



Improving the accuracy of segmentation masks using a generative-adversarial network model

Igor Victorovich **Vinokurov**

Financial University under the Government of the Russian Federation, Moscow, Russia

 iguvinokurov@fa.ru

Abstract. Masks obtained using the deep learning model Mask R-CNN may in some cases contain fragmented contours, uneven boundaries, false fusions of adjacent objects, and areas with missed segmentation. The more detection objects in the image and the smaller their size, the more often various types of defects in their masks are encountered. Examples of such images include aerial photographs of cottage and garden associations and cooperatives characterized by high building density. To correct these defects, it is proposed to use a generative adversarial network model that performs post-processing of the predicted Mask R-CNN masks.

A qualitative assessment of the model formed in the work demonstrated that it is capable of restoring the integrity of contours at an acceptable level, filling in missing areas, and separating erroneously merged objects. Quantitative analysis using the IoU, precision, recall, and F1-score metrics showed a statistically significant improvement in the segmentation quality compared to the original Mask R-CNN masks. The obtained results confirmed that the proposed approach allows to increase the accuracy of the formation of object masks to a level that satisfies the requirements of their practical application in automated aerial photograph analysis systems. (*Linked article texts in English and in Russian*).

Key words and phrases: Computer vision, image segmentation, object masks, generative adversarial networks, Mask R-CNN, PyTorch

2020 *Mathematics Subject Classification:* 68T20; 68T07, 68T45

For citation: Igor V. Vinokurov. *Improving the accuracy of segmentation masks using a generative-adversarial network model*. Program Systems: Theory and Applications, 2025, **16**:2(65), pp. 111–152. (*In English, in Russian*). https://psta.psiras.ru/read/psta2025_2_111-152.pdf

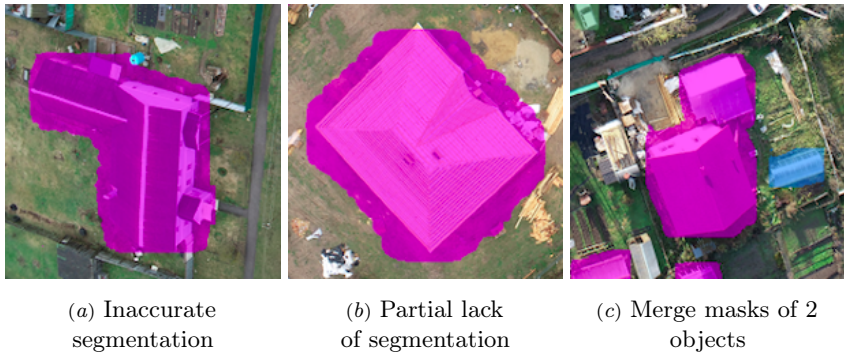


FIGURE 1. Errors in generated object masks

Introduction

Detecting correct property masks in dense aerial images is a challenge for deep learning models, Mask R-CNN, YOLO, and others. The main reason is that buildings are located close to each other, which makes it difficult to accurately separate objects, since their contours may intersect or merge in the image.

This is especially problematic in high-detail settings, where small architectural elements such as extensions, balconies, or common walls between buildings may be perceived by the model as a single object. In addition, complex building shapes, such as polygonal outlines or non-standard structures, require deeper analysis and adaptation from the model, which increases the likelihood of errors in mask generation. Several examples of such masks are shown in Figure 1.

Another reason is the presence of external factors, such as vegetation and shadows, which partially obscure objects in images. Trees, bushes, or even temporary objects such as cars can cover parts of buildings, creating gaps in the data available to the model. Shadows caused by different shooting angles or time of day also distort visual information, reducing contrast and making it difficult to distinguish clear boundaries of objects. These factors lead to the model either underestimating the area of an object, ignoring hidden areas, or overestimating it, including extraneous elements in the mask. As a result, there is an increased need for additional processing and improving the quality of predictions to ensure the correctness of the final data.

The problem of refining the boundaries of objects is especially relevant for the «Roskadastr» PLC information system (IS), where increased

requirements are placed on the accuracy of spatial data. Traditional post-processing methods, such as morphological operations or active contour algorithms, are often not effective enough for complex cases of dense development. In this regard, there is a need to solve the problem of automatic correction of segmentation results, while maintaining the topological correctness of objects.

The presented work proposes one of the possible solutions to this problem, based on the use of a generative adversarial network (GAN) for post-processing the segmentation results obtained using Mask R-CNN. The main idea of the study is that the GAN architecture, properly trained on a representative sample of development objects, typical for the «Roskadastr» PLC software package, will be able to effectively restore the integrity of broken boundaries, eliminate internal gaps in masks, separate erroneously merged objects and maintain the exact position of corner (turning) points of buildings.

The research implementation consisted of the following stages:

- (1) Formation of a representative and balanced dataset containing correspondences of the original masks from Mask R-CNN to reference masks extracted from files with object annotations and masks obtained as a result of their expert correction.
- (2) Development of a GAN architecture that takes into account the features of dense development in cottage and garden associations and cooperatives, requirements for the accuracy of boundaries for cadastral registration and the need to preserve the topology of objects.
- (3) Conducting experimental studies with an assessment of the qualitative characteristics of the resulting masks and quantitative indicators (IoU, precision, recall, F1-score).

1. Overview of work on using GAN to improve objects detection

There is currently a significant amount of research devoted to the use of generative adversarial networks (GANs) to improve the efficiency of image recognition and transformation.

The paper [3] discusses conditional GAN (cGAN) models that can be used to transform images of one type into another. For example, segmentation masks can be transformed into realistic images, or vice versa. The described models demonstrate high accuracy of object segmentation and improve the quality of their masks.

The Pix2Pix model described in [4] and based on cGAN is used to synthesize high-quality images based on input data such as segmentation masks. The paper shows that GAN can improve the quality of masks by training on (mask, exact mask) pairs.

The study [5] presents the MaskGAN model designed for interactive editing of face images using masks. The model improves the quality of masks and allows manipulation of segmented areas of images.

In [6], a SegAN model with a multi-level discriminator and L1 loss is proposed for medical segmentation. The model uses a multi-level discriminator to improve the quality of masks, especially in tasks with fuzzy object boundaries. The GAN model improves object boundaries, reduces noise, and eliminates object detection inaccuracies.

The attention mechanism for generating masks in Attention-GAN described in [7] is applied to transform objects in wildlife photographs. The model can eliminate noise and detection inaccuracies while preserving fine object details.

In [8], an image transformation method is proposed using CycleGAN and cyclic loss during training. The model can be adapted to improve segmentation masks, especially in tasks with inconsistent data. It eliminates noise and detection inaccuracies while preserving fine object details.

The Boundary-Aware GAN model with a multi-level discriminator described in [9] is designed to improve object boundaries in semantic segmentation problems. The discriminator is trained to distinguish between real and predicted boundaries. Improvements in object detection include identifying clear boundaries, reducing noise, and eliminating inaccuracies.

The paper [10] explores the application of GANs for binary semantic segmentation on imbalanced datasets. The authors propose a new architecture and approach to training GANs to improve the quality of object mask prediction under conditions of strong class imbalance. The main focus is on the problems of low prediction accuracy for the minority class, loss of fine object details, noise, and artifacts in the predicted masks. The authors demonstrate that GAN can improve the quality of masks due to more accurate modeling of object boundaries and restoration of fine details.

The article [11] describes the Power-Line GAN (PLGAN) model for segmentation of thin and long objects against a background of complex textures and noise. The main goal is to improve the accuracy of power line detection, which is important for infrastructure monitoring, mapping, and environmental analysis. The key ideas proposed in the paper include adapting GAN for thin object segmentation, improving detail, handling imbalanced data, and smoothing edges using active contours and spline interpolation.

The problem of unbalanced data distribution between the source domain (a set of images) and the target domain in the tasks of semantic segmentation of aerial photographs is considered in [12]. The authors propose a method for training a model without using labels for the target domain. For this purpose, a GAN is used that adapts the stylistic features of the source domain to the target one. The main goal of this work is to improve the segmentation accuracy on the target domain, minimizing the influence of differences between domains, such as changes in lighting, shooting angles, or object textures. The authors note the following improvements achieved using the proposed approach: better generalization of the model, eliminating the influence of differences between domains, preserving fine details, and reducing noise.

In [13], the application of cGAN to the task of semantic segmentation of medium-resolution satellite images is investigated. The authors focus on improving the accuracy of object extraction in such images, which often suffer from noise, lack of detail, and complex textures. The main goal of the work is to demonstrate how cGAN can be used to improve the quality of segmentation compared to traditional approaches such as U-Net or DeepLab. Particular attention is paid to the tasks of land use monitoring, urban planning, and environmental change analysis. The improvements achieved by the proposed approach include high accuracy for minority classes, sharp object boundaries, preservation of fine details, and noise removal.

In [14], the authors propose a new approach that uses GAN to improve the quality of predicted masks by more accurately modeling object boundaries and restoring fine details. The main focus is on tasks related to the analysis of high-resolution images such as aerial photographs, medical images, and urban landscapes. The aim of the work is to show that using

GAN can significantly improve the accuracy and detail of object instance segmentation compared to traditional methods such as Mask R-CNN or DeepLab. Improving the detail and smoothing the contours of masks are the main results of this work.

2. Dataset generation

The dataset used in the experimental studies is a set of (predicted mask, exact mask) pairs. The predicted mask is a binary mask of an object obtained using the Mask R-CNN model. The exact mask is a binary mask formed from a polygon of the same object extracted from a json file.

