Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

Using GAN Methods for Aerial Images Segmentation

Sara ALTUN GÜVEN[1*], Buket TOPTAŞ[2]

[1]Tarsus University, Computer Engineering Department, saraguven@tarsus.edu.tr, Orcid No: 0000-0003-2877-7105
[2]Bandırma Onyedi Eylül University, Software Engineering Department, btoptas@bandirma.edu.tr, Orcid No: 0000-0003-2556-8199

ARTICLE INFO

Article history:

Received 05 November 2023
Received in revised form 2 February 2023
Accepted 11 February 2024
Available online 29 March 2024

Keywords:

deep learning,
semantic segmentation,
aerial images,
GANs.

Doi: 10.24012/dumf.1386384

* Corresponding author

ABSTRACT

Object detection and segmentation in aerial images is currently a vibrant and significant field of research. The iSAID dataset has been created for object detection in images captured by aerial vehicles. In this study, image semantic segmentation was performed on the iSAID dataset using Generative Adversarial Networks (GANs). The compared GAN methods are CycleGAN, DCLGAN, SimDCL, and SSimDCL. All methods operate on unpaired images. DCLGAN and SimDCL methods are derived by taking inspiration from the CycleGAN method. In these methods, cost functions and network structures vary. This study thoroughly examines the methods, and their similarities and differences are observed. After semantic segmentation is performed, the results are presented using both visual and measurement metrics. Measurement metrics such as FID, KID, SCOOT, PSNR, FSIM, and SSIM are used. When looking at the metric results, the SSimDCL method ranks first with 132.62071 FID, 0.07825 KID, 0.6406 SCOOT, 0.85973 PSNR, 37.862 FSIM, and 0.82725 SSIM; the SimDCL method shows the second-best performance with 149.82306 FID, 0.10215 KID, 0.60142 SCOOT, 0.85224 PSNR, 37.4747 FSIM, and 0.82429 SSIM. The CycleGAN method, on the other hand, ranks last among the applied methods with results of 202.33857 FID, 0.16795 KID, 0.53218 SCOOT, 0.83408 PSNR, 35.7062 FSIM, and 0.7751 SSIM.Experimental studies show that SSimDCL and SimDCL methods outperform other methods in iSAID image semantic segmentation. CycleGAN method, on the other hand, is observed to be less successful compared to other methods. The aim of this study is to perform automatic semantic segmentation in aerial images.

Introduction

Aerial image, which includes images of objects captured from the air, is a popular technological tool used in various fields such as aviation, geographic information systems, and agriculture. Aerial photography, or aerial imaging, involves capturing images from an aircraft, drone, or other airborne platforms. When capturing moving images, it is referred to as aerial videography. Aerial and satellite images, known as remotely sensed images, allow for accurate mapping of land cover and enable the understanding of landscape features at regional, continental, and even global scales.

Semantic segmentation, a popular field that is accomplished with deep learning methods, is used to distinguish objects in aerial images. Semantic segmentation is the process of assigning each pixel in an image to predefined classes and is widely used in computer vision-related areas. Semantic segmentation has a extensive body of literature. Approaches to semantic segmentation can be categorized into traditional and innovative methods. Traditional segmentation methods exclusively utilize image processing techniques, while modern approaches leverage deep learning architectures. Traditional techniques for image segmentation include methods like thresholding, clustering, partial differential equation-based approaches, graph partitioning, watershed transformation, and so forth. These traditional segmentation techniques have widespread applications [1-3]. When we investigate current segmentation methods, we observe the utilization of convolutional neural networks [4-6] and Generative Adversarial Networks [7-9].

Generative Adversarial Networks (GANs), a deep learning algorithm, were proposed by Ian Goodfellow in 2014 for image synthesis [10]. GANs perform image synthesis in both supervised and unsupervised transformations. In the case of supervised transformation, it is necessary to have image pairs in two different domains. For learning the probability distribution, each input image is transformed into the corresponding output image in the other domain. In the literature, there are many studies using deep learning methods for semantic segmentation of aerial images.

In 2015, Saito et al., used CNN (Convolutional neural network) to train pixel labeling of building areas for the purpose of determining the semantic segmentation of aerial images, using Dijkstra's algorithm [11]. In 2018, Chen et al., proposed the periodic shuffling of aerial images for semantic segmentation, contributing to improving the field of view [12]. This model achieved effective results for two different datasets. In 2020, Chai et al., proposed a semantic segmentation model with Deep CNNs (DCNNs) to learn spatial context from high-resolution aerial images [13]. This model predicted distance maps to improve segmentation efficiency. In 2021, Abdollahi et al., proposed a GAN method for segmenting roads in high-resolution aerial images [14]. This model also used a modified UNet model (MUNet) to achieve satisfactory results. In 2021, Wang et al. designed a real-time semantic segmentation model for high-resolution aerial images called the Aerial-BiseNet [15]. Aerial-BiseNet used two modules called the "Feature Attention Module (FAM)" and the "Channel Attention-based Feature Fusion Module (CAFFM)" to analyze features.

In 2022, Koç and Özyurt proposed an examination of synthetic images produced with DCGAN based on the size of data and epoch [16]. The results indicated that the success of the generated fraudulent images was directly proportional to the number of data and the increase in epoch. In 2021, Şahin and Talu conducted a performance comparison of Generative Adversarial Networks (GANs) in in mustache pattern generation [17]. They utilized GAN architectures, including Pix2Pix, CycleGAN, DiscoGAN, and AttentionGAN. The study revealed that the generation speed of mustache patterns dropped below one second, while the production accuracy reached levels around 86%. In 2023, Şener and Ergen proposed "Enhancing Image Classification Performance through Discrete Cosine Transformation on Augmented Facial Images using GANs" [18]. The study found that the classification of faces could be improved by 30% compared to the normal classification model.

In 2022, Desai and Ghose suggested an active learning-based sampling strategy to select a highly representative labeled training dataset. Their proposed method resulted in a 27% improvement in mIoU with only 2% labeled data on two semantic segmentation datasets, including satellite images [19]. In 2022, Abdelfattah et al. introduced a simple yet effective method called PLGAN (Generative Adversarial Networks for Power-Line Segmentation in Aerial Images) to segment power lines from aerial images with different backgrounds. PLGAN, instead of directly using adversary networks to create segmentations, takes specific decoding features and places them in another semantic segmentation network, considering more context, geometry, and appearance information of power lines [20]. Comprehensive experiments and analyses showed that PLGAN outperformed previous state-of-the-art methods.

