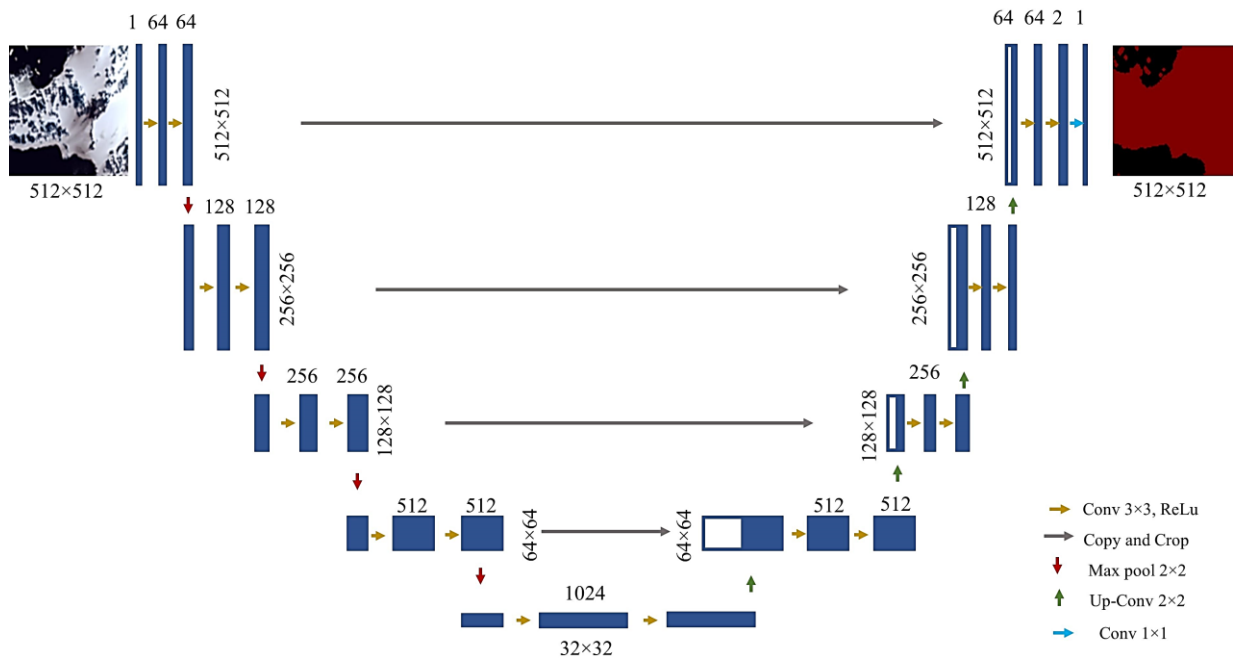
U-Net [10] is a convolutional neural network designed for biomedical image segmentation, providing high-accuracy segmentation by processing input images at various scales. The architecture consists of two symmetric paths named encoder path and decoder path.

In the encoder path, the input image is processed through 3x3 convolutional layers with ReLU activation

functions, followed by 2x2 max-pooling layers, down sampling the image at each step while doubling the number of feature maps.

In the decoder path, the feature maps are up sampled using 2x2 up-convolution layers and processed through 3x3 convolutional layers, then merged with the corresponding feature maps from the encoder path. At this stage, skip connections preserve high-level features from the encoder path, combining them with the up sampled output to prevent loss of detail. In the last layer, a 1x1 convolution network is used and the feature vector is mapped to the desired number of classes.

In traditional auto-encoder architecture, the input data size is progressively reduced through the initial layers, completing the encoder part of the architecture and then starting the decoder part. However, this process causes the model to learn a linear feature representation and leads to the 'bottleneck' problem as the size increases. U-Net differs at this point; it performs up-convolution on the decoder side and avoids the bottleneck problem through connections in the encoder path that prevent the loss of features. The symmetric structure between the encoder and decoder paths enables the model to learn both low-level details and high-level contextual information. The U-net architecture used for this study is as shown in Figure 7.



**Figure 7.** U-net architecture

### 2.2.3. LinkNet

LinkNet [11] is a lightweight convolutional neural network model developed to provide fast and efficient image segmentation. This model is ideal for applications requiring low computational cost and high speed. The architecture of LinkNet consists of two main parts named the encoder path and the decoder path.

The encoder path of LinkNet down samples the input image using various convolutional layers and max-

pooling layers. The first block contains a convolutional layer with a kernel size of 7x7 and a stride of 2, followed by a max-pooling layer with a window size of 2x2 and a stride of 2. Each convolutional block consists of two consecutive 3x3 convolutional layers and a ReLU activation function.

The decoder path expands the feature maps using up-convolution layers. Each up-sampling block merges feature maps from the previous block with feature maps from the corresponding blocks in the encoder path. This

merging process preserves the details from the encoder path and enhances the model's accuracy.

In LinkNet, skip connections are employed between the encoder and decoder paths. These connections allow the low-level features from the encoder path to be combined with the corresponding high-level features in

the decoder path. This increases the model's localization accuracy and minimizes detail loss. The final layer uses a  $1 \times 1$  convolutional layer to determine the class probabilities for each pixel. The LinkNet architecture used for this study is as shown in Figure 8.