The types and quantities of objects presented in the dataset, as well as the features of forming json files for image files, are described in [1]. For polygons of objects missing from the json files, the corresponding exact masks were formed (corrected) manually. This made it possible to create a representative and consistent dataset, including both automatically generated and manual masks. All masks in the dataset were reduced to a size of 256×256 pixels. Larger mask sizes lead to a sharp increase in computational load and, accordingly, computation time, smaller ones lead to uneven contours of predicted masks, after increasing their size. Increasing to the size of an image with real estate objects is necessary at the stage of visualization of the result in order to combine the mask of the object and the image containing it.

3. Model creation and exploration

3.1. GAN architecture features

GANs consist of two main components, a generator and a discriminator, which operate in an adversarial mode. The main goal of a GAN is to train the generator to produce data that is indistinguishable from real data.

The generator takes random noise (usually from a normal distribution) as input during training and transforms it into some data, such as an image. The generator's task is to produce data that looks real. To do this, it learns to minimize the losses that occur when the discriminator tries to detect a fake. The main generator architectures are the following:

DCGAN (Deep Convolutional GAN) [15] is simple and effective, but may produce less detailed results on complex data. The generator uses transposed convolutions to gradually increase the image resolution. It starts with a random vector (latent space) and successively increases the dimensionality of the data to the target resolution.

U-Net [16] is suitable for tasks where it is important to preserve spatial details, such as segmentation or mask enhancement. The generator is based on the U-Net architecture, which includes an encoder for dimensionality reduction and a decoder for image restoration. It uses a skip connection mechanism, which allows preserving fine details.

StyleGAN [17] provides high detail and control over style, but requires large computational resources. Allows control over different levels of image detail (e.g. overall shape and textures). The generator is divided into two parts for style matching and synthesis.

The discriminator evaluates the quality of the generated data. The goal of the discriminator is to help the generator generate more reliable data (images) by providing correct error signals. The main discriminator architectures are:

PatchGAN [3] is effective for tasks where object boundaries and fine details are important. It analyzes the image by small local areas ("patches") instead of the entire image. Each patch is classified as real or fake, which allows focusing on local details.

Multi-level discriminator [18] improves the quality of generation due to a more comprehensive assessment of the data. It uses multiple discriminators, each operating at different levels of resolution. This allows evaluating both global and local image characteristics.

Conditional GAN (cGAN) [19] is the most effective for tasks that require directed generation, such as translating images between groups. The discriminator accepts additional information (such as class labels or conditional data) along with the image, allowing control over the type of data generated.

Considering the features of the above-described generator and discriminator architectures, the U-Net and PatchGAN architectures were chosen to improve the accuracy of small real estate masks, respectively.

3.2. Model creation

The generator, implemented on the basis of the U-Net architecture, implements two main stages: compression (downsampling) and expansion (upsampling).

The compression stage reduces the spatial resolution of the data. Each of the 5 groups of elements reduces the image size by 2 times and consists of a convolutional layer (`Conv2d`) with a kernel size (`kernel_size`) of 4x4, a stride (`stride`) of 2 and a padding (`padding`) of 1. Then a `BatchNorm2d` layer can follow with the activation function `LeakyReLU` and a coefficient of 0.2.

Dilation elements, on the contrary, increase the spatial resolution of the data. There are also 5 expansion elements in the proposed generator. Each such element increases the image size by 2 times by performing a transposed convolution (`ConvTranspose2d`) with a kernel size of 4x4, a stride of 2, and a padding of 1. Then a `BatchNorm2d` layer follows (except for the last block) with the activation function `ReLU` (except for the last block, which uses `Tanh` to limit the output to the range [-1, 1]). The organization of the GAN generator is shown in Figure 2.

The expansion element combines its output data with the data from the corresponding image compression element via skip connectors. This allows preserving details and improving the quality of generation. The generator output is a single-channel image of the same size as the input. The compression and expansion elements of the level 1 generator described above are shown in Figure 3 and Figure 4, respectively.

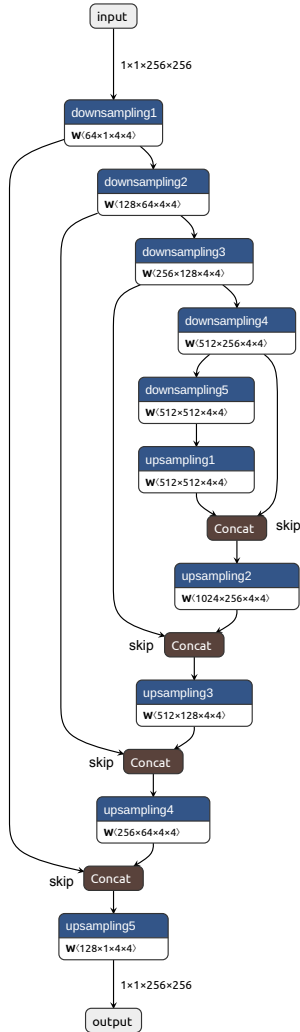
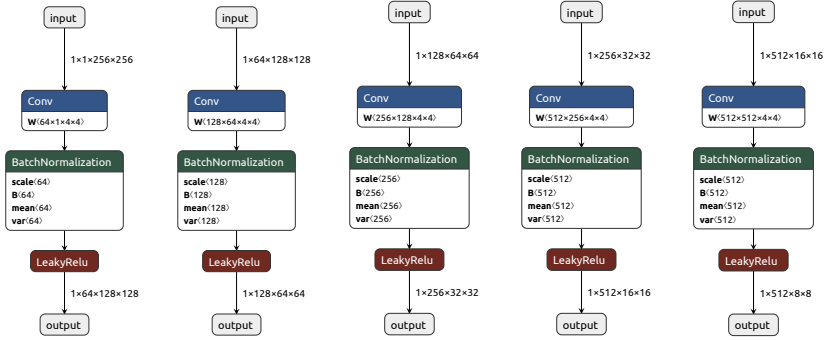


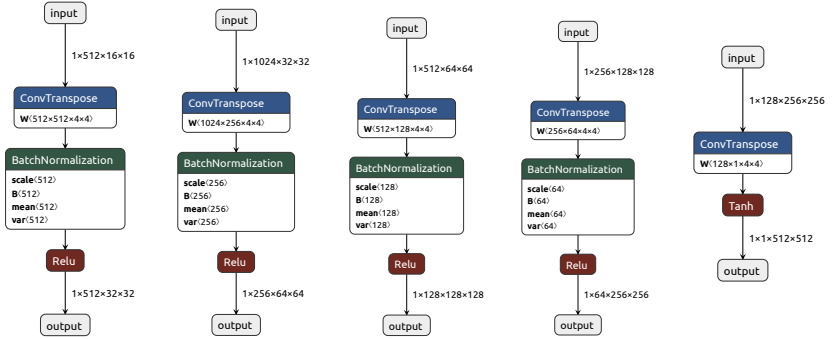
FIGURE 2. GAN generator

As noted above, the discriminator based on the PatchGAN architecture analyzes the image by small local regions (patches). Each patch is classified as real or fake, which allows focusing on local details. The discriminator



(a) downsampling1 (b) downsampling2 (c) downsampling3 (d) downsampling4 (e) downsampling5

FIGURE 3. GAN generator compression elements



(a) upsampling1 (b) upsampling2 (c) upsampling3 (d) upsampling4 (e) upsampling5

FIGURE 4. GAN generator extension elements

takes two input images: the predicted image (generated by the generator) and the real image (in our case, they are accurate object masks). Then these images are concatenated by channels. The next few convolutional layers (Conv2d) with a kernel size of 4x4, a stride of 2, and a padding of 1 reduce the spatial dimensions of the data.

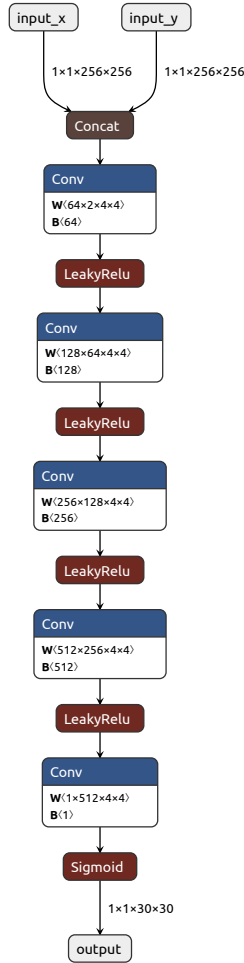


FIGURE 5. GAN discriminator

Starting from the second layer, BatchNorm2d is applied with the activation function LeakyReLU and a coefficient of 0.2. The last layer with the activation function Sigmoid calculates the probability that the input data is real. The structure of the discriminator is shown in Figure 5.

The generator uses the standard loss function for binary classification problems `BCEWithLogitsLoss` and `L1 Loss` to minimize the pixel difference between the generated and real masks. The first of these functions is used to “fool” the discriminator, the second helps the generator create masks that are close to the real ones in pixel values. An important parameter for `L1 Loss` is λ_{L1} , which controls the importance of the pixel difference between the generated and real masks. This allows you to balance between “fooling” the discriminator and the accuracy of detail reproduction. The generator loss function is `BCEWithLogitsLoss`.