In this study, the iSAID dataset was used for semantic segmentation of aerial images. Recently, in 2023, Zhou et al. [21] proposed a Weakly Supervised Semantic Segmentation (WSSS) method. When dealing with Remote Sensing (RS) images with complex backgrounds and multiple categories, it can be challenging to locate and distinguish the target categories. Based on extensive experiments, their WSSS framework has shown superiority over RS datasets and has become the first WSSS framework to achieve state-of-the-art results on the iSAID dataset, exploring cross-image semantics in multi-category RS scenes using only image-level labels.

In this study, semantic segmentation of the two-dimensional iSAID images was compared among state-of-the-art GAN architectures, including CycleGAN [22], DCLGAN [23], SimDCL [23], and SSimDCL [24]. Metric comparisons revealed that the recent SSimDCL method outperformed other methods in semantic segmentation, providing more superior and satisfactory results. It was observed that this method could be used as an automatic image segmentation system.

The main contributions of this study can be summarized as follows:

- Comparison of state-of-the-art GAN architectures, including CycleGAN, DCLGAN, SimDCL, and SSimDCL, for semantic segmentation of two-dimensional iSAID images.
- Observing that the SSimDCL method provides superior and satisfactory results in semantic segmentation when compared to other methods.
- SSimDCL method has the potential to be used as an automatic image segmentation system.
- It introduces a new perspective on conducting semantic segmentation analysis for object detection in aerial images.
- The study observed the results of the new and highly accurate SSimDCL model by transforming unsupervised image segmentation methods from state-of-the-art models into supervised ones.

These contributions highlight the advancement in the field of semantic segmentation for two-dimensional aerial images using GAN architectures, particularly the effectiveness of the SSimDCL method.

The remaining organization of the article is as follows: In Section 2, materials and methods (CycleGAN, DCLGAN, SimDCL, and SSimDCL) are presented; Section 3 discusses experimental studies and their results; Section 4 provides a discussion and conclusion to conclude the article.

Materials and Methods

Used Dataset

The iSAID (A Large-scale Dataset for Semantic Segmentation in Aerial Images) is a dataset [25]. Existing Earth Vision datasets are suitable for semantic segmentation or object detection. iSAID, for instance, is the first benchmark dataset for segmentation in aerial images. This large-scale and densely annotated dataset contains 655,451 object instances across 15 categories in 2,806 high-resolution images. The distinctive features of the iSAID dataset are as follows:

- A large number of images with high spatial resolution,
- 15 important and commonly occurring categories,
- Numerous examples per category,
- A substantial number of labeled examples per image that can assist in learning contextual information,
- Significant variations in object scale, including small, medium, and large objects within the same image,
- An unequal distribution of objects in different directions in images depicting real-world weather conditions,
- A few small-sized objects with uncertain appearances that can only be resolved through contextual reasoning,
- Precise annotations are available at the example level, cross-checked and verified by expert annotators following well-defined guidelines, and conducted by professional annotators.

These distinctive features make the iSAID dataset a valuable resource for semantic segmentation and object detection in aerial images.

In this study, the semantic segmentation images shown in Fig. 1 were utilized. The training dataset consisted of 1302 real images and 1302 corresponding ground truth images used for semantic segmentation. The test dataset included 109 real images, and it was used to obtain the results of semantic segmentation. These are 109 images captured from the air. The aerial images consist of a variety of scenes including roads, vehicles, rivers, airports, and seaports. The segmented results mainly focus on vehicles such as cars, trucks, and ships. Randomly selected training and test data were used in the iSAID dataset. The aim of this article is to perform automatic segmentation on the dataset.The image dimensions were resized to $256 \times 256 \times 3$ within the code for analysis.

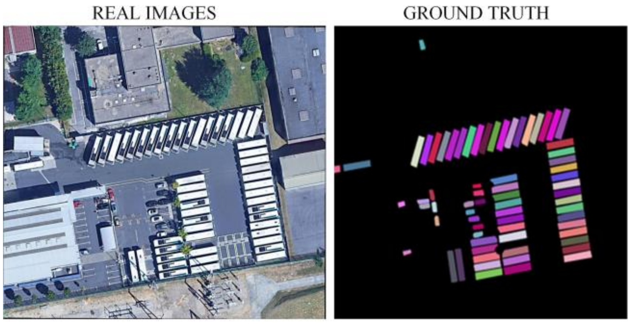

Figure 1. iSAID dataset image sample

Training Details

In the DCLGAN, SimDCL, and SSimDCL architectures, there are two generator networks, two feature-extracting layers, and two discriminator networks. In the CycleGAN architecture, there are two generator networks and one discriminator network. In this study, the internal structures of these methods include the same network content. The difference lies in the properties and numbers of the networks present in each method.

Fig. 2 illustrates the internal structure of the discriminator network. The discriminator network employs the PatchGAN architecture. Fig. 3 depicts the network structure used for feature extraction and embedding of the image in the DCLGAN, SimDCL, and SSimDCL methods. Fig. 4 provides an overview of the generator network's structure and network layers.