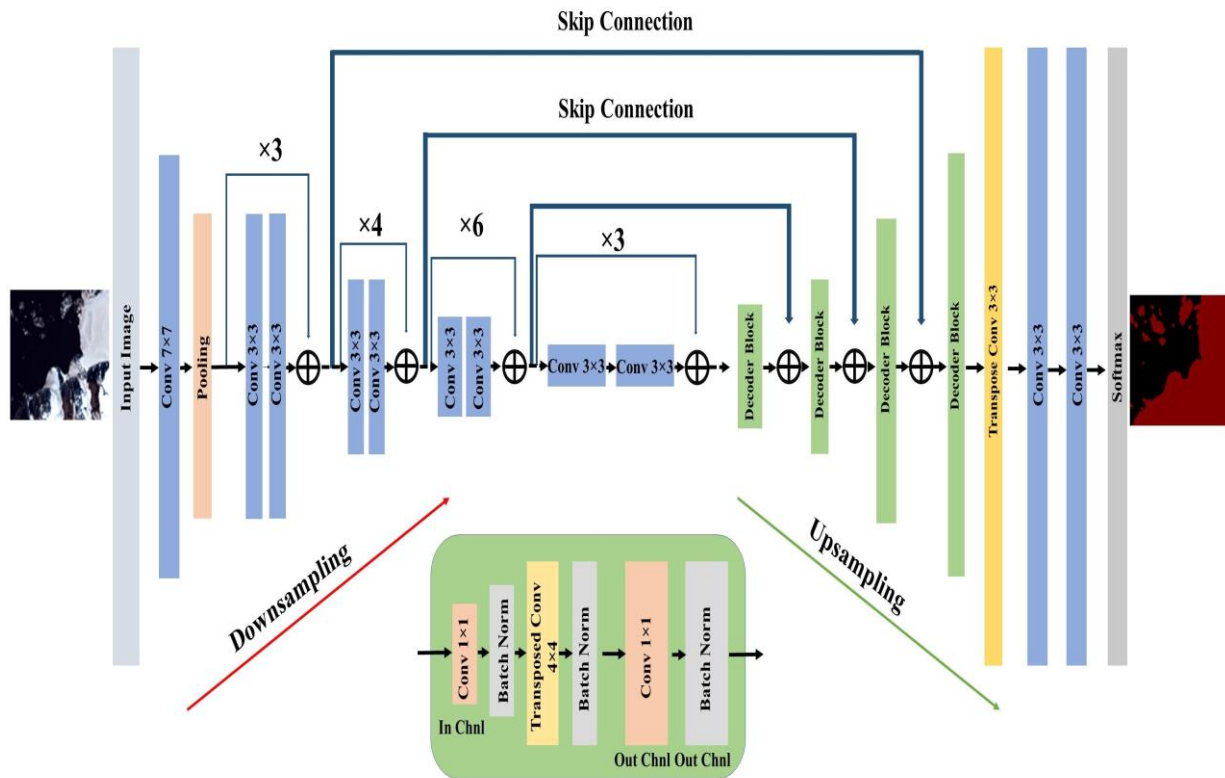


Figure 8. LinkNet architecture

#### 2.2.4. DeepLabV3+

DeepLab [12] architecture proposed by researchers at Google and has several variants. DeepLabV2 [30] model was presented by improving the encoder path of the DeepLab model. Conditional Random Field (CRF) was removed in DeepLabV3. Instead, Atrous Spatial Pyramid Pooling (ASPP) was implemented and DeepLabV3 addresses the problem of segmenting objects at different scales by using modules that capture multi-scale context through various atrous rates. Atrous convolutions, applied either sequentially or in parallel, capture multi-scale context. End-to-end framework restriction has been removed. DeepLabV3+ [13] is the latest version in this series.

The encoder path of DeepLabV3+ down samples the input image using Xception (inception with deep separable convolutional network) convolutional layers and atrous (dilated) convolutions. Atrous convolutions provide a large field of view while maintaining the resolution of feature maps, allowing the model to capture

more contextual information. A powerful CNN model Xception is used as the backbone. This robust backbone extracts detailed and meaningful features from the input image.

In the Decoder section, it is combined with the feature map of the same resolution coming from the encoder section, together with the up sampling. The decoder processes the feature maps with the ASPP module, capturing information at multiple scales. ASPP combines multiple atrous convolution layers with different dilation rates and uses a global average pooling layer. This structure allows the model to understand both broad and narrow contexts.

In DeepLabV3+, the use of up sampling and atrous convolutions at different rates in the ASPP module ensures the model captures both low and high-resolution feature maps details. Thus, the DeepLabV3+ making it ideal for tasks requiring high-resolution and detailed segmentation. The DeepLabV3+ architecture used for this study is as shown in Figure 9.