Training GANs is associated with certain problems, such as instability of the training process, mode collapse, and low generation quality, which require the use of special methods to address them. One such approach is spectral normalization, which limits the spectral norm of the weights in the generator and discriminator layers, thereby stabilizing training and preventing gradient explosion [20]. Another method is to use a gradient penalty, which adds regularization to the discriminator outputs, providing a smoother distribution of values and improving model convergence [21]. Self-attention, which allows the model to take into account global dependencies in the data, is also widely used, which is especially useful when working with high-dimensional images [22]. However, since restoring simple binary masks is a relatively simple task, and as the results of the study of the GAN model described above showed, the use of these approaches in this case is redundant.

3.3. Model Study Results

The GAN model was developed and studied using the `PyTorch` framework in Google Colab Pro. The Jupyter notebook (MIT License) is available at this link— jupyter. The main accuracy metrics were studied for different values of learning rate, number of epochs, and batch sizes. The accuracy metrics of the GAN generator were studied at $\lambda_{L1} = 100$, since

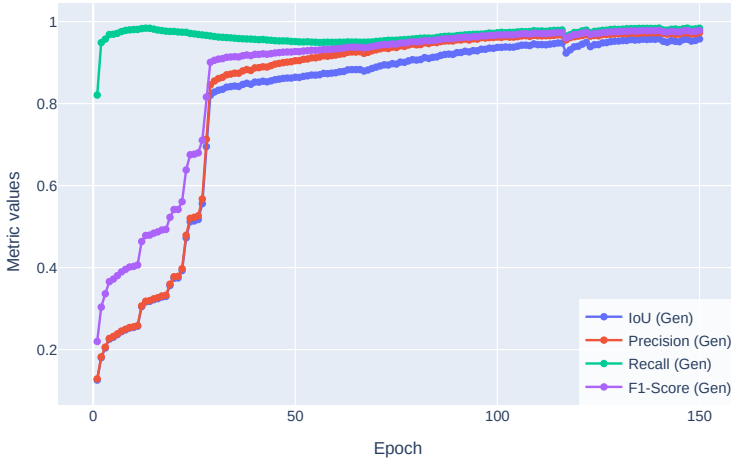


FIGURE 6. GAN generator accuracy metrics for $\lambda_{L1}=80$

at this value their curves have a sufficient degree of smoothness, which indicates the absence of explosive or decaying gradients. In Figure 6, Figure 7, and Figure 8 show the accuracy metrics curves for $\lambda_{L1} = 80, 100$ and 120.

Table 1 present the results of studies of the main accuracy metrics IoU, IoU0.5-95, Precision, Recall and F1-Score. As the studies have shown, the use of GAN can significantly improve the accuracy of real estate segmentation masks obtained using the Mask R-CNN [1] model.

Let's look at some examples of improving real estate masks obtained using the Mask R-CNN model. Figure 9 shows masks predicted by the Mask R-CNN model and improved masks generated by GAN. Based on the accuracy metrics, it can be concluded that using GAN can significantly improve the quality of masks by correcting errors such as contour discontinuities, noise, and inaccuracies in the boundaries of objects.

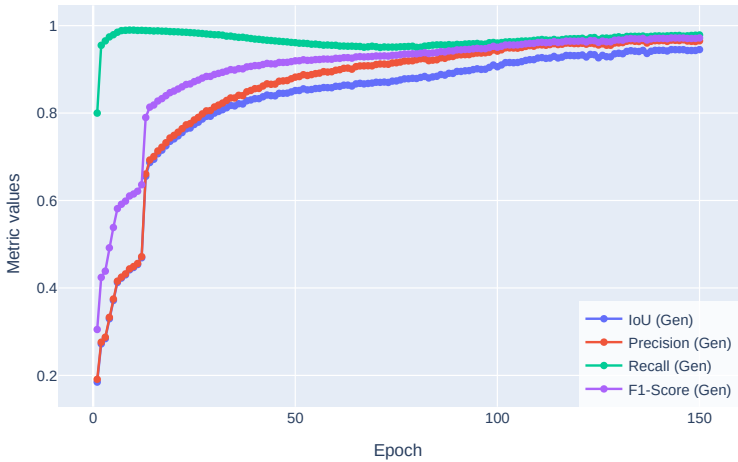


FIGURE 7. GAN generator accuracy metrics for $\lambda_{L1}=100$

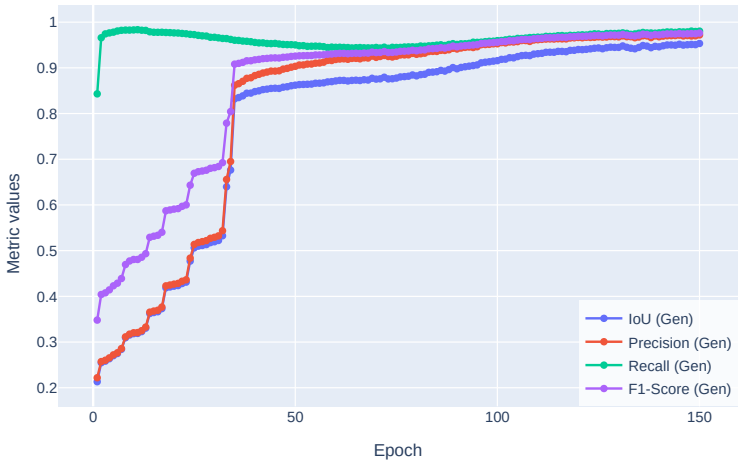


FIGURE 8. GAN generator accuracy metrics for $\lambda_{L1}=120$

TABLE 1. Evaluation of accuracy metrics for predicted GAN masks under different values of model training parameters

Learning Speed	Batch Size	Learning Epochs	IoU (Pred)	Precision (Pred)	Recall (Pred)	F1-Score (Pred)	IoU0.5-95 (Pred) (Mask R-CNN)	IoU0.5-95 (Pred) (GAN)
1e-6	10	500	0.87	0.92	0.91	0.91	0.67	0.90
5e-6	10	400	0.85	0.90	0.89	0.90	0.65	0.88
5e-6	20	350	0.84	0.89	0.88	0.88	0.64	0.87
1e-5	10	300	0.83	0.88	0.87	0.87	0.63	0.86
1e-5	20	250	0.84	0.89	0.88	0.88	0.64	0.87
5e-5	8	200	0.79	0.84	0.83	0.83	0.59	0.81
5e-5	10	300	0.82	0.87	0.86	0.86	0.62	0.85
5e-5	16	250	0.81	0.86	0.85	0.85	0.61	0.84
4e-5	8	150	0.80	0.85	0.84	0.84	0.60	0.82
4e-5	16	200	0.78	0.83	0.82	0.82	0.58	0.80
1e-4	8	100	0.75	0.80	0.78	0.79	0.55	0.78
1e-4	16	150	0.77	0.82	0.81	0.81	0.57	0.80
1e-4	20	200	0.78	0.83	0.82	0.82	0.58	0.81
2e-4	8	80	0.70	0.75	0.73	0.74	0.50	0.73
2e-4	16	100	0.72	0.78	0.76	0.77	0.52	0.76



FIGURE 9. The results of GAN enhancement are shown under the Mask R-CNN predicted real estate object masks

An example of improving object masks on an aerial photograph, where the complexity of the task increases due to the high density of objects and their diversity, is shown in Figure 10.

When generating masks using GAN, the problem of uneven and stepped edges often arises, see Figure 9. This is due to the specifics of GAN operation, which is associated with the discreteness of pixels when forming images and the complexity of accurately reproducing small details of object boundaries. To improve the contours of segmentation masks, various morphological smoothing functions can be applied, which allow you to eliminate noise, fill in gaps and make object boundaries smoother.

One of the basic approaches is to use morphological opening and closing operations with a given structuring element (for example, a square matrix (5×5)). Opening helps to remove small noise artifacts, and closing fills small gaps inside the mask. For additional smoothing of contours, you can apply Gaussian blur (`GaussianBlur`), which softens sharp transitions between the background and the object.

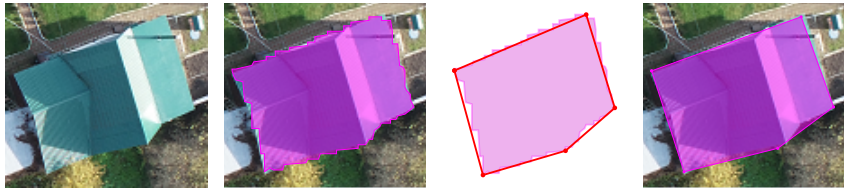


(a) Results of detecting real estate masks using the Mask R-CNN model



(b) The masks are improved by GAN generator

FIGURE 10. Improvement of objects masks



(a) Real estate (b) Jagged edges (c) Aligned edges (d) Resulting mask

FIGURE 11. Eliminating mask aliasing using the Douglas-Peucker algorithm

Another effective method is to use **active contours (snakes)**, which adaptively adjust to the shape of the object by minimizing the elastic energy and smoothness. To achieve even smoother contours, **spline interpolation** can be used, which approximates the original contour using cubic splines, ensuring uniform distribution of points and smooth boundaries.

Another effective way to eliminate aliasing is to use the Douglas-Peucker algorithm, which specializes in simplifying polylines and smoothing contours [23]. The peculiarity of this algorithm is its ability to analyze each point of the mask boundary and evaluate its significance for the overall shape of the object. Points that make a minimal contribution to the contour configuration are removed, which allows you to get smoother and more natural boundaries without losing key geometric features. An example of using the Douglas-Peucker algorithm is shown in Figure 11.

These methods or their combinations allow in some cases to improve the quality of masks, making them more suitable for subsequent analysis or visualization. As studies have shown, the use of morphological smoothing methods and improvement of mask boundaries using the Douglas-Peucker algorithm in post-processing of masks makes it possible to increase the accuracy of their formation by an average of 5–7%, which is especially important for computer vision tasks and segmentation of objects on aerial photographs in particular.