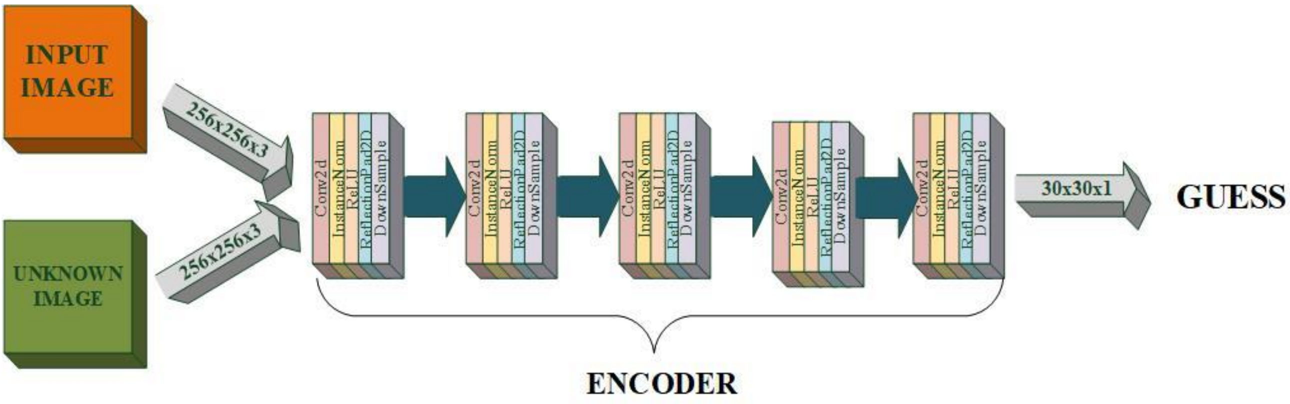

Figure 2. Discriminator Network Architecture

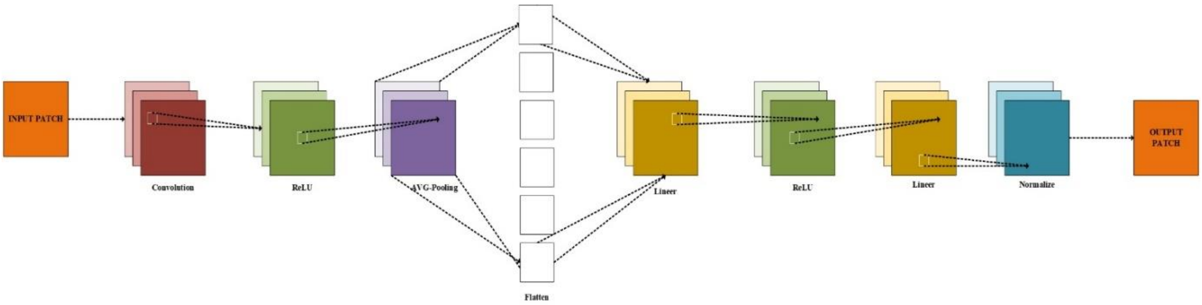

Figure 3. Feature Extracting Embedding Network Architecture

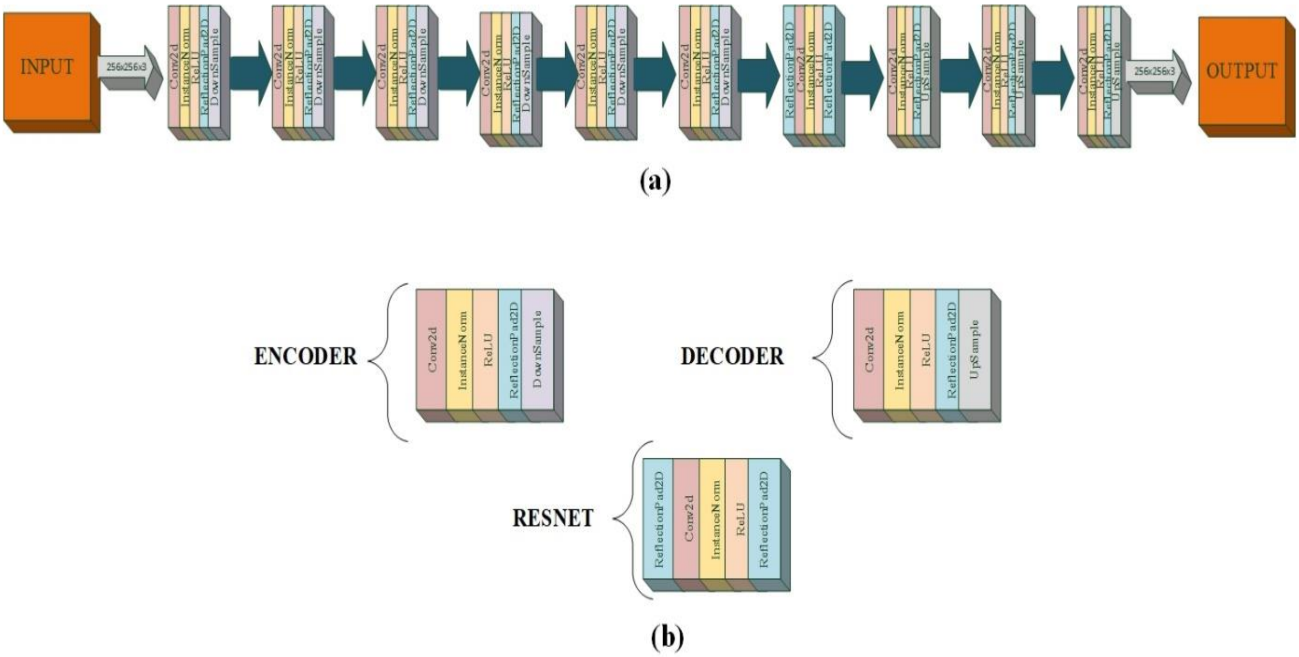

Figure 4. Generator Network Architecture (a) Inner Structure of the Network (b) Representation of Network Layers in the Inner Structure.

Image Segmentation Architectures

In this study, new segmentation architectures, namely CycleGAN [22], DCLGAN [23], SimDCL [23], and SSimDCL [23], have been utilized as segmentation architectures and extensively explained.

CycleGAN

CycleGAN [22] is a Generative Adversarial Network (GAN) architecture that employs bidirectional transformation. It utilizes two separate generative networks, denoted as $G: X \to Y$ and $F: Y \to X$, to convert an input image from the X domain into an output image in the Y domain, and vice versa. In the training process of this architecture, instance normalization is applied in lieu of batch normalization, and the generator network's internal design incorporates ResNET blocks. Figure 6 provides a visual representation of the overall structure of CycleGAN.

The loss function of CycleGAN consists of two distinct loss components: 1) cyclic consistency loss and 2) identity loss.

Cyclic Consistency Loss: When computing the cyclic loss, transformations$X \to \mathrm{G(X)} \to Y' \to \mathrm{F}(Y') \to \hat{X}$ and $Y \to \mathrm{F(Y)} \to X' \to \mathrm{G}(X') \to \hat{Y}$ are executed, and the aim is to minimize the summation of the difference values $X - \hat{X}$ and $Y - \hat{Y}$ The calculation of the cyclic loss is detailed in Table 1.

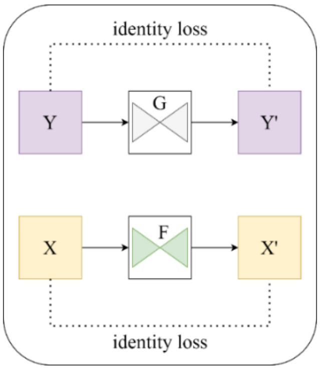

Figure 5. Identity Loss [24]

Identity Loss: When computing the transformations $X \rightarrow \hat{X}$ and $Y \rightarrow \hat{Y}$, intermediate outputs $X \rightarrow Y'$ and $Y \rightarrow X'$ are created. The purpose of these intermediate outputs is to closely resemble the original images. To perform this operation, it utilizes the F and G generative networks. Figure 5 illustrates the concept of identity loss. The identity loss is formulated as described in Table 1.

