International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

DEEP LEARNING-BASED IMAGE FUSION FOR HIGH-RESOLUTION REMOTE SENSING APPLICATIONS

Anita Chaudhary[1,2], Dr. Navdeep Kaur[3]

[1]Department of Computer Science, Sri Guru Granth Sahib World University, Fatehgarh Sahib (anitachaudhary111@gmail.com)

[2]Department of Computer Science, Akal College of Engineering & Technology, Eternal University, Baru Sahib (anita@eternaluniversity.edu.in)

[3]Department of Computer Science, Sri Guru Granth Sahib World University, Fatehgarh Sahib (drnavdeep.sggswu@gmail.com)

Abstract

Since it can integrate multimodal data and improve spatial, spectral, and temporal resolution, image fusion has become a key tool in remote sensing. Although they have been useful, traditional picture fusion techniques like Wavelet Transform with Principal Component Analysis (PCA) sometimes have trouble maintaining both spatial features and spectral accuracy. Transformer-based models, Generative Adversarial Networks (GANs), and Convolutional Neural Networks (CNNs) are among the deep learning-based techniques that have recently emerged has revolutionized image fusion by offering improved feature extraction, noise reduction, and information retention. This research investigates the advancements in deep learning-driven image fusion techniques, comparing them with conventional methods in terms of accuracy, computational effectiveness, and real-world applicability. The results highlight the efficiency of the suggested approach in current research problem in given context.

Keywords: Deep learning, convolutional neural networks, generative adversarial networks, image fusion, and remote sensing

1. Introduction

The ability to see and analyze the Earth's surface without making direct physical touch has made remote sensing an essential tool in many scientific and economic domains. The availability of remote sensing data has increased exponentially due to the quick advancements in sensor technology and data collection methods. However, restrictions in spatial, spectral, and temporal resolutions frequently make it difficult to use this data effectively. picture fusion has become a potent strategy for overcoming these obstacles, combining data from several picture sources to provide a more thorough and accurate depiction of the seen world [1].

1.1 Background on Remote Sensing and Image Fusion

Because it allows for the capture of vital information about the Earth's surface without direct touch, distant detection has revolutionized a number of scientific and economic domains. Applications such as catastrophe management, urban planning, environmental monitoring,

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
and military surveillance are made easier by remote sensing, which uses satellite and airborne sensors. However, individual sensors often suffer from limitations in spatial, spectral, or temporal resolution, making it necessary to develop advanced techniques like image fusion to enhance data quality and interpretation [2].

Image fusion is a crucial technique that integrates information from multiple image sources to provide a single, more instructive picture. It improves spatial resolution, enhances spectral integrity, and minimizes uncertainty in image interpretation, making it a useful instrument for applications involving remote sensing [3]. Conventional techniques for picture fusion, such as Wavelet Transform and Principal Component Analysis (PCA), have played a pivotal role in combining information from different spectral bands and sensor modalities. However, these methods often struggle with maintaining both spatial and spectral fidelity, leading to the emergence of deep learning-based fusion approaches that offer superior performance [4].

High-resolution imagery is particularly important for information extraction in remote sensing, as it provides finer details for categorization of land cover, analysis of change detection, and object detection. The accuracy of remote sensing analysis has been further increased by researchers' effective reconstruction of high-resolution pictures from low-resolution inputs thanks to developments in super-resolution methods [5]. The capacity to derive significant insights from remotely sensed pictures has also improved thanks to methods for digital image processing such as feature extraction, segmentation, and edge detection [6].

1.2 Importance of High-Resolution Images in Information Extraction

The geographical, spectral, and temporal resolutions of remote sensing data significantly impact its quality of the acquired images. High-resolution images provide finer details, which are crucial for various applications, including urban infrastructure planning, environmental sustainability, and precision agriculture [7]. For example, flood monitoring and damage assessment rely heavily on high-resolution satellite imagery to detect inundated areas and evaluate infrastructure damage [2]. Similarly, super-resolution methods improve the clarity of images used for facial recognition, surveillance, and other security applications [3].

