International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://www.theaspd.com/ijes.php

Development Of An Intelligent Analysis, Search, And Recognition Method For Anthropogenic Target Objects In Remote Sensing Data Using Generative Adversarial Networks

P.A. Gusev
Student, Izhevsk State Technical University named after M.T. Kalashnikov st. Studencheskaya, 7, Udmurt Republic, Izhevsk, 426069, Russia

Abstract
This paper presents an algorithm for the semantic segmentation of image regions containing anthropogenic objects. The method is based on the use of generative adversarial networks (GANs). The proposed detector operates by identifying a specific spectral artifact that arises when image resolution is enhanced using a generative adversarial network.
The study analyzes a spectral neural network-based detector that relies on detecting characteristic spectral artifacts produced by the resolution enhancement module of generative neural networks. This detector was chosen for investigation because its developers claim high efficiency in analyzing various types of images, including the detection of artificially generated images obtained via remote sensing (RS). However, previous research has only tested this detector on fully synthesized images. It remains unclear how effective it would be in scenarios where only specific regions of RS images require refinement. Additionally, prior studies utilized only a single type of satellite imagery, and the generator set did not account for adversarial networks with a fixed long-element generator.
This research focuses on studying semantic analysis algorithms for Earth observation satellite imagery. Currently, numerous image segmentation methods exist, and the development of algorithms for processing remote sensing data continues. Some of these methods are based on generative adversarial neural networks.
The proposed algorithm is designed for the segmentation of remote sensing images of varying types and resolutions. The processed image fragments differ in size, shape, and may be non-contiguous.
Keywords: *Image segmentation using neural networks, generative adversarial networks.*

The results demonstrate the effectiveness of the proposed spectral neural detector in identifying segmented areas in remote sensing (RS) images. Image segmentation is a necessary step before extracting target objects. To extract texture features, the empirical mode decomposition method is applied, followed by Hilbert spectral analysis [1]. The two-dimensional transformation allows efficient decomposition of the image into frequency components using only the frequencies present in it. The obtained analytical functions undergo transformation to determine the instantaneous frequencies and amplitudes used as texture characteristics. Approximation with a radial basis function was employed to construct the envelopes.

In [2], a method for forming an ensemble of hierarchical grid clustering algorithms for the segmentation of multispectral satellite images is presented. The study [2] proposes a way to create a combined hierarchical grid-based clustering algorithm for multispectral satellite image segmentation. This approach is based on an algorithm that combines the strengths of grid-based and density-based methods, ensuring high computational efficiency and the ability to detect clusters of complex shapes [3]. The ensemble algorithm has a hierarchical structure in which the final decision is made using a standard agglomerative dendrogram construction method with complete linkage, utilizing pairwise object distances as input data. The ensemble approach improves the quality of clustering and enhances its robustness to variations in grid parameters.

In [4], a method for detecting extended geospatial objects in satellite images is introduced, based on searching for areas that correspond to the spatial characteristics of target-layer objects [5]. The

road network localization algorithm employs a finite impulse response (FIR) filter to approximate the impulse response, which helps determine the pixel membership of road sections. The digital filter is characterized by the following transfer function:

$$H(z) = h_0 + h_1 z^{-1} + \cdots + h_{N-1} z^{-(N-1)} = \sum_{k=0}^{N-1} h(k) z^{-k}$$

The method identifies extended geospatial objects by sequentially traversing the image in order of increasing coordinates and verifying the points for their membership in the target layer. Spectral separation, such as matched filtering, is one of the methods for rapidly extracting target objects from large volumes of remote sensing (RS) data. Matched filters enhance the contrast of the target object relative to the background [6].

The processing result represents the percentage content of the desired reference object in the image (ranging from 0 to 1). Another approach to improving the distinction between the background and target objects is the principal component analysis (PCA) method, which performs image decorrelation.

This study explores algorithms for the semantic segmentation of RS images. To date, numerous image segmentation algorithms have been developed, and research continues into methods for segmenting remote sensing imagery. Some of these approaches are based on generative adversarial networks (GANs) in various ways.

Currently, convolutional neural network (CNN) technologies have demonstrated strong performance in wildfire detection due to their ability to automatically extract data from images in real time. A pre-trained object detection neural network, ReNet, enables the rapid identification of wildfires and their components. Another emerging application of neural networks for the automated processing of RS data is the prediction of wildfire consequences and the assessment of forested areas affected by fires.