Furthermore, the errors in the discriminator architectures within the generative networks ($\mathcal{L}_{GAN}^{G}$ and $\mathcal{L}_{GAN}^{F}$) are computed, contributing to the formation of the target function as detailed in Table 1 [22].

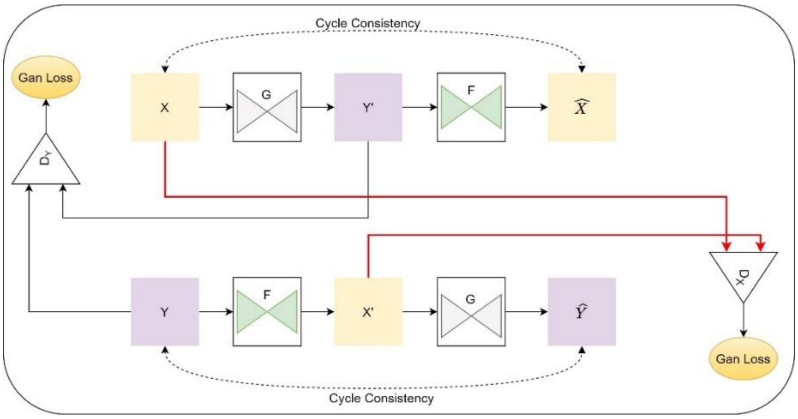

Figure 6. CycleGAN Architecture [24]

DCLGAN and SimDCL

DCLGAN [23] aims to enhance feature extraction between input and output image patches by utilizing two separate embedded systems to maximize mutual information. DCLGAN [23] seeks to maximize mutual information by improving feature extraction between input and output image patches through the use of two distinct embedded systems. The stability of this approach is enhanced through binary learning training. Certain design decisions for mutual learning have been assessed. In the implementation of the PatchNCE loss. The removal of RGB pixels corresponding to tiny patches has led to enhanced results. It has been demonstrated that enforcing cycle consistency is unnecessary. SimDCL [23] is a variant of DCLGAN that effectively mitigates mode collapse.

DCLGAN utilizes adversarial loss, identity loss, and patch-wise noise-contrastive estimation (PatchNCE) loss as its loss components. Furthermore, SimDCL incorporates similarity loss ($\mathcal{L}_{sim}$) to prevent mode collapse. The essential target function for DCLGAN is presented in Table 1.

A similarity loss has been integrated into the target function of DCLGAN, leading to the name SimDCL. In this context, "sim" signifies the similarity loss, while "DCL" signifies binary comparative learning. SimDCL incorporates this similarity loss ($\mathcal{L}_{sim}$)alongside the existing loss functions to mitigate mode collapse

Similarity loss: In essence, images originating from the same domain exhibit certain resemblances. These images may possess distinct semantics but still exhibit a common stylistic element. In binary learning, there is one genuine and one synthetic image within the same field. There are two domains, denoted as X and Y. In brief, the architecture comprises a total of two genuine and two synthetic images.

Utilizing a similarity loss on deep features promotes a resemblance between the generated images and real images at the deep feature level. This, in turn, makes the generated images more lifelike." In Table 1, the target function for SimDCL is provided. Figure 7 depicts the DCLGAN and SimDCL architectures.

SSimDCL

SSimDCL [24] utilizes two embedded systems To improve mutual information. In place of using unmatched images during the training process, matched images are employed. The objective is to transform the architecture into a supervised and matched state. To accomplish this, an L_1 metric is introduced between real and generated images within the SSimDCL framework.

SSimDCL [24] employs adversarial loss, identity loss, and patch-wise noise-contrastive estimation (PatchNCE) loss, which is also applied in the CUT method, as loss functions. It also incorporates a similarity loss ($\mathcal{L}_{sim}$), similar to SimDCL, to mitigate mode collapse. Moreover, in contrast to other techniques, it includes an L_1loss to measure the difference between real and generated images.

When looking at the training results of the SSimDCL method, it is observed that the identity results in higher-resolution generated images compared to real images [24]. The method that is transformed into supervised learning and works on matched images closely resembles real images.

$\mathbf{L_1}$ loss: In order to add a more supervised aspect to the unsupervised system, it is computed between the genuine image and the generated image. The specific calculations for the L_1 loss can be found in Table 1. The target function for the SSimDCL method is also presented in Table 1.

Fig. 8 illustrates the architecture of SSimDCL. SSimDCL results in less pixel loss and higher resolution in the images generated with $\mathcal{L}_{identity}(G,F)$ loss. The SSimDCL method has been employed to create a new dataset using the images generated with $\mathcal{L}_{identity}(G,F)$ loss.

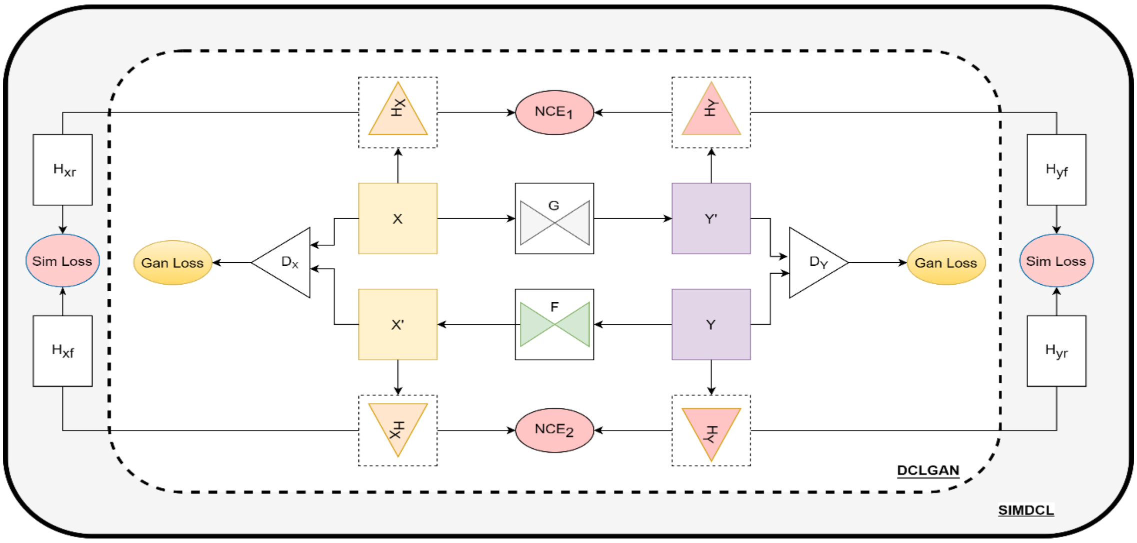

Figure 7. DCLGAN and SimDCL architecture [24]