Conclusion

In this paper, a GAN model was proposed and investigated to improve the accuracy of segmentation masks obtained using the Mask R-CNN model. Experiments have shown that the use of GAN can significantly improve the quality of detected real estate masks by correcting errors such as gaps in masks, noise artifacts, and jagged edges. The improvement is achieved by using the L1 loss function to minimize pixel differences between the target and generated masks, as well as by using a discriminator that contributes to more realistic mask generation.

The results of the studies demonstrate an improvement in accuracy metrics such as IoU, Precision, Recall, and F1-Score for the improved masks compared to the original Mask R-CNN predictions. The average IoU0.5-95 value increased by an average of 30-35%, which confirms the effectiveness of the proposed approach. It was additionally noted that the use of morphological operations and contour smoothing methods will further improve the smoothness and accuracy of mask boundaries.

The obtained results are of great practical importance and will be used in the «Roskadastr» PLC IS for automatic conversion of raster images with real estate objects into vector maps. This will improve the accuracy and speed of digital map generation, which is especially important for cadastral registration and territory management tasks. The proposed approach can be adapted for other applications related to image processing and computer vision, where high accuracy of object segmentation is required. The study conducted in the work confirms the prospects of using GAN to improve the quality of real estate masks, which allows it to be effectively used to automate the processes of data analysis and transformation in geographic information systems. Further research can be aimed at optimizing the model parameters, introducing new methods of mask post-processing and expanding the scope of GAN.

References

- [1] Vinokurov I. V.. “Using the Mask R-CNN model for segmentation of real estate objects in aerial photographs”, *Program Systems: Theory and Applications*, **16**:1(64) (2025), pp. 3-44.   [↑116, 123](#)

- [2] G. Cohen, R. Giryes. *Generative adversarial networks*, 2024, 28 pp.   2203.00667 
- [3] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros. *Image-to-image translation with conditional adversarial networks*, 2016, 17 pp.   1611.07004 ¹¹³, ¹¹⁷
- [4] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro. *High-resolution image synthesis and semantic manipulation with conditional GANs*, 2017, 14 pp.   1711.11585 ¹¹⁴
- [5] C.-H. Lee, Z. Liu, L. Wu, P. Luo. *MaskGAN: Towards diverse and interactive facial image manipulation*, 2019, 20 pp.   1907.11922 ¹¹⁴
- [6] Y. Xue, T. Xu, H. Zhang, L. Rodney Long, X. Huang. *SegAN: Adversarial network with multi-scale L_1 loss for medical image segmentation*, 2017, 9 pp.   1706.01805 ¹¹⁴
- [7] X. Chen, C. Xu, X. Yang, D. Tao. *Attention-GAN for object transfiguration in wild images*, 2018, 18 pp.   1803.06798 ¹¹⁴
- [8] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros. *Unpaired image-to-image translation using cycle-consistent adversarial networks*, 2017, 18 pp.   1703.10593 ¹¹⁴
- [9] J. Gong, J. Xu, X. Tan, J. Zhou, Y. Qu, Y. Xie, L. Ma. *Boundary-aware geometric encoding for semantic segmentation of point clouds*, 2021, 9 pp.   2101.02381 ¹¹⁴
- [10] L. Xu, M. Gabbouj. *Revisiting generative adversarial networks for binary semantic segmentation on imbalanced datasets*, 2024, 14 pp.   2402.02245 ¹¹⁴
- [11] R. Abdelfattah, X. Wang, S. Wang. *JPLGAN: Generative adversarial networks for power-line segmentation in aerial images*, 2022, 11 pp.   2204.07243 ¹¹⁵
- [12] B. Benjdira, Y. Bazi, A. Koubaa, K. Ouni. “Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images”, *Remote Sens.*, **11**:11 (2019), id. 1369, 23 pp.   1905.03198 ¹¹⁵
- [13] A. Kulkarni, T. Mohandoss, D. Northrup, E. Mwebaze, H. Alemohammad. *Semantic segmentation of medium-resolution satellite imagery using conditional generative adversarial networks*, 2020, 7 pp.   2012.03093 ¹¹⁵
- [14] Q. H. Le, K. Youcef-Toumi, D. Tsetserukou, A. Jahanian. *GAN Mask R-CNN: Instance semantic segmentation benefits from generative adversarial networks*, 2020, 13 pp.   2010.13757 ¹¹⁵
- [15] A. Radford, L. Metz, S. Chintala. *Unsupervised representation learning with deep convolutional generative adversarial networks*, 2016, 16 pp.   1511.06434 ¹¹⁷

- [16] O. Ronneberger, Ph. Fischer, Th. Brox. *U-Net: Convolutional networks for biomedical image segmentation*, 2015, 8 pp. [arXiv](#) 1505.04597 ¹¹⁷
- [17] T. Karras, S. Laine, T. Aila. *A style-based generator architecture for generative adversarial networks*, 2018, 12 pp. [arXiv](#) 1812.04948 ¹¹⁷
- [18] T. Karras, T. Aila, S. Laine, J. Lehtinen. *Progressive growing of GANs for improved quality, stability, and variation*, 2017, 26 pp. [arXiv](#) 1710.10196 ¹¹⁷
- [19] M. Mirza, S. Osindero. *Conditional generative adversarial nets*, 2014, 7 pp. [arXiv](#) 1411.1784 ¹¹⁷
- [20] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida. *Spectral normalization for generative adversarial networks*, 2018, 26 pp. [arXiv](#) 1802.05957 ¹²²
- [21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. C. Courville. *Improved training of Wasserstein GANs*, 2017, 20 pp. [arXiv](#) 1704.00028 ¹²²
- [22] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena. *Self-attention generative adversarial networks*, 2019, 10 pp. [arXiv](#) 1805.08318 ¹²²
- [23] H. Chen. “An improved Douglas-Peucker algorithm applied in coastline generalization”, Fourth International Conference on Geology, Mapping, and Remote Sensing (ICGMRS 2023) (14–16 April 2023, Wuhan, China), Proc. SPIE, vol. **12978**, 2024, id. 1297804. ¹²⁸

Received	21.04.2025;
approved after reviewing	10.06.2025;
accepted for publication	11.06.2025;
published online	28.06.2025.

Recommended by

prof. A. M. Elizarov

Information about the author:



Igor Victorovich Vinokurov

Candidate of Technical Sciences (PhD), Associate Professor at the Financial University under the Government of the Russian Federation. Research interests: information systems, information technologies, data processing technologies

0000-0001-8697-1032
 e-mail: igvvinokurov@fa.ru

The author declare no conflicts of interests.

УДК 004.932.75'1, 004.89

 10.25209/2079-3316-2025-16-2-111-152

Повышение точности сегментирования объектов с использованием генеративно-состязательной сети

Игорь Викторович **Винокуров** 

Финансовый Университет при Правительстве Российской Федерации, Москва, Россия

 igvvinokurov@fa.ru

Аннотация. Маски, полученные с использованием модели глубокого обучения Mask R-CNN, в ряде случаев могут содержать фрагментированные контуры, неровные границы, ложные сращения соседних объектов и участки с пропущенной сегментацией. Чем больше объектов детектирования на изображении и меньше их размер, тем более чаще встречаются различного вида недостатки их масок. Примерами таких изображений могут являться аэрофотоснимки коттеджных и садовых товариществ и кооперативов, характеризующихся высокой плотности застройки. Для коррекции указанных недостатков предлагается использовать модель генеративно-состязательной сети, выполняющую постобработку предсказанных Mask R-CNN масок.

Качественная оценка сформированной в работе модели продемонстрировала, что она способна на приемлемом уровне восстанавливать целостность контуров, заполняет пропущенные области и разделять ошибочно объединенные объекты. Количественный анализ по метрикам IoU, precision, recall и F1-score показал статистически значимое улучшение качества сегментации по сравнению с исходными масками Mask R-CNN. Полученные результаты подтвердили, что предложенный подход позволяет довести точность формирования масок объектов до уровня, удовлетворяющего требованиям их практического применения в системах автоматизированного анализа аэрофотоснимков. *(Связанные тексты статьи на английском и на русском языках)*

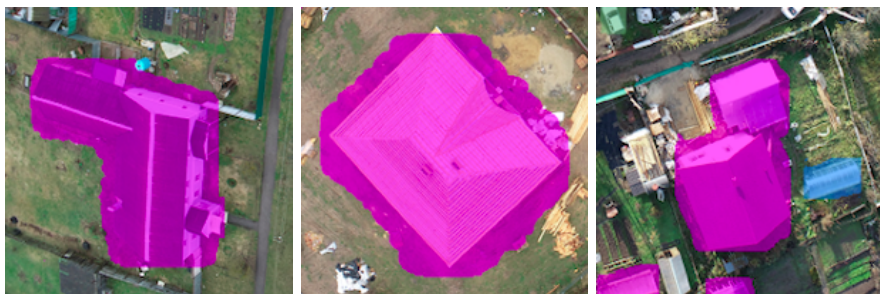
Ключевые слова и фразы: Компьютерное зрение, сегментация изображений, маски объектов, генеративно-состязательные сети, Mask R-CNN, PyTorch

Для цитирования: Винокуров И.В. *Повышение точности сегментирования объектов с использованием генеративно-состязательной сети* // Программные системы: теория и приложения. 2025. Т. 16. № 2(65). С. 111–152. (Англ.+русс.) https://psta.psiras.ru/read/psta2025_2_111-152.pdf

Введение

Выявление точных масок объектов недвижимости на аэрофотоснимках с плотной застройкой представляет определённую сложность для моделей глубокого обучения, Mask R-CNN, YOLO и других. Основной причиной является близкое расположение зданий друг к другу, что затрудняет точное разделение объектов, поскольку их контуры могут пересекаться или сливаться на изображении.