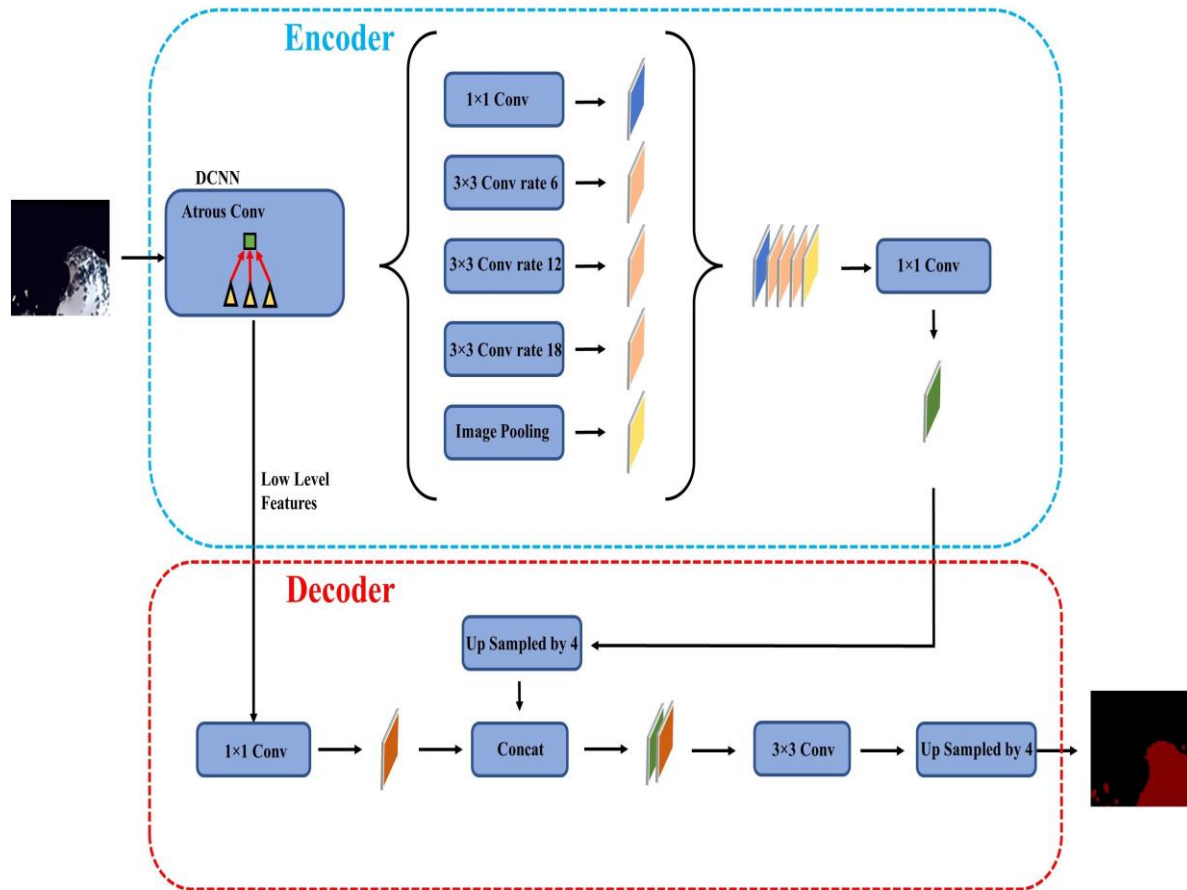


Figure 9. DeepLabV3+ architecture

### 2.2.5. Hyperparameters

Training for the U-Net, LinkNet, and DeepLabV3+ segmentation models has initiated using PyTorch. To enable effective comparison, the same hyperparameters are applied across all models. Training is performed on an Nvidia GeForce 3060 graphics card, with a learning rate of 0.00005 and the ADAM optimizer for optimization. DiceLoss is used as the loss function, while the IoU metric evaluates model performance throughout the training process.

### 2.2.6. Evaluation Criteria

#### 2.2.6.1. Confusion Matrix

A Confusion Matrix is a table that evaluates classification model performance by visualizing correct and incorrect predictions. It includes True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. TP counts correctly predicted positive instances, TN counts correctly predicted negatives, FP indicates false alarms, and FN shows missed positives. These components help assess accuracy, precision, specificity, and overall model performance, offering insights into its reliability in real-world applications.

#### 2.2.6.2. Intersection over Union

The IoU score is commonly used to evaluate model performance in segmentation and object detection tasks. It is obtained as given with Eq 7.

$$IoU = \frac{TP}{TP+FP+FN} \quad (7)$$

#### 2.2.6.3. Dice Loss and Dice Coefficient

The Dice Coefficient measures the overlap between the predicted segmentation region and the ground truth segmentation region.

$$Dice\ Coefficient = \frac{2 \times TP}{(TP+FP)+(TP+FN)} \quad (8)$$

Dice Loss measures how well the model predictions overlap with the ground truth values and encourages the model to perform accurate segmentation during the training process.

$$Dice\ Loss = 1 - Dice\ Coefficient \quad (9)$$

#### 2.2.6.4. Precision

Precision measures how many of the pixels predicted as positive by the model are actually positive.

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

#### 2.2.6.5. Recall

Recall measures how many of the actual positive pixels the model correctly predicts.

$$Recall = \frac{TP}{TP+FN} \quad (11)$$



### 2.2.6.6. Accuracy

Accuracy measures the model's rate of correctly predicting all pixels. It takes into account both true positives and true negatives.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

A high Accuracy indicates that the model generally makes correct predictions; however, it may not fully reflect the balance between positive and negative classes.

### 2.2.6.7. Specificity

Specificity measures how many of the pixels predicted as negative by the model are actually negative. It aims to reduce the rate of false positives.

$$Specificity = \frac{TN}{TN+FP} \quad (13)$$

### 2.2.6.8. F1 Score

The F1 Score is the harmonic mean of Precision and Recall and is used to measure the overall performance of the model.

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

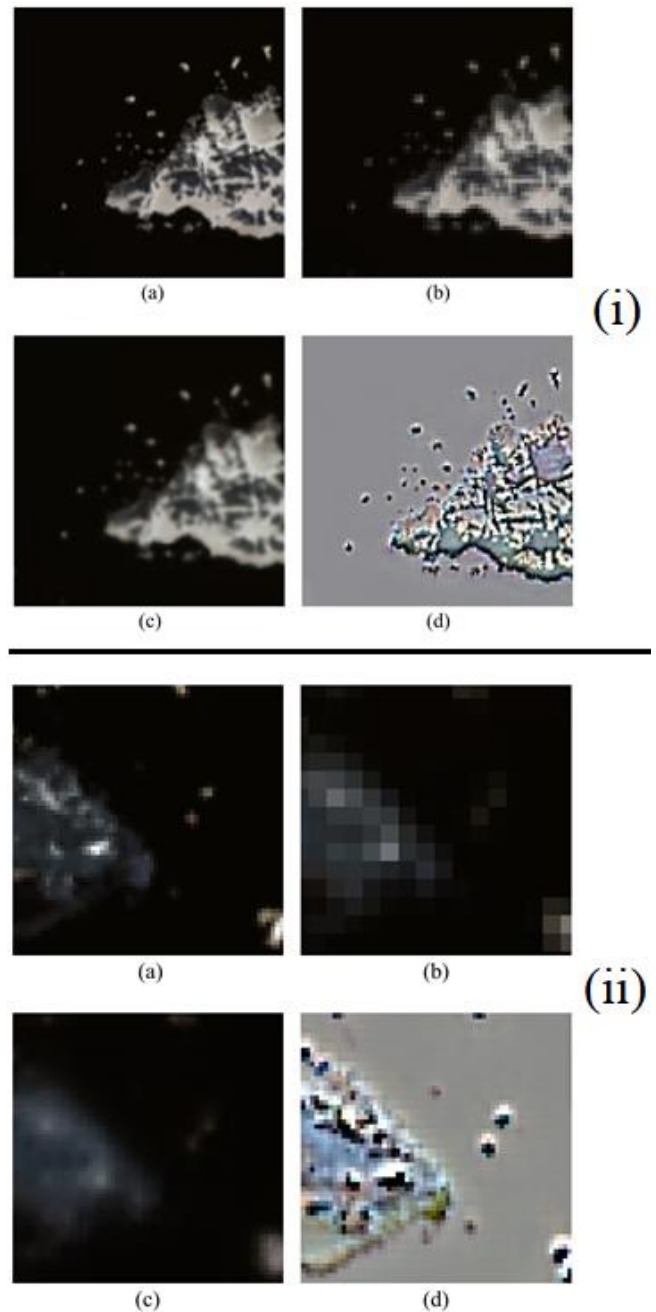
A high F1 Score indicates that the model has both high Precision and high Recall values. This reflects the model's overall performance in capturing true positives and minimizing false positives.