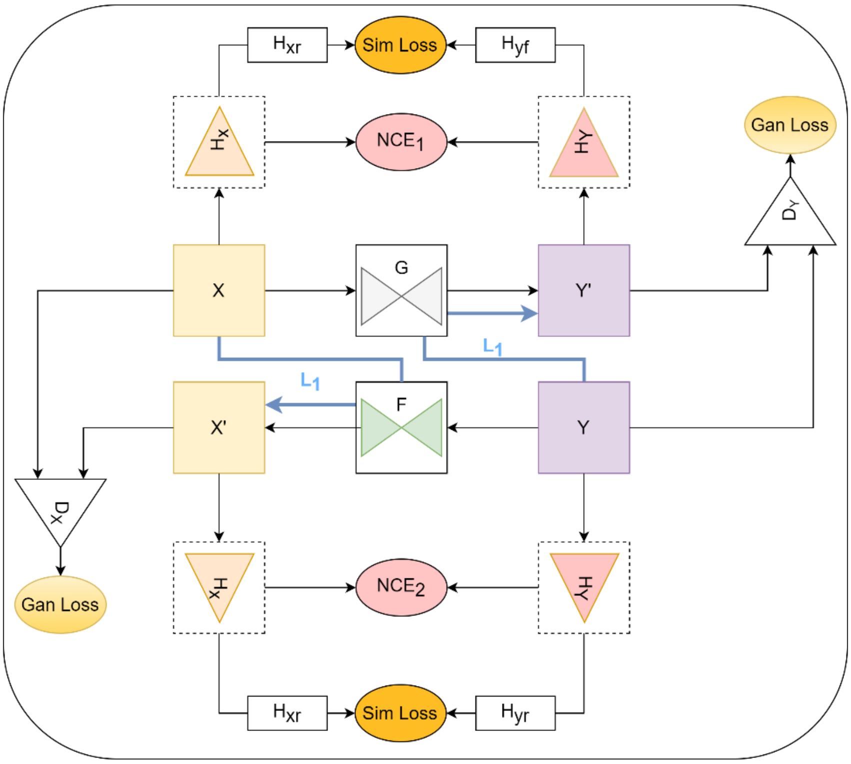

Figure 8. SSimDCL architecture[24]

Table 1. GAN Methos and Target Function

GAN methods	Target Function
CYCLEGAN [22]	$\mathcal{L}_{cyc} = \lvert X - \hat{X} \rvert + \lvert Y - \hat{Y} \rvert$ $\mathcal{L}_{identity} = L_1(G(Y) - Y) + L_1(F(X) - X)$ $G^*, F^* = \underset{G,F}{argmin}\,\underset{D_X,D_Y}{max}\,\mathcal{L}(G, F, D_X, D_Y)$ $= \mathcal{L}_{GAN}^{G}(G, D_Y, X, Y) + \mathcal{L}_{GAN}^{F}(F, D_X, Y, X)$ $+\lambda\left(\mathcal{L}_{cyc}(G, F, X, Y) + \mathcal{L}_{identity}(G, F, X, Y)\right)$
DCLGAN [23]	$G^*, F^* = \underset{G,F}{argmin}\,\underset{D_X,D_Y}{max}\,\mathcal{L}(G, F, D_X, D_Y, H_X, H_Y)$ $= \lambda_{GAN}\left(\mathcal{L}_{GAN}^{G}(G, D_Y, X, Y) + \mathcal{L}_{GAN}^{F}(F, D_X, Y, X)\right)$ $+ \lambda_{NCE}\left(\mathcal{L}_{PatchNCE_X}(G, H_X, H_Y, X)\right.$ $\left.+ \mathcal{L}_{PatchNCE_Y}(F, H_X, H_Y, Y)\right)$ $+ \lambda_{idt}(\mathcal{L}_{identity}(G, F))$
SimDCL [23]	$G^*, F^* = \underset{G,F}{argmin}\,\underset{D_X,D_Y}{max}\,\mathcal{L}(G, F, D_X, D_Y, H_X, H_Y)$ $= \lambda_{GAN}\left(\mathcal{L}_{GAN}^{G}(G, D_Y, X, Y) + \mathcal{L}_{GAN}^{F}(F, D_X, Y, X)\right)$ $+ \lambda_{NCE}\left(\mathcal{L}_{PatchNCE_X}(G, H_X, H_Y, X)\right.$ $\left.+ \mathcal{L}_{PatchNCE_Y}(F, H_X, H_Y, Y)\right)$ $+ \lambda_{sim}\mathcal{L}_{sim}\left(G, F, H_X, H_Y, H_{xr}, H_{xf}, H_{yr}, H_{yf}\right)$ $+ \lambda_{idt}\mathcal{L}_{identity}(G, F)$
SSimDCL [24]	$G^*, F^* = \underset{G,F}{argmin}\,\underset{D_X,D_Y}{max}\,\mathcal{L}(G, F, D_X, D_Y, H_X, H_Y)$ $= \lambda_{GAN}\left(\mathcal{L}_{GAN}^{G}(G, D_Y, X, Y) + \mathcal{L}_{GAN}^{F}(F, D_X, Y, X)\right)$ $+ \lambda_{NCE}\left(\mathcal{L}_{PatchNCE_X}(G, H_X, H_Y, X)\right.$ $\left.+ \lambda_{NCE}\mathcal{L}_{PatchNCE_Y}(F, H_X, H_Y, Y)\right)$ $+ \lambda_{sim}\mathcal{L}_{sim}\left(G, F, H_X, H_Y, H_{xr}, H_{xf}, H_{yr}, H_{yf}\right)$ $+ \lambda_{idt}\left(\mathcal{L}_{identity}(G, F) + \mathcal{L}_{identity}(G, F, X, Y)\right)$

Here $\lambda_{GAN} = 1$, $\lambda_{NCE} = 2$ and $\lambda_{sim} = 10$ and $\lambda_{idt} = 1$ are the hyperparameters used in the method.