Это особенно проблематично в условиях высокой детализации, где небольшие элементы архитектуры, такие как пристройки, балконы или общие стены между строениями, могут быть восприняты моделью как единый объект. Кроме того, сложная форма зданий, например, многоугольные очертания или нестандартные конструкции требует от модели более глубокого анализа и адаптации, что увеличивает вероятность ошибок при формировании масок. Несколько примеров таких масок приведены на рисунке 1.



(а) Неточная сегментация (б) Частичное отсутствие сегментации (в) Слияние масок объектов

Рисунок 1. Ошибки при формировании масок объектов

Ещё одной причиной является наличие внешних факторов, таких как растительность и тени, которые частично скрывают объекты на снимках. Деревья, кустарники или даже временные объекты, такие как автомобили, могут перекрывать части зданий, создавая разрывы в данных, доступных модели. Тени, вызванные различными углами съемки или временем суток, также искажают визуальную информацию, снижая контраст и затрудняя выделение четких границ объектов. Эти факторы приводят к тому, что модель может либо недооценивать площадь объекта, игнорируя скрытые участки, либо переоценивать ее, включая посторонние элементы в маску. В результате повышается потребность в дополнительной обработке и улучшении качества предсказаний, чтобы обеспечить корректность итоговых данных.

Проблема уточнения границ объектов особенно актуальна для информационной системы (ИС) ППК «Роскадастр», где к точности пространственных данных предъявляются повышенные требования. Традиционные методы постобработки, такие как морфологические операции или алгоритмы активных контуров, часто оказываются недостаточно эффективными для сложных случаев плотной застройки. В связи с этим возникает необходимость в решении задачи автоматического корректирования результатов сегментации, сохраняя при этом топологическую правильность объектов.

В представленной работе предлагается одно из возможных решений этой задачи, основанное на применении генеративно-сопоставительной сети (GAN) для постобработки результатов сегментации, полученных с помощью Mask R-CNN. Основная идея исследования заключается в том, что GAN-архитектура, должным образом обученная на репрезентативной выборке объектов застройки, характерной для ППК «Роскадастр», сможет эффективно восстанавливать целостность нарушенных границ, устранять внутренние разрывы масок, разделять ошибочно слившиеся объекты и сохранять точное положение угловых (поворотных) точек зданий.

Реализация исследований состояла из следующих этапов:

- (1) Формирование репрезентативного и сбалансированного датасета, содержащего соответствия исходных масок от Mask R-CNN эталонным маскам, извлечённым из файлов с аннотациями объектов и маскам, полученные в результате их экспертной коррекции.
- (2) Разработка архитектуры GAN, учитывающей особенности плотной застройки в коттеджных и садовых товариществах и кооперативах, требования к точности границ для кадастрового учета и необходимость сохранения топологии объектов.
- (3) Проведение экспериментальных исследований с оценкой качественных характеристик получаемых масок и количественных показателей (IoU, precision, recall, F1-score).

1. Обзор работ по использованию GAN для улучшения результатов распознавания объектов

В настоящее время существует значительное количество научных исследований, посвященных применению генеративно-сопоставительных сетей (GAN) для повышения эффективности распознавания изображений и их трансформации.

В работе [3] рассматриваются модели условно-сопоставительных сетей (conditional GAN, cGAN), которые могут быть применены для преобразования изображений одного типа в другой. Например, маски сегментации могут быть преобразованы в реалистичные изображения, или наоборот. Описанные модели демонстрируют высокую точность сегментации объектов и улучшают качество их масок.

Модель Pix2Pix, описанная в [4] и основанная на cGAN, используется для синтеза высококачественных изображений на основе входных данных, таких как маски сегментации. В статье показано, что GAN может улучшить качество масок путем обучения на парах «маска-точная маска».

В исследовании [5] представлена модель MaskGAN, предназначенная для интерактивного редактирования изображений лиц с использованием масок. Модель улучшает качество масок и позволяет манипулировать сегментированными областями изображений.

В [6] предлагается модель SegAN с многоуровневым дискриминатором и L1-потерями для медицинской сегментации. Модель использует многоуровневый дискриминатор для улучшения качества масок, особенно в задачах с размытыми границами объектов. GAN-модель улучшает границы объектов, снижает уровень шума и устраняет неточности детектирования объектов.

Механизм внимания для генерации масок в Attention-GAN, описанный в [7], применяется для трансформации объектов на фотографиях дикой природы. Модель позволяет устранять шумы и неточности детектирования, сохраняя мелкие детали объектов.

В статье [8] предлагается метод преобразования изображений с использованием CycleGAN и циклических потерь при обучении. Модель может быть адаптирована для улучшения масок сегментации, особенно в задачах с несоответствующими данными. Она устраняет шумы и неточности детектирования объектов, сохраняя их мелкие детали.

Модель Boundary-Aware GAN с многоуровневым дискриминатором, описанная в [9], предназначена для улучшения границ объектов в задачах семантической сегментации. Дискриминатор обучается различать реальные и предсказанные границы. Улучшения детектирования объектов заключаются в выявлении четких границ, снижении уровня шума и устранении неточностей.

В статье [10] исследуется применение GAN для бинарной семантической сегментации на несбалансированных датасетах. Авторы предлагают новую архитектуру и подход к обучению GAN для улучшения качества предсказания масок объектов в условиях сильной разбалансировки классов. Основное внимание уделяется проблемам низкой точности предсказания для миноритарного класса, потере мелких деталей объектов, шуму и артефактам в предсказанных масках. Авторы демонстрируют, что GAN позволяет улучшить качество масок за счёт более точного моделирования границ объектов и восстановления мелких деталей.

В работе [11] описывается модель PLGAN (Power-Line GAN), предназначенная для сегментации тонких и длинных объектов на фоне сложных текстур и шума. Основная цель — улучшить точность выделения линий электропередач, что важно для мониторинга инфраструктуры, картографирования и анализа окружающей среды. В статье предложены ключевые идеи, включая адаптацию GAN для сегментации тонких объектов, улучшение детализации, обработку несбалансированных данных и сглаживание контуров с использованием активных контуров и сплайн-интерполяции.

Проблема несбалансированного распределения данных между исходным доменом (совокупность изображений) и целевым доменом в задачах семантической сегментации аэрофотоснимков рассматривается в [12]. Авторы предлагают метод обучения модели без использования разметки для целевого домена. Для этого используется GAN, которая адаптирует стилистические особенности исходного домена к целевому. Основная цель этой работы — улучшить точность сегментации на целевом домене, минимизируя влияние различий между доменами, таких как изменения в освещении, ракурсы съемки или текстуры объектов. Авторы отмечают следующие улучшения, достигнутые с помощью предложенного подхода: лучшие обобщения модели, устранение влияния различий между доменами, сохранение мелких деталей и снижение шума.

В статье [13] исследуется применение cGAN для задач семантической сегментации спутниковых изображений среднего разрешения. Авторы фокусируются на улучшении точности выделения объектов на таких изображениях, которые часто страдают от шума, недостаточной детализации и сложных текстур. Основная цель работы — продемонстрировать, как cGAN может быть использована для повышения качества сегментации по сравнению с традиционными подходами, такими как U-Net или DeepLab. Особое внимание уделяется задачам мониторинга землепользования, городского планирования и анализа изменений в окружающей среде. Улучшениями, достигнутыми с помощью предложенного подхода, являются высокая точность для миноритарных классов, чёткие границы объектов, сохранение мелких деталей и устранение шума.

В [14] авторы предлагают новый подход, который использует GAN для улучшения качества предсказанных масок за счёт более точного моделирования границ объектов и восстановления мелких деталей. Основное внимание уделяется задачам, связанным с анализом изображений с высоким разрешением, таких как аэрофотоснимки, медицинские изображения и городские ландшафты. Цель работы — показать, что использование GAN может значительно повысить точность и детализацию

сегментации экземпляров объектов по сравнению с традиционными методами, такими как Mask R-CNN или DeepLab. Улучшение детализации и сглаживание контуров масок являются основными результатами этой работы.

2. Формирование набора данных

Использованный в экспериментальных исследованиях датасет представляет собой совокупность пар (предсказанная маска, точная маска). Предсказанная маска это бинарная маска объекта, полученная с использованием модели Mask R-CNN. Точная маска представляет собой бинарную маску, сформированную из полигона того же самого объекта, извлечённого из json-файла.

Типы и количества объектов, представленных в датасете, а также особенности формирования json-файлов к файлам с изображениями, описаны в [1]. Для отсутствующих в json-файлах полигонов объектов, соответствующие им точные маски формировались (корректировались) вручную. Это позволило создать репрезентативный и согласованный набор данных, включающий как автоматически сгенерированные, так и ручные маски. Все маски датасета были приведены к размеру 256×256 пикселей. Большой размер масок приводит к резкому увеличению вычислительной нагрузки и, соответственно, времени вычислений, меньший — к неровностям контуров предсказанных масок, после увеличения их размера. Увеличение до размера изображения с объектами недвижимости необходимо на этапе визуализации результата с целью совмещения маски объекта и содержащего его изображения.