## 3. Results

In this study, the performance and effectiveness of the applied models and techniques were systematically evaluated. The findings obtained in this process revealed the strengths and weaknesses of the models. Satellite images were enhanced with the SRGAN model and U-net, LinkNet, DeepLabV3+ segmentation models were tested with images of different resolutions. Test results showed that segmentation success increases as the resolution increases. However, due to the advantages and disadvantages of the models, the success of the models varies as the resolution changes.

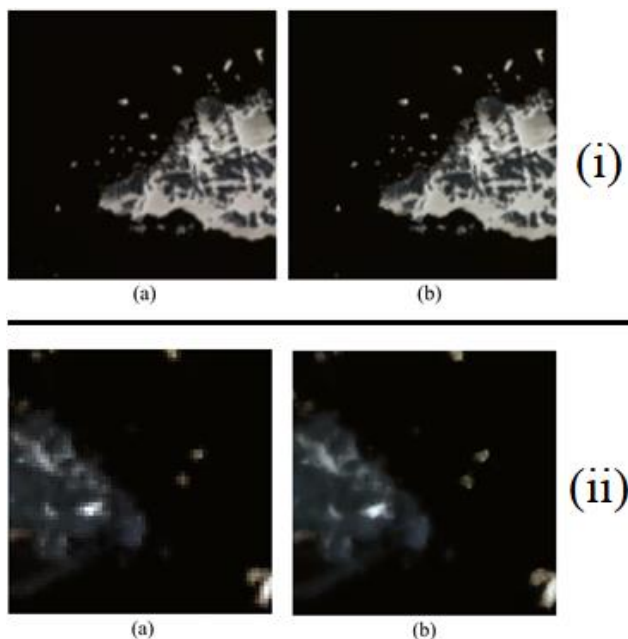
According to the results, low resolution, super resolution and GT-1 satellite images. U-net and LinkNet models achieved high scores on various metrics.

Figure 10 (i) shows the comparison of the results of the SR technique in reduced resolution domain for scale factor 2. The difference image shows that the discrepancy between the remote sensing image and the super-resolved image is minimal. Figure 10(ii) shows the comparison of the results of the SR technique in reduced resolution domain for a scale factor of 4. As mentioned in the method section, improvement for a scale factor of 4 is more challenging. Nevertheless, the obtained super-resolved image demonstrates that structure and details can be recovered using SR. The difference image shows a high similarity in the sea areas and coastline. This is a significant result for coastline segmentation.



**Figure 10.** Remote sensing image (a), down sampled image (b), super-resolved image of down sampled Image (c), difference between remote sensing image and super-resolved image (d); for (i) 2x and (ii) 4x scale factor.

Figure 11 (i) shows a remote sensing image alongside its super-resolved version at a scale factor of 2. The super-resolved image retains the structure of the original while enhancing details. Figure 11 (ii) presents the remote sensing image and its super-resolved counterpart at a scale factor of 4, highlighting the SRGAN model's effectiveness in improving image quality. Super-resolution clearly enhances coastal features, making inland snow masses more distinguishable and improving the recognition of environmental conditions.



**Figure 11.** Original remote sensing image (a), super-resolved image (b); for (i) 2x and (ii)4x scale factor.

Table 1 provides a comprehensive evaluation of the SRGAN model and histogram equalization applied in this study for a scale factor of 2. The results demonstrate that the SRGAN model can produce highly correlated images. Enhancing the down sampled image with the SRGAN model yielded impressive results, achieving a PSNR value of 30.35 and an SSIM value of 0.93, indicating a high degree of similarity. Histogram equalization further increased the PSNR value from 30.35 to 31.56 and the SSIM value from 0.93 to 0.94, suggesting an improvement in similarity. Similar results were obtained when the satellite image was tested with the model, with an SSIM value of 0.99 indicating that the enhancement of the satellite image's resolution was performed with minimal loss of information.

**Table 1.** Remote sensing image (a), super-resolved image of down sampled image (b), super-resolved image of the down sampled image and histogram equalization (c), super-resolved image of the remote sensing image (d), super-resolved image of the remote sensing image and histogram equalization (e) for scale factor 2.