Image Quality Metric

Inception Distance (FID) [26] assesses the similarity between the distribution of real images and that of generated images.

Kernel Inception Distance (KID) [27] is akin to FID but relies on the Mean Squared Error (MSE) between the generated and genuine images. KID offers an advantage over FID as it incorporates the ReLU activation function.

Feature Similarity Index Measurement (FSIM) [28] compares the phase consistency and gradient magnitude features of image pairs.

The Structural Similarity Index Metric (SSIM) [29] uses several simple statistical moments such as the mean (μ) and standard deviation (σ) of image pairs to obtain a similarity score.

Peak Signal-to-Noise Ratio (PSNR) [30] is the prevailing objective measurement for assessing the quality of image signals. However, PSNR values do not correlate well with perceived image quality due to the complex, highly nonlinear nature of the human visual system.

Structure Co-Occurrence Texture (SCOOT) [31] uses the Scoot metric to measure the similarity between real and synthesized images. SCOOT provides results that are very close to human perception. It systematically evaluates different texture-based/edge-based features in the Scoot architecture.

In terms of measurement metrics, higher values indicate better results for SCOOT [31], FSIM [28], SSIM [29], and PSNR [30]. For FID [26] and KID [27], lower values indicate better results.

Application Environment and Experimental Setup

The application was trained on an NVIDIA® GeForce® RTX 4060 Max-Performance 8GB GDDR6 128-Bit DX12 graphics card with a power of 115 watts + 25 watts for Dynamic Boost 2.0. The system also used an Intel® Raptor Lake Core™ i7 processor.

Work has been conducted on the Python programming language using PyCharm and Anaconda IDEs. Within the environment, Python 3.7 and the following libraries have been utilized: torch, torchvision, dominate, visdom, packaging, GPUtil, scipy, Pillow, and numpy.

All the methods were run in a Python environment with 300 iterations on a computer with an 8 GB GPU. The measurement metrics used in the evaluation included the classic GAN methods FID [26] and KID [27], and the traditional method FSIM [28], SSIM [29], PSNR [30], SCOOT [31].

For training GAN methods, the settings of DCLGAN [23] were used as a reference. The hyperparameters used in these settings are shown in Table 2. The training process utilized the Adam optimization method [32]. The generative network was based on ResNet [33], and a PatchGAN [34] discriminator was used. Semantic normalization was also applied.

Table 2. Training Details

	β_1	β_2	Epoch	Lr	Batch Size	Image Size
Training hyperparameter	0.5	0.999	300	0.0001	1	256×256

Experimental Results

Fig. 9 displays the visual results of CycleGAN, DCLGAN, SimDCL, and SSimDCL methods in the context of iSAID semantic segmentation. When comparing the visual output results of semantic segmentation on iSAID images, it is observed that SimDCL and SSimDCL methods provide the best results.