3. Формирование и исследование модели

3.1. Особенности архитектуры GAN

GAN состоят из двух основных компонент: генератора и дискриминатора, которые работают в состязательном режиме. Основная цель GAN заключается в обучении генератора создавать данные, неотличимые от реальных.

Генератор в процессе обучения принимает случайный шум (обычно из нормального распределения) в качестве входных данных и преобразует его в некоторые данные, например, изображения. Задача генератора — сформировать данные, которые выглядят как реальные. Для этого он обучается минимизировать потери, возникающие при попытке дискриминатора определить подделку. Основными архитектурами генератора являются следующие:

DCGAN (Deep Convolutional GAN) [15]. Простой и эффективный, но может выдавать менее детализированные результаты на сложных данных. Генератор использует транспонированные свёртки для постепенного увеличения разрешения изображения. Начинает со случайного вектора (латентного пространства) и последовательно увеличивает размерность данных до целевого разрешения.

U-Net [16]. Подходит для задач, где важно сохранить пространственные детали, например, сегментация или улучшение масок. Генератор основан на архитектуре U-Net, которая включает энкодер для понижения размерности и декодер для восстановления изображения. Использует механизм пропусков (skip connections), что позволяет сохранять мелкие детали.

StyleGAN [17]. Высокая детализация и контроль над стилем, но требует больших вычислительных ресурсов. Позволяет контролировать различные уровни детализации изображения (например, общую форму и текстуры). Генератор разделен на две части для сопоставления стилей и для синтеза.

Дискриминатор оценивает качество сгенерированных данных. Задача дискриминатора – помочь генератору генерировать более правдоподобные данные (изображения), предоставляя корректные сигналы ошибки. Основные архитектуры дискриминатора:

PatchGAN [3]. Эффективен для задач, где важны границы объектов и мелкие детали. Анализирует изображение по небольшим локальным областям («патчам») вместо всего изображения. Каждый патч классифицируется как реальный или поддельный, что позволяет фокусироваться на локальных деталях.

Многоуровневый дискриминатор [18]. Улучшает качество генерации за счёт более комплексной оценки данных. Использует несколько дискриминаторов, каждый из которых работает на разных уровнях разрешения. Это позволяет оценивать как глобальные, так и локальные характеристики изображения.

Conditional GAN (cGAN) [19]. Наиболее эффективен для задач, где требуется направленная генерация, например, перевод изображений между их группами. Дискриминатор принимает дополнительную информацию (например, метки классов или условные данные) вместе с изображением, что позволяет контролировать тип генерируемых данных.

Учитывая особенности описанных выше архитектур генератора и дискриминатора, для повышения точности масок мелких объектов недвижимости были выбраны архитектуры U-Net и PatchGAN соответственно.

3.2. Формирование модели

Генератор, реализованный на основе архитектуры U-Net, реализует два основных этапа: сжатие (downsampling) и расширение (upsampling).

Этап сжатия уменьшает пространственное разрешение данных. Каждая из 5-ти групп элементов уменьшает размеры изображения в 2 раза и состоит из свёрточного слоя (`Conv2d`) с размером ядра (`kernel_size`) 4x4, шагом (`stride`) 2 и дополнением (`padding`) 1. За ним может следовать слой `BatchNorm2d` с функцией активации `LeakyReLU` и коэффициентом 0.2.

Элементы расширения наоборот, увеличивают пространственное разрешение данных. Элементов расширения в предлагаемом генераторе тоже 5. Каждый такой элемент увеличивает размеры изображения в 2 раза за счёт выполнения транспонированной свёртки (`ConvTranspose2d`) с размером ядра 4x4, шагом 2 и дополнением 1. Затем следует слой `BatchNorm2d` (кроме последнего блока) с функцией активации: `ReLU` (кроме последнего блока, где используется `Tanh` для ограничения выхода в диапазон $[-1, 1]$). Организация генератора GAN показана на рисунке 2.

Элемент расширения объединяет свои выходные данные с данными из соответствующего ему элемента сжатия изображения посредством скип-коннекторов. Это позволяет сохранять детали и улучшать качество генерации. На выходе генератора формируется одноканальное изображение размером, равным входному. Описанные выше элементы сжатия и расширения генератора 1-го уровня приведены на рисунках 3 и 4 соответственно.

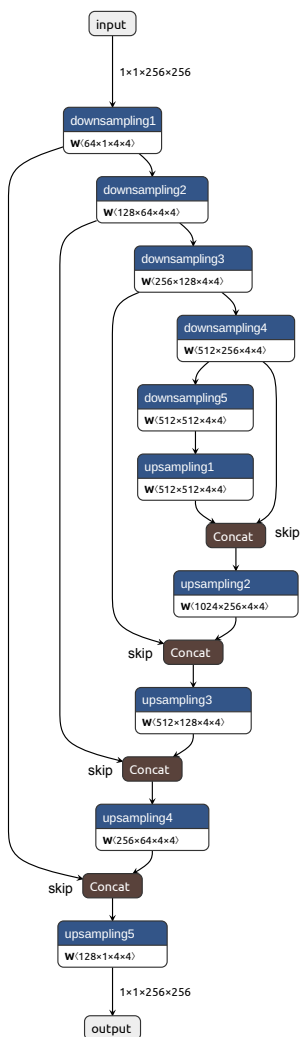


Рисунок 2. Генератор GAN

Как уже было отмечено выше, дискриминатор на основе архитектуры PatchGAN анализирует изображение по небольшим локальным областям (патчам). Каждый патч классифицируется им как реальный

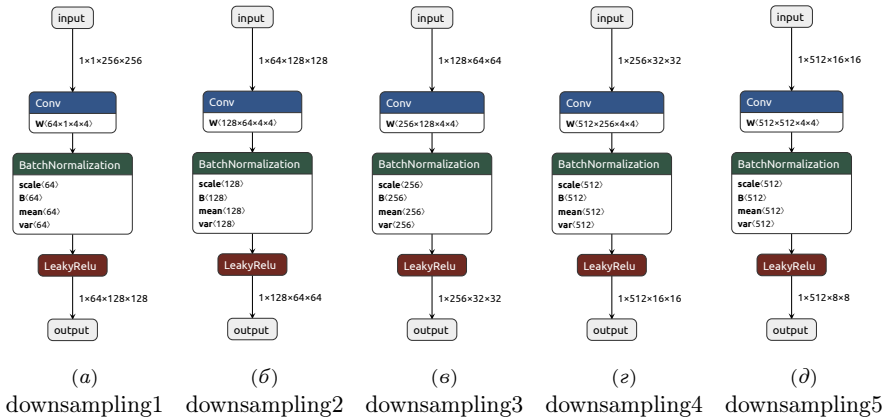


РИСУНОК 3. Элементы сжатия генератора GAN

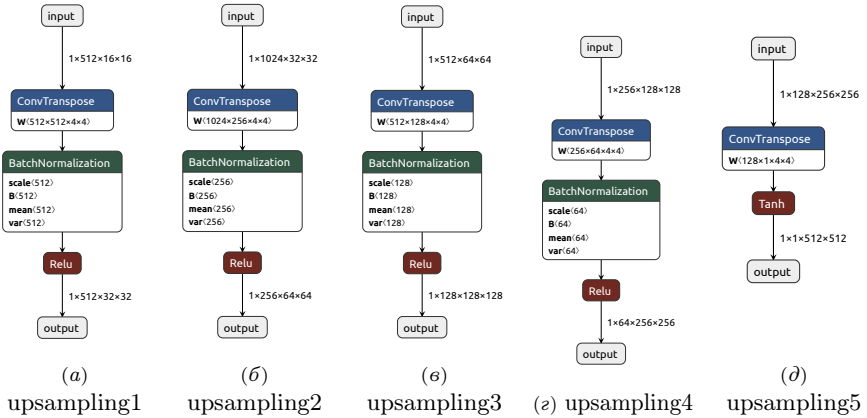


РИСУНОК 4. Элементы расширения генератора GAN

или поддельный, что позволяет фокусироваться на локальных деталях. Дискриминатор принимает два входных изображения: предсказанное изображение (сгенерированное генератором) и реальное изображение (в нашем случае они представляют собой точные маски объектов). Затем эти изображения объединяются по каналам. Последующие несколько свёрточных слоёв (Conv2d) с размером ядра 4×4 , шагом 2 и дополнением 1 уменьшают пространственные размеры данных.

Начиная со второго слоя применяется BatchNorm2d с функцией активации LeakyReLU и коэффициентом 0.2. Последний слой с функцией активации Sigmoid вычисляет вероятность того, что входные данные являются реальными. Структура дискриминатора приведена на рисунке 5.

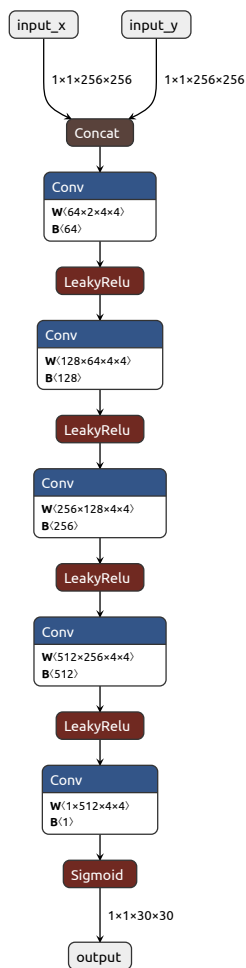


РИСУНОК 5. Дискриминатор GAN

В генераторе используется стандартная функция потерь для задач бинарной классификации `BCEWithLogitsLoss` и `L1 Loss` для минимизации пиксельной разницы между сгенерированными и реальными масками. Первая из этих функций используется для «обмана» дискриминатора, вторая помогает генератору создавать маски, которые близки к реальным по пиксельным значениям. Важным параметром при `L1 Loss` является λ_{L1} , который управляет важностью пиксельной разницы между сгенерированными и реальными масками. Это позволяет балансировать между «обман» дискриминатора и точностью воспроизведения деталей. Функция потерь генератора – `BCEWithLogitsLoss`.