×2	MSE	PSNR	SSIM	CS	CC	ERGAS
(a)/(b)	0.0020	30.355	0.930	0.994	0.991	17.213
(a)/(c)	0.0015	31.564	0.945	0.995	0.993	14.938
(a)/(d)	0.0006	35.053	0.979	0.998	0.997	10.073
(a)/(e)	0.0001	41.635	0.990	0.999	0.999	4.7084

Table 2 presents a comprehensive evaluation of the SRGAN model and histogram equalization applied in this study for scale factor 4. The results obtained for scale factor 4 fall behind those for scale factor 2. This is expected, as the difficulty increases with a higher scale factor. However, a high correlation was still observed. With histogram equalization, despite the high scale factor, high similarity results were achieved, with a CC of 0.98 and an SSIM of 0.84. By comparing the satellite image with the super-resolved satellite image with

histogram equalization, PSNR of 34.37 and SSIM of 0.978 values were obtained. This demonstrated that the model could enhance satellite images without distorting the texture and structure.

**Table 2.** Remote sensing image (a), super-resolved image of down sampled image (b), super-resolved image of the down sampled image and histogram equalization (c), super-resolved image of the remote sensing image (d), super-resolved image of the remote sensing image and histogram equalization (e) for scale factor 4.

×4	MSE	PSNR	SSIM	CS	CC	ERGAS
(a)/(b)	0.0045	26.794	0.814	0.987	0.978	25.894
(a)/(c)	0.0041	27.157	0.841	0.988	0.980	24.809
(a)/(d)	0.0007	34.684	0.972	0.997	0.996	10.504
(a)/(e)	0.0007	34.370	0.978	0.997	0.996	10.965

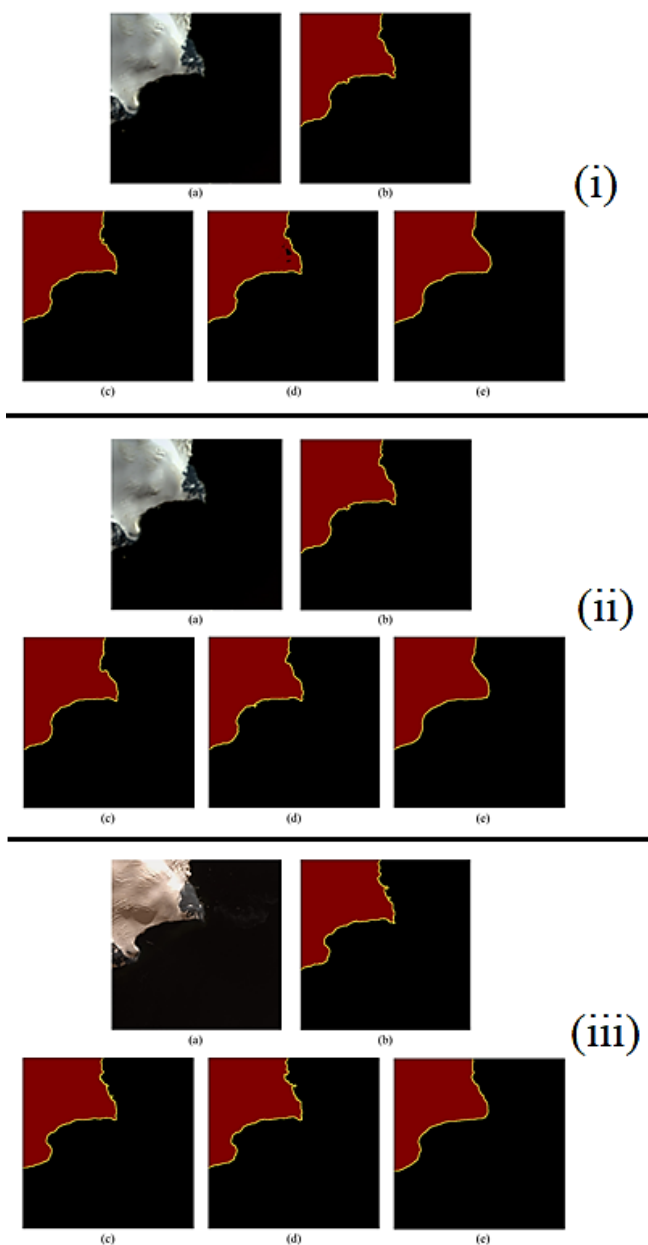
The data presented in Table 1 and Table 2 provide a comparative evaluation of the results of scale factor 2 and scale factor 4. The obtained results have been analyzed using different metrics in conjunction with image enhancement and histogram equalization methods. These data demonstrate that the enhancement and histogram equalization methods used in both scale factor 2 and scale factor 4 significantly improve image quality and have positive effects on the metrics. It is evident that the histogram equalization process yields better results at both levels of scaling. The results confirm that these methods can be effectively used in image processing field to achieve higher quality results.

Results of the segmentation experiments are evaluated both visually and quantitatively with use of metrics.

Figure 12(i) shows the LR image, its mask, and the segmentation results from the U-net, LinkNet, and DeepLabV3+ models. While all models accurately separate the coastline, LinkNet mislabels areas within the land due to environmental factors, likely because it focuses too much on fine details. LinkNet performs well in segmenting small regions like islands, while U-net is less affected by environmental conditions. DeepLabV3+ produces a shallower prediction than the others.

Figure 12(ii) presents the SR image, its mask, and segmentation results for the same models. The models' performance improves with higher resolution, particularly LinkNet, which resolves the land mislabeling issue. U-net also shows improvement, and while DeepLabV3+ doesn't visibly change between LR and SR images, evaluation metrics show better segmentation accuracy.

Figure 12(iii) shows the GT-1 image, its mask, and the models' segmentation results. Both U-net and LinkNet perform better with the high-resolution GT-1 image, closely aligning with the mask, while DeepLabV3+ remains inconsistent. Overall, U-net and LinkNet show clear superiority in segmenting high-resolution images.



**Figure 12.** Image (a), mask (b), Segmentation Results of U-net (c), LinkNet (d), DeepLabV3+ (e), for (i)LR, (ii) SR and (iii) GT-1 images.