Satellite platforms capture the same territory multiple times per day, enabling the monitoring of wildfire outbreaks and their progression. Unlike large-scale active wildfires, which are measured in hectares of forested land, maximizing the efficiency of fire detection requires defining key selection criteria for satellite data sources:

- The data source must be equipped with infrared (IR) imaging capabilities.
- The spatial resolution of the selected data source must be as close as possible to the minimum threshold required for fire hotspot identification (not exceeding 100 m^2).
- To maintain data relevance, the potential update frequency should be no less than twice per day.

For the subsequent analysis, satellite data from the Kanopus-V series were used, as they met the aforementioned criteria. The images are provided in GeoTIFF format with a resolution of 7600×7600 pixels and 16 bits per pixel per channel. Each multispectral pixel corresponds to a 30-meter spatial resolution.

The application of deep learning methods for assessing the Earth's surface condition based on satellite observation data is becoming an actively developing research area. The primary challenges in implementing this approach involve determining the neural network architecture most effective for identifying the necessary artifacts in the analyzed images and finding large-scale datasets for its training.

Compared to classical neural networks, convolutional neural network (CNN) architectures are more suitable for satellite data analysis, as they represent one of the most accessible and well-developed solutions. CNNs take into account the spatial structure of images by analyzing neighboring pixels not only horizontally but also vertically. Additionally, they incorporate dimensionality reduction techniques, enabling object recognition regardless of scale and position within the image, while also reducing the volume of input data through the addition of pooling layers.

One of the key advantages of CNNs is the use of shared weights: convolutional kernels apply the same weights to different regions of the image, significantly reducing the number of trainable parameters and, consequently, making the training process more efficient. This is why CNNs became one of the first deep learning architectures successfully trained using the backpropagation algorithm. CNNs have likely succeeded in areas where fully connected networks with backpropagation proved ineffective due to their computational efficiency. This advantage facilitates experimentation and hyperparameter tuning. As noted earlier, CNNs currently demonstrate the best results in image processing. Before proceeding to a description of CNN models, it is important to first understand what CNNs are and how they function.

The primary distinction between CNN architectures and classical neural networks lies in the presence of convolutional layers, which apply filters to perform image convolution, as well as pooling layers for subsampling and fully connected layers for final classification.

The image patches from the dataset were randomly split into three subsets for training, validation, and testing, comprising 40%, 10%, and 50% of the total patches, respectively. The selected neural network architecture was trained for 50 epochs, with an early stopping mechanism in case a specified performance metric ceased to improve. In our case, the goal of training was to minimize the loss function, which resulted in the training process stopping at epoch 37 out of 50.

To enhance computational efficiency, the Adam optimization algorithm for first-order stochastic objective functions was employed. The choice of this method was motivated by several advantages, including computational efficiency, low memory requirements, and invariance to diagonal rescaling of gradients.

An evaluation of the implemented architecture was conducted based on key performance metrics commonly used in semantic segmentation tasks. The results are presented in Table 1.

One of the most promising approaches to generating artificial images is the use of a generative adversarial network (GAN) with a dedicated generator for extended elements. This method relies on a specialized algorithm for supplementing such elements in images. Studies have demonstrated that this generator consistently achieves high performance in the creation of artificial remote sensing (RS) images. This paper addresses the problem of detecting artificially generated RS images using examples of images enhanced by the aforementioned generator. The continuous advancement of artificial RS image generation techniques has made the development of detection algorithms increasingly relevant.

Convolutional neural networks (CNNs) have demonstrated strong performance in wildfire detection due to their ability to automatically extract data from images in real time. The pre-trained ReNet network enables rapid detection of wildfires and their components. A particularly relevant research direction involves predicting wildfire consequences and assessing forested areas affected by fire.

Deep learning methods are actively evolving for the assessment of Earth's surface conditions based on satellite data. The primary challenges include selecting an appropriate neural network architecture and identifying large datasets for training. CNNs account for the spatial structure of images by analyzing neighboring pixels both horizontally and vertically, allowing for object recognition regardless of scale and position.

This study examines the detector in the context of identifying artificial images generated by a GAN incorporating a dedicated generator for extended elements. A distinctive feature of the analyzed images is the presence of pre-defined segments of various shapes and sizes, which replace specific regions of real RS images. Additionally, RS images from different sources with varying resolutions are considered.

Algorithms for Generating Artificial RS Images

Generative adversarial networks (GANs) play a crucial role in the creation of artificial RS images. In recent years, generative adversarial networks (GANs) have demonstrated superior performance in generating artificial images. A GAN consists of two neural subnetworks: a generator and a discriminator. The generator creates artificial images, which are then mixed with real data and

fed into the discriminator. The discriminator's objective is to distinguish artificial data from real data by classifying the input image into one of two categories: real or synthetic.