Обучение GAN связано с определёнными проблемами, такими как нестабильность процесса обучения, коллапс мод и низкое качество генерации, что требует применения специальных методов для их решения. Одним из таких подходов является спектральная нормализация, которая ограничивает спектральную норму весов в слоях генератора и дискриминатора, тем самым стабилизируя обучение и предотвращая взрыв градиентов [20]. Другой метод заключается в использовании штрафа за градиент (gradient penalty), который добавляет регуляризацию к выходам дискриминатора, обеспечивая более плавное распределение значений и улучшая сходимость модели [21]. Также широкое применение нашёл механизм самовнимания (self-attention), позволяющий модели учитывать глобальные зависимости в данных, что особенно полезно при работе с высокоразмерными изображениями [22]. Однако, поскольку восстановление простых бинарных масок является относительно простой задачей, и как показали результаты исследования описанной выше модели GAN, использование этих подходов в данном случае является излишним.

3.3. Результаты исследования модели

Формирование и исследование модели GAN проводилось с использованием фреймворка `PyTorch` в среде `Google Colab Pro`. `Jupyter`-ноутбук (лицензия MIT) [доступен по этой ссылке](#):  `Jupyter`. Исследования основных метрик точности были выполнены для различных значений скорости обучения, количества эпох и размеров батча. Метрики точности генератора

GAN исследовались при $\lambda_{L1} = 100$, поскольку при таком значении их кривые обладают достаточной степенью гладкости, что свидетельствует об отсутствии взрывных или затухающих градиентов. На рисунках 6, 7 и 8 показаны кривые метрик точности для значений $\lambda_{L1} = 80, 100$ и 120 .

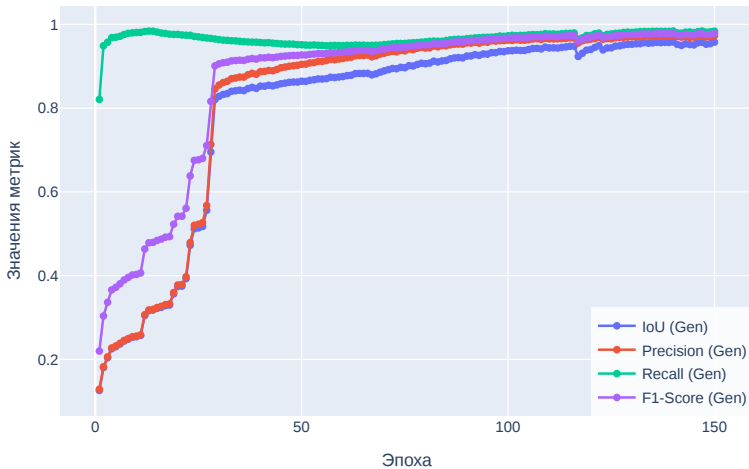


Рисунок 6. Графики метрик точности генератора GAN для $\lambda_{L1}=80$

В таблице 1 приведены результаты исследований основных метрик точности IoU, IoU0.5-95, Precision, Recall и F1-Score. Как показали проведённые исследования, использование GAN позволяет значительно улучшить точность масок сегментации объектов недвижимости, полученных с использованием модели Mask R-CNN [1].

Рассмотрим несколько примеров улучшения масок объектов недвижимости, полученных с использованием модели Mask R-CNN. На рисунке 9 показаны маски, предсказанные моделью Mask R-CNN, и улучшенные маски, сгенерированные с помощью GAN. По значениям метрик точности можно сделать вывод, что использование GAN позволяет значительно улучшить качество масок за счёт коррекции ошибок, таких как разрывы контуров, шум и неточности в границах объектов.

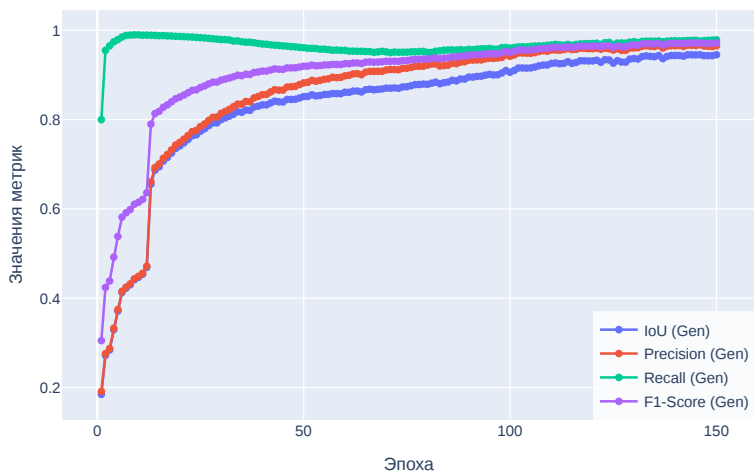


Рисунок 7. Графики метрик точности генератора GAN для $\lambda_{L1}=100$

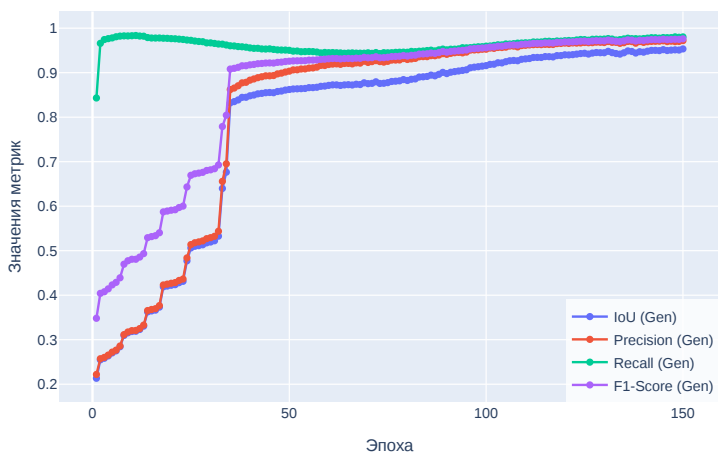


Рисунок 8. Графики метрик точности генератора GAN для $\lambda_{L1}=120$

Таблица 1. Оценка метрик точности для предсказанных GAN масок в зависимости от параметров обучения модели

Скорость обучения	Размер батча	Эпох обучения	IoU (Pred)	Precision (Pred)	Recall (Pred)	F1-Score (Pred)	IoU0.5-95 (Pred) (Mask R-CNN)	IoU0.5-95 (Pred) (GAN)
1e-6	10	500	0.87	0.92	0.91	0.91	0.67	0.90
5e-6	10	400	0.85	0.90	0.89	0.90	0.65	0.88
5e-6	20	350	0.84	0.89	0.88	0.88	0.64	0.87
1e-5	10	300	0.83	0.88	0.87	0.87	0.63	0.86
1e-5	20	250	0.84	0.89	0.88	0.88	0.64	0.87
5e-5	8	200	0.79	0.84	0.83	0.83	0.59	0.81
5e-5	10	300	0.82	0.87	0.86	0.86	0.62	0.85
5e-5	16	250	0.81	0.86	0.85	0.85	0.61	0.84
4e-5	8	150	0.80	0.85	0.84	0.84	0.60	0.82
4e-5	16	200	0.78	0.83	0.82	0.82	0.58	0.80
1e-4	8	100	0.75	0.80	0.78	0.79	0.55	0.78
1e-4	16	150	0.77	0.82	0.81	0.81	0.57	0.80
1e-4	20	200	0.78	0.83	0.82	0.82	0.58	0.81
2e-4	8	80	0.70	0.75	0.73	0.74	0.50	0.73
2e-4	16	100	0.72	0.78	0.76	0.77	0.52	0.76



Рисунок 9. Результаты улучшения с помощью GAN показаны под предсказанными Mask R-CNN масками объектов недвижимости

Пример улучшения масок объектов на аэрофотоснимке, на которых сложность задачи возрастает из-за высокой плотности объектов и их разнообразия приведён на рисунках 10.

При генерации масок с помощью GAN часто возникает проблема ступенчатых краёв, см. рисунок 9. Это обусловлено спецификой работы GAN, которая связана с дискретностью пикселей при формировании изображений и сложностью точного воспроизведения мелких деталей границ объектов. Для улучшения контуров масок сегментации могут быть применены различные функции морфологического сглаживания, которые позволяют устранить шум, заполнить пробелы и сделать границы объектов более гладкими.

Одним из базовых подходов является использование операций морфологического открытия и закрытия с заданным структурирующим элементом (например, квадратной матрицей (5×5)). Открытие помогает удалить мелкие шумовые артефакты, а закрытие заполняет мелкие разрывы внутри маски. Для дополнительного сглаживания контуров



(а) Обнаруженные моделью Mask R-CNN маски объектов недвижимости



(б) Маски улучшены генератором GAN

РИСУНОК 10. Улучшение масок объектов недвижимости

можно применить Гауссово размытие (**GaussianBlur**), которое смягчает резкие переходы между фоном и объектом.