The segmentation performance of DeepLabV3+, LinkNet, and U-net models for the LR, SR and GT-1 image was compared using the metrics in [31,32,33], and results were presented in Table 3. For LR image, DeepLabV3+ achieved the highest accuracy, with 0.973 IoU, 0.986 Dice, 0.989 Precision, 0.983 Recall, 0.997 Accuracy, 0.999 Specificity, and 0.9866 F1 Score. LinkNet ranked second with 0.967 IoU, 0.983 Dice, 0.967 Precision, 0.999 Recall, 0.997 Accuracy, 0.997 Specificity, and 0.983 F1 Score, excelling in Recall. U-net, while achieving the highest Recall, slightly trailed in other metrics, with 0.961 IoU, 0.980 Dice, 0.961 Precision, 1.0 Recall, 0.997 Accuracy, 0.996 Specificity, and 0.980 F1 Score. Overall, DeepLabV3+ provided the best results, with LinkNet and U-net also performing well, showing that these models are effective for high-accuracy coastline segmentation from satellite images.

For the SR image, LinkNet achieved the highest accuracy with 0.983 IoU, 0.991 Dice, 0.983 Precision, 0.999 Recall, 0.998 Accuracy, 0.998 Specificity, and 0.991 F1 Score. DeepLabV3+ followed with 0.978 IoU, 0.989 Dice, 0.987 Precision, 0.990 Recall, 0.998 Accuracy, 0.998 Specificity, and 0.989 F1 Score, showing a balanced performance. U-net, with 0.973 IoU, 0.986 Dice, 0.973 Precision, 0.999 Recall, 0.997 Accuracy, 0.997 Specificity, and 0.986 F1 Score, performed well but slightly lagged behind the other models. LinkNet benefited the most from SR, showing improved performance over LR.

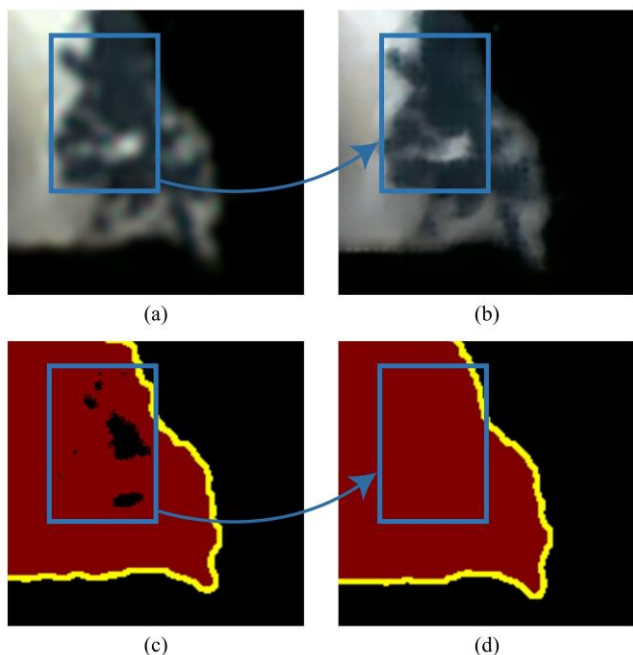
For the GT-1 image, U-net achieved the best results with 0.985 IoU, 0.992 Dice, 0.991 Precision, 0.994 Recall, 0.998 Accuracy, 0.999 Specificity, and 0.992 F1 Score. LinkNet was close behind with 0.984 IoU, 0.992 Dice, 0.985 Precision, 0.999 Recall, 0.998 Accuracy, 0.998 Specificity, and 0.992 F1 Score. DeepLabV3+, with 0.967 IoU, 0.983 Dice, 0.985 Precision, 0.981 Recall, 0.997 Accuracy, 0.998 Specificity, and 0.983 F1 Score, lagged slightly but still performed well. U-net led with the highest resolution, with LinkNet close behind, while DeepLabV3+ fell behind both.

**Table 3.** Metric based evaluation results of segmentation models across LR, SR and GT-1 images.

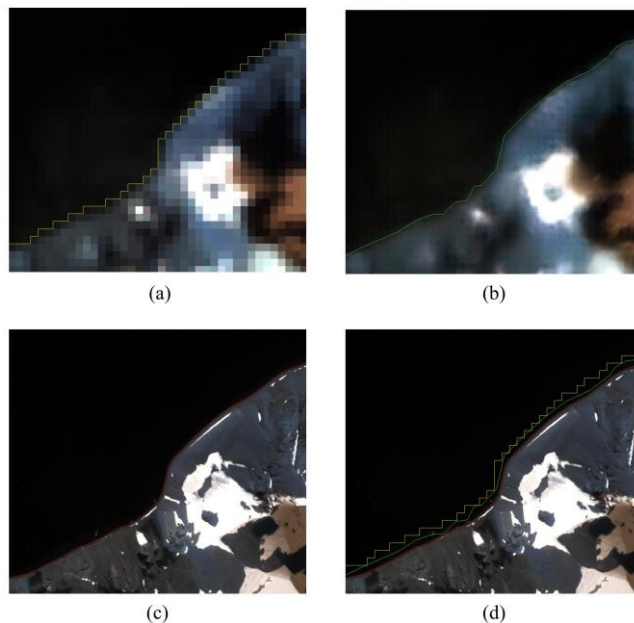
LR	IoU	DiceCoeff.	Precision	Recall	Accuracy	Specificity	F1-Score
U-net	0,961	0,980	0,961	<b>1</b>	0,996	0,996	0,980
LinkNet	0,967	0,983	0,967	0,999	<b>0,997</b>	0,997	0,983
DeepLabv3+	<b>0,973</b>	<b>0,986</b>	<b>0,989</b>	0,983	<b>0,997</b>	<b>0,999</b>	<b>0,986</b>
SR	IoU	DiceCoeff.	Precision	Recall	Accuracy	Specificity	F1-Score
U-net	0,973	0,986	0,973	<b>0,999</b>	0,997	0,997	0,986
LinkNet	<b>0,983</b>	<b>0,991</b>	0,983	<b>0,999</b>	<b>0,998</b>	<b>0,998</b>	<b>0,991</b>
DeepLabv3+	0,978	0,989	<b>0,987</b>	0,990	0,998	0,998	0,989
GT-1	IoU	DiceCoeff.	Precision	Recall	Accuracy	Specificity	F1-Score
U-net	<b>0,985</b>	<b>0,992</b>	<b>0,991</b>	0,994	<b>0,998</b>	<b>0,999</b>	<b>0,992</b>
LinkNet	0,984	<b>0,992</b>	0,985	<b>0,998</b>	<b>0,998</b>	0,998	<b>0,992</b>
DeepLabv3+	0,967	0,983	0,985	0,981	0,997	0,998	0,983