The success or failure of this classification process enables adjustments to the weight coefficients of both the generative and discriminative subnetworks. As a result, the generator learns to create increasingly realistic artificial images, while the discriminator improves its ability to detect synthetic data. This process leads to the mutual enhancement of both networks within the framework of unsupervised learning, which does not require a labeled training dataset.

In this study, GANs are employed as components of more complex algorithms designed for the generation and detection of artificial remote sensing (RS) images.

CycleGANs for Artificial Image Generation in Remote Sensing

Cycle-consistent generative adversarial networks (CycleGANs) are designed for image style transfer. Some of the most well-known applications of such networks include:

- Seasonal transformation in photographs (e.g., converting a summer scene into a winter scene),
- Recoloring objects (e.g., transforming a horse into a zebra),
- Converting photographs into paintings in the style of a specific artist.

A particularly notable application of CycleGANs is image compression, as these networks enable the generation of highly detailed, full-color digital images—which typically require significant storage—based on compact, low-detail, low-bit-depth sketches or maps.

In this study, a CycleGAN is utilized to convert digital map fragments into remote sensing (RS) image fragments, effectively generating artificial RS images based on digital maps. The training dataset consists of paired samples, where each digital map fragment is matched with a corresponding RS image fragment.

The CycleGAN architecture itself comprises two adversarial subnetworks. The first adversarial subnetwork is trained to generate RS image fragments from digital map fragments.

Adversarial Network with an Extended Element Generator for Artificial Remote Sensing Image Generation

Simultaneously, the second adversarial subnetwork solves the inverse problem: it learns to generate a digital map from remote sensing (RS) images. The output of this process consists of artificially generated RS images that correspond to a given digital map (see the Experimental Results section).

In this study, an adversarial neural network with a dedicated extended element generator was employed to refine and complete RS images for the purpose of generating artificial remote sensing imagery.

The specified generator comprises two adversarial subnetworks:

1. The extended element generator, responsible for producing elongated elements within incomplete regions of RS images.
2. The image refiner, which finalizes the generation process.

The input data for the extended element generator includes: a grayscale version of the incomplete RS image, a binary mask marking unfilled regions, and a binary image containing only the extended elements of the RS image.

The goal of the extended element generator is to synthesize realistic extended structures within missing areas of RS images. The generator produces an intermediate output—an image containing only the newly generated extended elements. This image, along with the color version of the incomplete RS image, is then passed to the second adversarial subnetwork, which performs final refinement of the RS image.

This adversarial network with an extended element generator was applied to various types of RS image fragmentswith different resolutions [7, 8]. The method was used to replace missing areas of various sizes and shapes, including disconnected regions.

Dataset Preparation and Neural Network Training

Both real RS images and the generator's output were compiled into Dataset No. 2, which included:

- 928 real RS images
- 928 artificially modified RS images containing synthetic fragments

Examples of the dataset are shown in Figures 2 and 3.

The spectral neural detector was trained on Dataset No. 1 and tested on Dataset No. 2. To prevent overfitting, Dataset No. 2 was divided into training and validation subsets:

- The training set consisted of 742 real RS images and 742 RS images with artificially inserted fragments.
- The validation set consisted of 186 real RS images and 186 RS images with modified areas.

This segmented dataset was utilized for a separate experiment aimed at training and evaluating the spectral neural detector.

The described adversarial neural network with a separate generator for extended elements is highly effective [9] in generating artificial RS images, as ensuring high-quality extended elements is especially crucial when solving this task.

Artifacts in the Output of Adversarial Neural Networks

Despite significant structural differences in various adversarial neural networks, most of these networks feature a resolution enhancement block. The presence of this block leads to the appearance of specific types of artifacts in the generated images [10].

Let us explain in more detail the mechanism of these artifacts. Any block of twofold resolution enhancement can, in the 1D case, be represented as two consecutive operations (the generalization to 2D is trivial).

Operation 1: Placing zeros between the samples of the original image *f*(*k*) of size *N* to obtain an image $g(k)$ of double size *2N*:

$$g(k) = \{f\left(\frac{k}{2}\right), \qquad k = 2m\ 0, \qquad k = \ 2m + 1$$

Operation 2: Interpolation of the zero samples of the image with odd coordinates $k = 2m + 1$ using some filter, usually a low-pass filter, to obtain the resulting higher-resolution image.