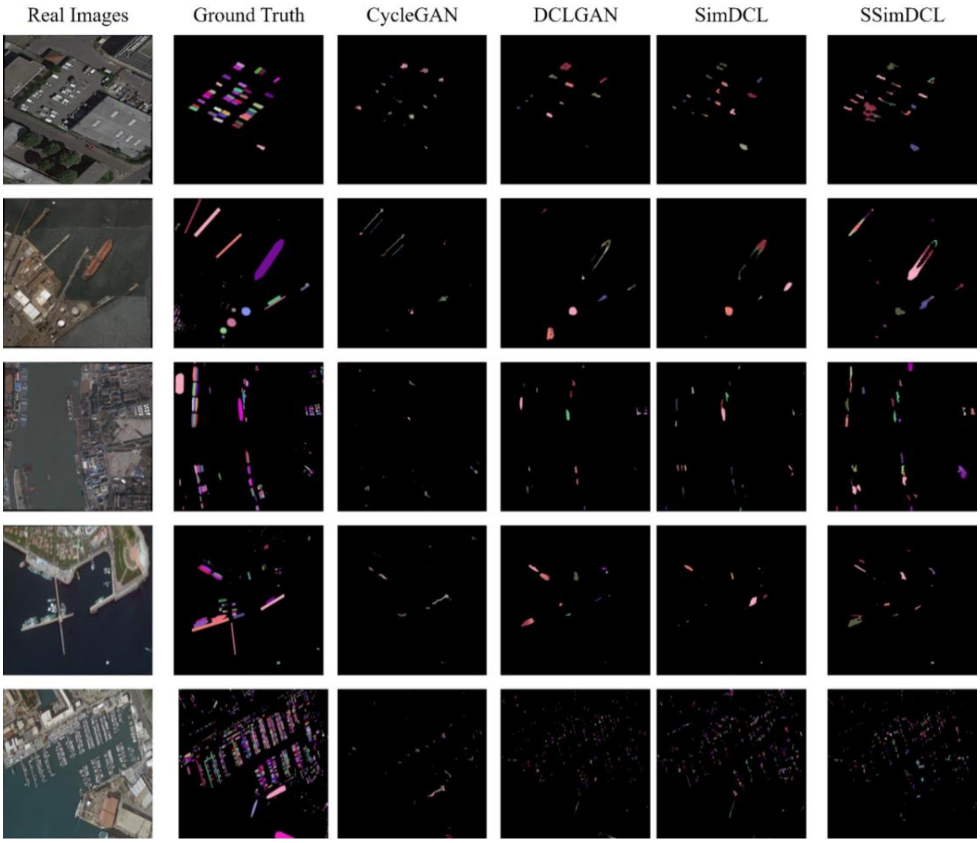

Figure 9. Visual results of the methods for semantic segmentation

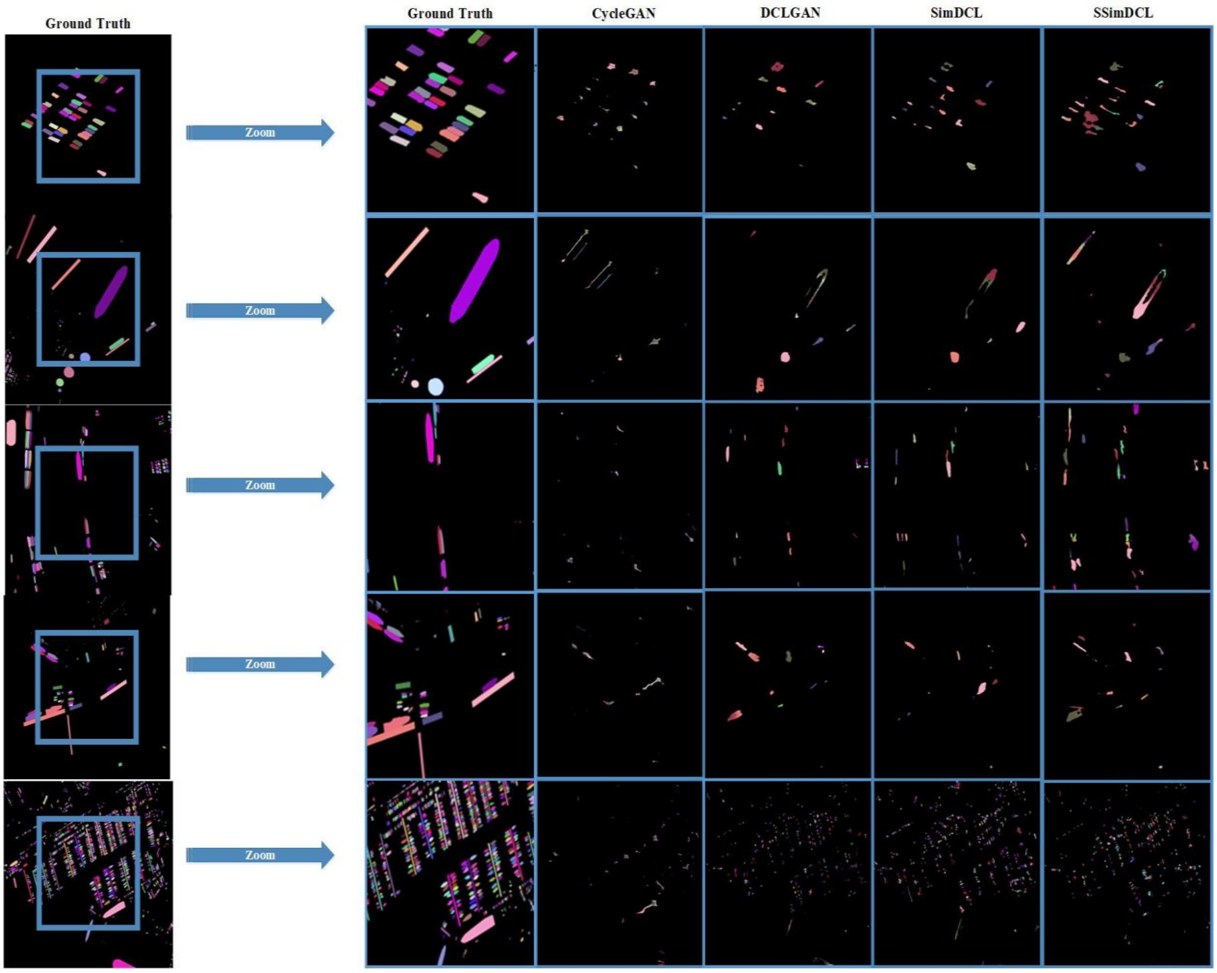

Figure 10. iSAID image semantic segmentation analysis according to the methods. Original image size $(256 \times 256 \times 3)$. , Zoomed-in view of the colored regions shown in first column.

Table 3. Image Similarity Metrics Results

	FID↓	KID↓	SCOOT↑	PSNR↑	SSIM↑	FSIM↑
CYCLEGAN[22]	202.33857	0.16795	0.53218	0.83408	35.7062	0.7751
DCLGAN [23]	164.83025	0.12890	0.5980	0.85089	37.5326	0.82128
SimDCL [23]	149.82306	0.10215	0.60142	0.85224	37.4747	0.82429
SSimDCL [24]	**132.62071**	**0.07825**	**0.6406**	**0.85973**	**37.862**	**0.82725**

Table 3 shows the bold font indicating the first successful method, and the underlined font indicating the second successful method. Metrically, the lower the FID and KID, the higher the performance rate. Among the examined methods, the SSimDCL method achieved the highest success with the lowest FID of 132.32071 and KID of 0.07825; the SimDCL method secured the second position with FID of 149.82306 and KID of 0.10215, while the CycleGAN method was observed to be the least successful with FID of 202.33857 and KID of 0.16795.

In terms of SCOOT, PSNR, SSIM, and FSIM, the higher the values, the better the performance. Among the examined methods, the SSimDCL method ranked first with the lowest

0.6406 SCOOT, 0.85973 PSNR, 37.862 SSIM, and 0.82725 FSIM; the SimDCL method secured the second position with 0.60142 SCOOT, 0.85224 PSNR, 37.4747 SSIM, and 0.82429 FSIM. The CycleGAN method, with 0.53218 SCOOT, 0.83408 PSNR, 35.7062 SSIM, and 0.7751 FSIM, was observed to be the least successful among the examined methods.

When examining Table 3, it can be observed that according to the image similarity metrics FID, KID, SCOOT, PSNR, SSIM, and FSIM, the method most similar to the Ground Truth image is SSimDCL (indicated in bold). SimDCL (indicated with an underline) follows in the second place, and the DCLGAN method ranks third. When looking at the image measurement metric results, the CycleGAN method appears to be less successful compared to the other methods.

Discussion and Conclusion

The main subject of this article is to compare the efficiency of methods that initially use the CycleGAN method, which is commonly used in the problem of image semantic segmentation. The CycleGAN method has been modified to develop new GAN architectures, namely DCLGAN and SimDCL. The SSimDCL method is derived from the SimDCL method. The iSAID dataset is used for evaluating the visual semantic segmentation efficiency of CycleGAN, DCLGAN, SimDCL, and SSimDCL methods.

When looking at the metric results, the SSimDCL method ranks first with 132.62071 FID, 0.07825 KID, 0.6406 SCOOT, 0.85973 PSNR, 37.862 FSIM, and 0.82725 SSIM; the SimDCL method shows the second-best performance with 149.82306 FID, 0.10215 KID, 0.60142 SCOOT, 0.85224 PSNR, 37.4747 FSIM, and 0.82429 SSIM. The CycleGAN method, on the other hand, ranks last among the applied methods with results of 202.33857 FID, 0.16795 KID, 0.53218 SCOOT, 0.83408 PSNR, 35.7062 FSIM, and 0.7751 SSIM.