Overall, the segmentation results from GT-1 show the highest accuracy, followed by the SR segmentation results, and then the LR segmentation results. These findings suggest that GT-1 provides the most accurate segmentation. The SR results also demonstrate high performance, but the LR results exhibit slightly lower performance. As illustrated in Figure 13, the area misclassified in the LR image is accurately predicted in the SR image by the same model (LinkNet), indicating an improvement in segmentation success with SR.

Resolution is an important factor for the segmentation of images. The effects of this factor can be observed by using the high-resolution GT-1 satellite image as a reference for the same region. By comparing the GT-1 satellite image (Assumed Ground Truth) with the Landsat-8 satellite image and the super-resolved image of Landsat-8 image, it is evident that as the resolution increases, accurate, smoother and more realistic coastline can be delineated as shown in Figure 14.



**Figure 13.** Super Resolution effect on LinkNet model, LR image (a), SR image (b), results of LinkNet segmentation model with LR image (c) and SR image (d).



**Figure 14.** Coastline (yellow) of LR image (a), coastline (green) of SR image (b), coastline (red) of GT-1 image (c), all coastlines on GT-1 image (d).

The comparison of coastline lengths obtained from different models and for all dataset is provided in Table 6. For relative error analysis, a manual digitization was applied on GT-1 image to act as ground truth. The coastal length for the region in accuracy analysis was 3412.00 m according to this digitization. According to the results, although the U-net and DeepLabV3+ models produce successful results in segmentation process according the metrics, it was observed that the LinkNet model determines the coastline length closest to the ground truth value. In particular, by segmenting the images improved with the SRGAN model with the LinkNet model, the closest coastline length to the real one could be calculated with Landsat-8 and Sentinel-2 satellite images.

**Table 4.** Segmentation Models vs coastal length.

	Coastal Distance (m)			Relative Error (%)		
	U-Net	Link Net	Deep labv3+	U-Net	Link Net	Deep labv3+
LR	3,126.67	3,163.38	2,951.16	8.36	7.29	13.51
SR	3,157.98	3,196.15	2,991.42	7.44	6.33	12.33
GT1	3,187.21	3,388.66	3,226.27	6.59	0.68	5.44

#### 4. Discussion

The SRGAN model has been adapted to work with 32-bit satellite images for Super Resolution. The obtained results show that resolution enhancement at 2x and 4x scales has been successful for Landsat-8 and Sentinel-2 images. The main factors contributing to this outcome are preventing data loss by adjusting the number of bands and maintaining the data type required by the algorithms for the model's operation. Various tests have indicated that preserving the characteristics of original data is crucial. Additionally, matching the histograms of the obtained high-resolution images with the original images has increased the accuracy of the

analysis. Providing a homogeneously distributed data enhances the accuracy of the training. It is important to select the images considering the different characteristics of the study area and to test the training based on these differences can be satisfied this way.

Segmentation model tests have been repeated for different images and regions. According to the obtained results, the DeepLabV3+ model is successful in regional predictions but lags behind the U-net and LinkNet models in capturing details. The U-net model is generally successful but falls behind LinkNet in accurately calculating the actual coastline length. The LinkNet model has been found to be more successful in segmenting details. With this advantage, the model can better segment small land areas, although its detailed segmentation feature may lead to incorrect predictions due to environmental factors such as snow and shadow within the land. However, this problem can be overcome with SR techniques. It can be stated that, the combination of SRGAN and the LinkNet model can calculate values closest to the actual coastline length while preserving coastline details according to outputs of this study.

## 5. Conclusion

The integration of SR and advanced segmentation techniques is highly effective in enhancing the resolution of satellite images and accurately delineating coastline features. The challenges posed by low-resolution images can be overcome with SR. It is evident that the performance of segmentation models is directly proportional to resolution. In this study, it was observed that the SRGAN model enhances details and improves segmentation performance. By applying SR to Landsat-8 and Sentinel-2 satellite images, the details of the coastline were enhanced, and the adverse effects of environmental factors on segmentation were mitigated. It was observed that the LinkNet segmentation model is the most effective in segmenting details. These results demonstrate that relatively low-resolution images obtained at low cost can achieve near-real values when combined with the SRGAN model and LinkNet segmentation model.

## Acknowledgement

This study was funded by the Scientific and Technological Research Council of Turkey (TUBITAK), 1001 program, Project No: 121Y366

## Author contributions

**Ilhan PALA:** Data curation, Writing, Data Processing, Data Analysis, Literature Review, Visualization, Software, Methodology.

**Ugur ALGANCI:** Conceptualization, Methodology, Results Interpretation, Critical Revision, Guidance, Advisor.

## Conflicts of interest

The authors declare no conflicts of interest.