A frequency spectrum was used for training the classifier. It is easy to show that the first half of the spectrum {G(j), j < N} of the enhanced resolution image g(k) coincides with the spectrum F(j) of the low-resolution image f (j):

$$G(j) = \sum_{k=0}^{2N-1} g(k)exp\left(\frac{-2i\pi kj}{2N}\right) - 2i = \sum_{m=n}^{n-1} g(2m)exp\left(\frac{-2i\pi(2m)j}{2N}\right) =$$

$$\sum_{m=0}^{N-1} f(m)exp\left(\frac{-2i\pi mj}{N}\right) = F(j), j < \ N$$

Meanwhile, the second half of the spectrum G (j) coincides with the shifted spectrum of the image f (j):

$$G(j) = \sum_{m=0}^{N-1} f(m)exp\left(\frac{-2i\pi m(j - N + N)}{N}\right) =$$

$$= \sum_{m=n}^{n-1} f(m)exp\left(\frac{-2i\pi m(j - N)}{N}\right)exp\left(\frac{-2i\pi mN}{N}\right) =$$

$$= F(j - N), j \geq N$$

Thus, the spectrum of the high-resolution image contains two copies of the spectrum of the low-resolution image. In practice, it has been shown that the overall operation of the adversarial neural network #2 is not a full low-pass filter and, therefore, cannot completely remove the described artifacts. Moreover, these artifacts are sufficiently well detected by neural networks in the spectral domain.

A real image (I) is used as input and passed through the generator (G), which has a structure similar to the generator used in GANs for image generation. The decoder contains upsampling modules, such as transposed convolution or nearest-neighbor interpolation.

It is important to note that the only knowledge assumed to be available is the general architecture, but not the specific details (such as model weights and hyperparameters) of the GAN models used in the generation of fake images. Conceptually, this can be seen as solving a "gray-box" problem, as opposed to a "white-box" solution, where all details of the model are known, or a "black-box" solution, where no knowledge of the attack model exists. AutoGAN includes a discriminator (D) and an L_1 norm. Rather than making the output distribution from the generator resemble the distribution of images from another semantic category, as shown in the general image translation process (Fig. 1), the generator's output is compared with the original image. Formally, assuming there is a training set $\{I_1, \cdots, I_n\}$ containing n images, the final loss function L can be written as:

$$L = \sum_{i=1}^{n} \quad log(D(I_i)) + log(1 - D(G(I_i))) + \lambda \|I_i - G(I_i)\|$$

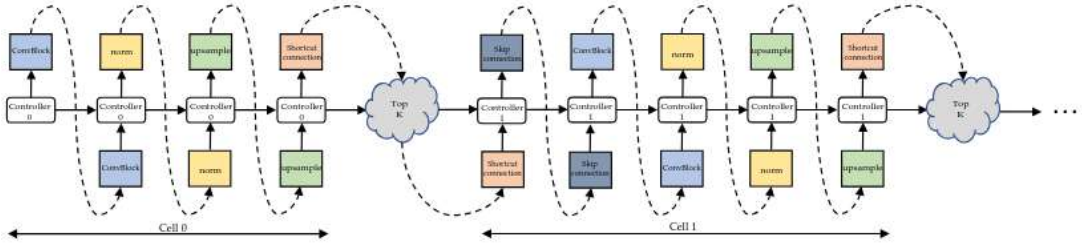

Fig. 1. The AutoGAN process.

In this study, ResNet-34, pre-trained on the ImageNet image dataset [10], was used. The last layer of this neural network was replaced with a layer with binary output (real or fake image), after which the resulting detector was further trained on remote sensing (RS) images, as described in more detail in the experimental section of this article.

For each of the three channels of the original color image, the amplitude spectrum G'_a was calculated (the phase part of the spectrum was not used). Spectral features were then computed by logarithmizing the spectra to equalize the amplitudes and normalizing them to the range [-1, 1]:

$$G'_a = \frac{ln\ lnG_a\ - min\{lnG_a\}}{max\{lnG_a\} - min\{lnG_a\}}$$

This spectral feature was then input into the artificial image detector based on the ResNet-34 neural network [11], which effectively addresses the issue of reduced accuracy when the number of layers in the neural network increases. Neural networks of this type include so-called "shortcut connections" that allow "skipping" one or more layers in a deep neural network.

As a result, ResNet-type neural networks are generally easier to scale and optimize compared to other deep neural networks.

The experimental study of algorithms for detecting fragments of RS images involves the use of a spectral neural network detector for detecting anthropogenic objects, performing the necessary number of experiments [12].