When examining the visual results of the compared methods, it is observed that SimDCL and SSimDCL methods achieved the best results. CycleGAN, on the other hand, was found to be less effective in segmentation compared to the other methods. Looking at the results with image evaluation metrics in Table 3, it can be observed that SSimDCL and SimDCL methods have produced the best results, in that order. According to Table 3, it can be observed that the CycleGAN method received lower metric results compared to the other methods. As a result of this study, it can be said that SSimDCL and SimDCL can be used for instance segmentation and achieve more efficient results.

References

[1] R. S. A. V. K. V. Shrimali, "Current trends in segmentation of medical ultrasound B-mode images: A review," IETE Tech. Rev., cilt 1, no. 817, ss. 26, 2009.

[2] G. Hu and Mageras, "Survey of recent volumetric medical image segmentation techniques," Biomedical Engineering, Vukovar, Croatia: In-Tech, ss. 3216, 2009.

[3] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," BMC medical imaging, cilt 15, no. 1, ss. 1-28, 2015.

[4] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen et al., "Interactive medical image segmentation using deep learning with image-specific fine tuning," IEEE transactions on medical imaging, cilt 37, no. 7, ss. 1562-1573, 2018.

[5] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao et al., "Ce-net: Context encoder network for 2d medical image segmentation," IEEE transactions on medical imaging, cilt 38, no. 10, ss. 2281-2292, 2019.

[6] B. Kayalibay, G. Jensen and P. van der Smagt, "CNN-based segmentation of medical imaging data," arXiv preprint arXiv:1701.03056, 2017.

[7] Y. Xue, T. Xu, H. Zhang, L. R. Long, X. Huang, "SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation," Neuroinformatics, cilt 16, ss. 383–392, 2018.

[8] N. Khosravan, A. Mortazi, M. Wallace, and U. Bagci, "Pan: Projective adversarial network for medical image segmentation," Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, ss. 68–76, 2019.

[9] M. Zhao, L. Wang, J. Chen, D. Nie, Y. Cong, S. Ahmad et al., "Craniomaxillofacial bony structures segmentation from MRI with deep-supervision adversarial learning," Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, ss. 720–727, 2018.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair et al., "Generative adversarial nets," Advances in neural information processing systems, ss. 2672-2680, 2014.

[11] S. Saito, R. Arai and Y. Aoki, "Seamline determination based on semantic segmentation for aerial image mosaicking," IEEE Access, cilt 3, ss. 2847–2856, 2015.

[12] B. Yu, L. Yang and F. Chen, "Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, cilt 11, no. 9, ss. 3252–3261, 2018.

[13] D. Chai, S. Newsam and J. Huang, "Aerial image semantic segmentation using DCNN predicted distance maps," ISPRS Journal of Photogrammetry and Remote Sensing, cilt 161, ss. 309–322, 2020.

[14] A. Abdollahi, B. Pradhan, G. Sharma, K. N. A. Maulud and A. Alamri, "Improving road semantic segmentation using generative adversarial network," IEEE Access, cilt 9, ss. 64381–64392, 2021.

[15] F. Wang, X. Luo, Q. Wang and L. Li, "Aerial-BiSeNet: A real-time semantic segmentation network

for high resolution aerial imagery," Chinese Journal of Aeronautics, cilt 34, no. 9, ss. 47–59, 2021.

[16] C., KOÇ., &, F. Özyurt., An examination of synthetic images produced with DCGAN according to the size of data and epoch. Firat University Journal of Experimental and Computational Engineering, 2(1), 32-37, 2023.

[17] E. Şahin,, & Talu, M. F. Talu, Bıyık Deseni Üretiminde Çekişmeli Üretici Ağların Performans Karşılaştırması. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 10(4), 1575-1589, 2022.

[18] A., ŞENER, & B. ERGEN, Enhancing Image Classification Performance through Discrete Cosine Transformation on Augmented Facial Images using GANs. Computer Science, (IDAP-2023), 7-18, 2023.

[19] S. Desai and D. Ghose, "Active learning for improved semi-supervised semantic segmentation in satellite images," Proceedings of the IEEE/CVF winter conference on applications of computer vision, ss. 553-563, 2022.

[20] R. Abdelfattah, X. Wang and S. Wang, "Plgan: Generative adversarial networks for power-line segmentation in aerial images," arXiv preprint arXiv:2204.07243, 2022.

[21] R. Zhou, Z. Yuan, X. Rong, W. Ma, X. Sun, K. Fu et al., "Weakly Supervised Semantic Segmentation in Aerial Imagery via Cross-Image Semantic Mining," Remote Sensing, cilt 15, no. 4, ss. 986, 2023.

[22] J. Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," Proceedings of the IEEE international conference on computer vision, ss. 2223-2232, 2017.

[23] J. Han, M. Shoeiby, L. Petersson and M. A. Armin, "Dual Contrastive Learning for Unsupervised Image-to-Image Translation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, ss. 746-755, 2021.

[24] S. A. Güven and M. F. Talu, "Brain MRI high resolution image creation and segmentation with the new GAN method," Biomedical Signal Processing and Control, cilt 80, ss. 104246, 2023.

[25] S. Waqas Zamir, A. Arora, A. Gupta, S. Khan, G. Sun, F. Shahbaz Khan et al., "iSAID: A large-scale dataset for instance segmentation in aerial images," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, ss. 28-37, 2019.

[26] M. Heusel, H. Ramsauer, T. Unterthiner et al., "Gans trained by a two time-scale update rule converge to a local nash equilibrium," Advances in neural information processing systems, cilt 30, 2017.

[27] M. Bińkowski, D. J. Sutherland, M. Arbel and A. Gretton, "Demystifying mmd gans," arXiv preprint arXiv:1801.01401, 2018.

[28] L. Zhang, L. Zhang, X. Mou et al., "FSIM: A feature similarity index for image quality assessment," IEEE transactions on Image Processing, cilt 20, no. 8, ss. 2378-2386, 2011.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh et al., "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing, cilt 13, no. 4, ss. 600-612, 2004.

[30] PSNR (Peak Signal-to-Noise Ratio), IEEE transactions on Image Processing, cilt 20, no. 8, ss. 2378-2386, 2011.

[31] D. P. Fan, S. C. Zhang, Y. H. Wu et al., "Scoot: A perceptual metric for facial sketches," Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, ss. 5612-5622.

[32] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," International Conference on Learning Representations (ICLR), 2014.

[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, "Deep residual learning for image recognition," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), sayfalar 770-778, 2016.

[34] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros, "Image-to-image translation with conditional adversarial networks," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.