## References

- Saleem, A., Mahmood, S. (2023). Spatio-temporal assessment of urban growth using multi-stage satellite imageries in Faisalabad, Pakistan. *Advanced Remote Sensing*, 3(1), 10–18.
- Zadbagher, E., Marangoz, A.M, Becek, K. (2023). Characterizing and estimating forest structure using active remote sensing: An overview. *Advanced Remote Sensing*, 3(1), 38–46.
- Efe, E., Algancı, U. (2023). Çok zamanlı Sentinel 2 uydu görüntüleri ve makine öğrenmesi tabanlı algoritmalar ile arazi örtüsü değişiminin belirlenmesi. *Geomatik*, 8(1), 27-34.
- Yiğit, A.Y., Şenol, H.İ., Kaya, Y. (2022). Çok zamanlı multispektral uydu verilerinin Marmara Gölü kıyı değişimi analizinde kullanılması. *Geomatik*. 2022, 7(3), 253-260.
- Bakırman, T., Sertel, E. (2023). A benchmark dataset for deep learning-based airplane detection: HRPlanes. *International Journal of Engineering and Geosciences*, 8(3), 212-223.
- Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative Adversarial Networks (Version 1). arXiv, <https://doi.org/10.48550/ARXIV.1406.2661>
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2016). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network (Version 5). arXiv, <https://doi.org/10.48550/ARXIV.1609.04802>
- Salgueiro Romero, L., Marcello, J., Vilaplana, V. (2020). Super-Resolution of Sentinel-2 Imagery Using Generative Adversarial Networks. *Remote Sensing*, 12(15), 2424.
- Xiong, Y., Guo, S., Chen, J., Deng, X., Sun, L., Zheng, X., Xu, W. (2020). Improved SRGAN for Remote Sensing Image Super-Resolution Across Locations and Sensors. *Remote Sensing*, 12(8), 1263.
- Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234-241, [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Chaurasia, A., Culurciello, E. (2017). LinkNet: Exploiting encoder representations for efficient semantic segmentation. *IEEE Visual Communications and Image Processing (VCIP)*, St. Petersburg, FL, USA, 1- 4. <https://doi.org/10.1109/vcip.2017.8305148>
- Chen, L-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L. (2014). Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs (Version 4). arXiv, <https://doi.org/10.48550/ARXIV.1412.7062>
- Chen, L-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation (Version 3). arXiv, <https://doi.org/10.48550/ARXIV.1802.02611>
- Alsabhan, W., Alotaiby, T. (2022). Automatic Building Extraction on Satellite Images Using Unet and

- ResNet50. Computational Intelligence and Neuroscience, (1), 5008854.
15. Sariturk, B., Seker, D.Z., Ozturk, O., Bayram, B. (2022). Performance evaluation of shallow and deep CNN architectures on building segmentation from high-resolution images. *Earth Science Informatics*, 15, 1801–1823.
  16. Zhang, Z., Liu, Q., Wang, Y. (2018). Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749-753.
  17. Han, J., Wang, Z., Wang, Y., Hou, W. (2022). Building extraction algorithm from remote sensing images based on improved DeepLabv3+ network. *Journal of Physics: Conference Series*, 2303, 012010.
  18. Kaya, Y., Şenol, H. İ., Yiğit, A. Y., & Yakar, M. (2023). Car detection from very high-resolution UAV images using deep learning algorithms. *Photogrammetric Engineering & Remote Sensing*, 89(2), 117-123.
  19. Şenol, H. İ., Kaya, Y., Yiğit, A. Y., & Yakar, M. (2024). Extraction and geospatial analysis of the Hersek Lagoon shoreline with Sentinel-2 satellite data. *Survey Review*, 56(397), 367-382.
  20. Yang, T., Jiang, S., Hong, Z., Zhang, Y., Han, Y., Zhou, R., ... & Kuc, T. (2020). Sea-Land Segmentation Using Deep Learning Techniques for Landsat-8 OLI Imagery. *Marine Geodesy*, 43(2), 105-133.
  21. Panuntun, I. A., Chen, Y-N., Jamaluddin, I., Tran, T. L. C. (2024). Evaluation of Deep Learning Semantic Segmentation for Land Cover Mapping on Multispectral, Hyperspectral and High Spatial Aerial Imagery. arXiv, <https://doi.org/10.48550/ARXIV.2406.14220>
  22. Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481 - 2495.
  23. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J. (2016). Pyramid Scene Parsing Network (Version 2). arXiv, <https://doi.org/10.48550/ARXIV.1612.01105>
  24. Jégou, S., Drozdal, M., Vazquez, D., Romero, A., Bengio, Y. (2016). The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation (Version 3), arXiv, <https://doi.org/10.48550/ARXIV.1611.09326>
  25. Lin, G., Milan, A., Shen, C., Reid, I. (2016). RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation (Version 3). arXiv, <https://doi.org/10.48550/ARXIV.1611.06612>
  26. Lin, T-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. (2016). Feature Pyramid Networks for Object Detection (Version 2). arXiv, <https://doi.org/10.48550/ARXIV.1612.03144>
  27. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 770-778
  28. He, K., Zhang, X., Ren, S., Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification (Version 1). arXiv, <https://doi.org/10.48550/ARXIV.1502.01852>
  29. Simonyan, K., Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition (Version 6). arXiv, <https://doi.org/10.48550/ARXIV.1409.1556>
  30. Chen, L-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(10), 834 - 848.
  31. Akar, Ö., Saralioğlu, E., Güngör, O., Bayata, H. F. (2024). Semantic segmentation of very-high spatial resolution satellite images: A comparative analysis of 3D-CNN and traditional machine learning algorithms for automatic vineyard detection. *International Journal of Engineering and Geosciences*, 9(1), 12-24.
  32. Mogaraju, J. K. (2024). Machine learning assisted prediction of land surface temperature (LST) based on major air pollutants over the Annamayya District of India. *International Journal of Engineering and Geosciences*, 9(2), 233-246.
  33. Çetin, Ş. B. (2023). Real-ESRGAN: A deep learning approach for general image restoration and its application to aerial images. *Advanced Remote Sensing*, 3(2), 90–99.



© Author(s) 2025. This work is distributed under <https://creativecommons.org/licenses/by-sa/4.0/>