Также эффективным методом является использование активных контуров (**snakes**), которые адаптивно подстраиваются под форму объекта за счёт минимизации энергии упругости и гладкости. Для достижения еще более гладких контуров может быть применена сплайн-интерполяция, которая аппроксимирует исходный контур с помощью кубических сплайнов, обеспечивая равномерное распределение точек и плавность границ.

Ещё одним из эффективных способов устранения ступенчатости является использование алгоритма Дугласа-Пекера, который специализируется на упрощении полилиний и сглаживании контуров [23]. Особенность данного алгоритма заключается в его способности анализировать каждую точку границы маски и оценивать ее значимость для общей формы объекта. Точки, которые вносят минимальный вклад в конфигурацию контура, удаляются, что позволяет получить более плавные и естественные границы без потери ключевых геометрических особенностей. Пример использования алгоритма Дугласа-Пекера приведён на рисунке 11.

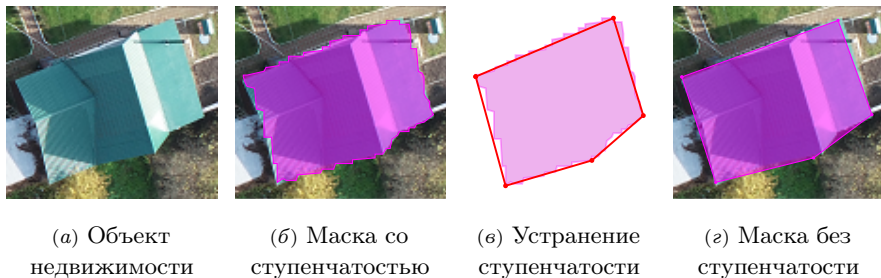


РИСУНОК 11. Устранения ступенчатости маски с использованием алгоритма Дугласа-Пекера

Эти методы или их комбинация позволяют в ряде случаев улучшить качество масок, делая их более пригодными для последующего анализа или визуализации. Как показали проведённые исследования, использования методов морфологического сглаживания и улучшения границ масок с использованием алгоритма Дугласа-Пекера в постобработке масок, даёт возможность повысить точность их формирования в среднем на 5–7%, что особенно важно для задач компьютерного зрения и сегментации объектов на аэрофотоснимках в частности.

Заключение

В данной работе была предложена и исследована модель GAN, предназначенная для повышения точности масок сегментации, полученных с использованием модели Mask R-CNN. Эксперименты показали, что применение GAN позволяет существенно улучшить качество детектированных масок объектов недвижимости за счёт коррекции ошибок, таких как разрывы в масках, шумовые артефакты и неровные контуры. Улучшение достигается за счёт использования функции потерь L1 для минимизации пиксельных различий между целевыми и сгенерированными масками, а также за счёт работы дискриминатора, который способствует более реалистичной генерации масок.

Результаты проведенных исследований демонстрируют улучшение метрик точности, таких как IoU, Precision, Recall и F1-Score, для улучшенных масок по сравнению с исходными предсказаниями Mask R-CNN. Среднее значение IoU_{0.5-95} увеличилось в среднем на 30-35%, что подтверждает эффективность предложенного подхода. Дополнительно было отмечено, что использование морфологических операций и методов сглаживания контуров позволит дополнительно улучшить гладкость и точность границ масок.

Полученные результаты имеют важное практическое значение и будут использованы в ИС ППК «Роскадастр» для автоматического преобразования растровых изображений с объектами недвижимости в векторные карты. Это позволит повысить точность и скорость формирования цифровых карт, что особенно важно для задач кадастрового учета и управления территориями. Предложенный подход может быть адаптирован для других приложений, связанных с обработкой изображений и компьютерным зрением, где требуется высокая точность сегментации объектов. Проведённое в работе исследование подтверждает перспективность применения GAN с целью улучшения качества масок объектов недвижимости, что позволяет его эффективно использовать для автоматизации процессов анализа и преобразования данных в геоинформационных системах. Дальнейшие исследования могут быть направлены на оптимизацию параметров модели, внедрение новых методов постобработки масок и расширение области применения GAN.

Список использованных источников

- [1] Винокуров И. В. *Использование модели Mask R-CNN для сегментации объектов недвижимости на аэрофотоснимках* // Программные системы: теория и приложения.– 2025.– Т. 16.– № 1(64).– С. 3–44. ↑^{137, 144}
- [2] G. Cohen, R. Giryes *Generative adversarial networks*.– 2024.– 28 pp. 2203.00667 ↑
- [3] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros *Image-to-image translation with conditional adversarial networks*.– 2016.– 17 pp. 1611.07004 ↑^{134, 138}
- [4] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro *High-resolution image synthesis and semantic manipulation with conditional GANs*.– 2017.– 14 pp. 1711.11585 ↑¹³⁵
- [5] C.-H. Lee, Z. Liu, L. Wu, P. Luo *MaskGAN: Towards diverse and interactive facial image manipulation*.– 2019.– 20 pp. 1907.11922 ↑¹³⁵
- [6] Y. Xue, T. Xu, H. Zhang, L. Rodney Long, X. Huang *SegAN: Adversarial network with multi-scale L_1 loss for medical image segmentation*.– 2017.– 9 pp. 1706.01805 ↑¹³⁵
- [7] X. Chen, C. Xu, X. Yang, D. Tao *Attention-GAN for object transfiguration in wild images*.– 2018.– 18 pp. 1803.06798 ↑¹³⁵
- [8] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros *Unpaired image-to-image translation using cycle-consistent adversarial networks*.– 2017.– 18 pp. 1703.10593 ↑¹³⁵
- [9] J. Gong, J. Xu, X. Tan, J. Zhou, Y. Qu, Y. Xie, L. Ma *Boundary-aware geometric encoding for semantic segmentation of point clouds*.– 2021.– 9 pp. 2101.02381 ↑¹³⁵
- [10] L. Xu, M. Gabbouj *Revisiting generative adversarial networks for binary semantic segmentation on imbalanced datasets*.– 2024.– 14 pp. 2402.02245 ↑¹³⁵
- [11] R. Abdelfattah, X. Wang, S. Wang *JPLGAN: Generative adversarial networks for power-line segmentation in aerial images*.– 2022.– 11 pp. 2204.07243 ↑¹³⁶
- [12] B. Benjdira, Y. Bazi, A. Koubaa, K. Ouni *Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images* // Remote Sens..– 2019.– Vol. 11.– No. 11.– id. 1369.– 23 pp. 1905.03198 ↑¹³⁶
- [13] A. Kulkarni, T. Mohandoss, D. Northrup, E. Mwebaze, H. Alemohammad *Semantic segmentation of medium-resolution satellite imagery using conditional generative adversarial networks*.– 2020.– 7 pp. 2012.03093 ↑¹³⁶
- [14] Q. H. Le, K. Youcef-Toumi, D. Tsetserukou, A. Jahanian *GAN Mask R-CNN: Instance semantic segmentation benefits from generative adversarial networks*.– 2020.– 13 pp. 2010.13757 ↑¹³⁶
- [15] A. Radford, L. Metz, S. Chintala *Unsupervised representation learning with deep convolutional generative adversarial networks*.– 2016.– 16 pp. 1511.06434 ↑¹³⁸
- [16] O. Ronneberger, Ph. Fischer, Th. Brox *U-Net: Convolutional networks for biomedical image segmentation*.– 2015.– 8 pp. 1505.04597 ↑¹³⁸

- [17] T. Karras, S. Laine, T. Aila *A style-based generator architecture for generative adversarial networks.*– 2018.– 12 pp.  [arXiv](#)  1812.04948 ↑¹³⁸
- [18] T. Karras, T. Aila, S. Laine, J. Lehtinen *Progressive growing of GANs for improved quality, stability, and variation.*– 2017.– 26 pp.  [arXiv](#)  1710.10196 ↑¹³⁸
- [19] M. Mirza, S. Osindero *Conditional generative adversarial nets.*– 2014.– 7 pp.  [arXiv](#)  1411.1784 ↑¹³⁸
- [20] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida *Spectral normalization for generative adversarial networks.*– 2018.– 26 pp.  [arXiv](#)  1802.05957 ↑¹⁴³
- [21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. C. Courville *Improved training of Wasserstein GANs.*– 2017.– 20 pp.  [arXiv](#)  1704.00028 ↑¹⁴³
- [22] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena *Self-attention generative adversarial networks.*– 2019.– 10 pp.  [arXiv](#)  1805.08318 ↑¹⁴³
- [23] H. Chen *An improved Douglas-Peucker algorithm applied in coastline generalization*, Fourth International Conference on Geology, Mapping, and Remote Sensing (ICGMRS 2023) (14–16 April 2023, Wuhan, China), Proc. SPIE.– vol. **12978**.– 2024.– id. 1297804.  ↑¹⁴⁹

Поступила в редакцию 21.04.2025;
 одобрена после рецензирования 10.06.2025;
 принята к публикации 11.06.2025;
 опубликована онлайн 28.06.2025.

Рекомендовал к публикации

д.ф.-м.н. А. М. Елизаров

Информация об авторе:



Игорь Викторович Винокуров

Кандидат технических наук (PhD), ассоциированный профессор в Финансовом Университете при Правительстве Российской Федерации. Область научных интересов: информационные системы, информационные технологии, технологии обработки данных



0000-0001-8697-1032

e-mail: igvvinokurov@fa.ru

Декларация об отсутствии личной заинтересованности: *благополучие автора не зависит от результатов исследования.*