Using a cyclic adversarial neural network, we can create image datasets from different domains, such as CycleGAN [9], without direct correspondence between the images. The training is unsupervised, which allows for performing various tasks, such as photo enhancement, image colorization, and style transfer. To work with this network, you only need the source and target datasets in the form of image directories. The main principle of this technology is illustrated in the diagram, which demonstrates the process of creating new image datasets using a cyclic adversarial neural network.

This dataset contains fragments of a digital map and the corresponding fragments of remote sensing (RS) images [13]. The result of the generator's work was a set of artificial RS image fragments intended for training the detector. An example is shown in Fig. 1. The real RS images and the results of the generator's work formed Dataset No. 1, which includes 2194 real RS images and the same number of artificial (fully generated) ones.

Experimental Findings

The spectral neural network detector under consideration is capable of detecting artificial RS images with high accuracy, created by replacing parts of real RS images with those generated by the adversarial neural network with a generator of extended elements. The detector was trained not only to work with RS images of various types and resolutions but also to detect images where only part of the image was replaced with an artificial fragment [14].

Training the detector on RS images created by replacing parts of real RS images using the adversarial neural network with a separated generator of extended elements is meaningful as it significantly improves detection efficiency.

The spectral neural network detector under consideration is a promising algorithm for detecting artificial RS images, as it shows high results when counteracting a very effective forgery generator, particularly in the complex situation of partial information replacement in RS images [15].

Table 1. Neural Network Efficiency

	Dataset 1	Dataset 2
True Positive Rate	0.3485	0.9435
False Negative Rate	0.8527	0.8958

CONCLUSION

This paper addressed the issue of detecting anthropogenic target objects using remote sensing (RS) data. Special attention was given to the spectral artifact that arises in the works of adversarial image generators. A detector developed for detecting artificial RS images based on neural network technologies capable of identifying this artifact was analyzed. The detector was used to identify artificial RS images generated by an adversarial neural network that includes a generator of extended elements. The focus of the research was not on fully generated images but on scenes where specific areas of the image were replaced with generated data. The experiment applied real RS images of various types and resolutions, in which regions of different shapes and sizes were replaced with artificial information, including disconnected fragments. The experimental results demonstrated the effectiveness of the spectral neural network detector. These data confirm that the proposed detector is highly efficient in detecting artificial RS images, even in cases of partial information replacement.

REFERENCES

[1] Gao J, Yuan Q, Li J, Su X. Unsupervised missing information reconstruction for single remote sensing image with Deep Code Regression. Int J Appl Earth Obs Geoinf 2021; 105: 102599.

[2] Nazeri K, Ng E, Joseph T, Qureshi F, Ebrahimi M. EdgeConnect: Generative image inpainting with adversarial edge learning. arXiv preprint 2019. Source: https://arxiv.org/abs/1901.00212.

[3] Voronin V, Gapon N, Semenishchev E, Zelensky A, Agaian S. A block-based method for remote sensing images cloud detection and removal. Proc SPIE 2021; 11734: 117340K.

[4] Kuznetsov AV, Gashnikov MV. Remote sensing data retouching based on image inpainting algorithms in the forgery generation problem. Computer Optics 2020; 44(5): 763-771. DOI: 10.18287/2412-6179-CO-721.

[5] Songyuan L, Fan M, Chen R. Overview of generative adversarial networks. J Phys Conf Ser 2021; 1873(1): 12071.

[6] Songyuan L, Fan M, Chen R. Overview of generative adversarial networks. J Phys Conf Ser 2021; 1873(1): 012071.

[7] Roscosmos. Source: www.roscosmos.ru.
[8] Google Earth. Source: www.google.com/earth.
[9] Index of /~taesung_park/CycleGAN/datasets. Source: https://people.eecs.berkeley.edu/~taesung_park/CycleGAN/datasets/.
[10] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. IEEE Conf on Computer Vision and Pattern Recognition 2009:248-255.
[11] He F, Liu T, Tao D. Why resnet works? residuals generalize. IEEE Trans Neural Netw Learn Syst 2020; 31(12): 5349-5362.
[12] Zhang X, Karaman S, Chang SF. Detecting and simulating artifacts in GAN fake images. IEEE Int Workshop on Information Forensics and Security (WIFS) 2019: 1-6.
[13] Mallat S. A Wavelet Tour of Signal Processing. Academic Press, 1998. 550 p.
[14] Pratt W.K. Digital Image Processing. Hoboken, New Jersey, 2007. 807 pp.
[15] Xin M., Shuyan X., Guangxin L. Remote sensing image restoration with modulation transfer function compensation technology in-orbit. Proceedings of SPIE, 2013. Vol. 8768.