The compromise between spatial resolution and spectral resolution is among the main issues in remote sensing picture processing. Panchromatic images offer high spatial resolution but lack spectral diversity, whereas multispectral and hyperspectral images provide rich spectral information but suffer from lower spatial resolution [8]. Image fusion techniques bridge this gap by integrating spatial and spectral details from different sources, ensuring that both spatial sharpness and spectral integrity are preserved [9].

With the growing need for accurate and timely information, the demand for automated and efficient image fusion methods has increased. When it comes to extracting and combining characteristics from remote sensing photos, Two deep learning-based methods, Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), have shown improved performance. [10]. The ability of these models to pick up hierarchical

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
representations greatly improves the precision of tasks involving item recognition and categorization in high-resolution images [11].

1.3 Research Objectives

This study's primary goal is to evaluate the developments in deep learning-based image fusion techniques and compare their performance with traditional fusion methods. The key objectives of this research include:

1. **Reviewing Traditional Image Fusion Techniques:** Examining conventional fusion methods such as Wavelet Transform transformation, PCA, and Intensity-Hue-Saturation (IHS), assessing their strengths and limitations.
2. **Analysing Deep Learning-Based Image Fusion Approaches:** Investigating the effectiveness of CNNs, GANs, and Transformer-based models in enhancing picture fusion for distant sensing.
3. **Evaluating the Role of High-Resolution Image Fusion in Various Applications:** Exploring how image fusion contributes to urban planning, disaster management, environmental monitoring, and defense applications.
4. **Identifying Challenges and Future Research Directions:** Discussing important issues including model interpretability, data accessibility, and computational complexity, along with potential future improvements in image fusion methodologies.

2. Related Work

Over time, the area of image fusion has seen tremendous change, moving from conventional statistical and mathematical techniques to sophisticated deep learning-based methods. Integrating information from numerous sources to produce a more informative, higher-resolution image that enhances spectral and spatial quality is the main goal of image fusion. Conventional techniques like Principal Component Analysis (PCA), Wavelet Transform, and Intensity-Hue-Saturation (IHS) transformation have been extensively employed in remote sensing for image fusion. Although these methods have proven effective in combining pictures from several sources, they often struggle with spectral distortion, information loss, and reduced feature retention.

Deep learning has revolutionized picture fusion by using neural networks to automatically remove the complex spatial and spectral characteristics. Convolutional neural networks, or CNNs, have been essential in enhancing fusion quality by learning hierarchical representations of input images. Additionally, GANs, or Generative Adversarial Networks, have enabled the possibility to generate highly realistic fused images by employing adversarial training mechanisms. More recently, Transformer-based models have gained traction, offering superior feature attention and global context learning capabilities, further enhancing the accuracy and effectiveness of image fusion.

This section offers an extensive analysis of the body of research on both conventional and deep learning-based image fusion techniques. It examines the strengths and limitations of

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
various approaches, comparing their effectiveness in different remote sensing applications. By analyzing previous research, this study identifies gaps in current methodologies and highlights potential improvements that can be integrated into future fusion frameworks.

The integration of various picture sources to enhance spatial, spectral, and temporal resolution has been made possible by image fusion, a crucial approach in remote sensing. Conventional techniques for image fusion, such as Intensity-Hue-Saturation (IHS), Wavelet Transform, and Principal Component Analysis (PCA) transformation, have been widely used for enhancing image quality. These methods are computationally efficient and have been successfully utilized in satellite and aerial pictures. However, they often struggle with maintaining a balance between spatial and spectral integrity, leading to the emergence of more advanced fusion techniques **(Ha et al., 2018) [13].**

Deep learning-based image fusion techniques have grown in popularity recently because of their capacity to extract hierarchical representations from data. CNNs, or convolutional neural networks have established themselves as quite effective in image fusion tasks, providing improved feature extraction and spatial enhancement **(Zhang et al., 2018) [14].** Generative Adversarial Networks (GANs) have further advanced image fusion by generating high-quality, realistic images that preserve fine details while reducing artefacts **(Ravì et al., 2018) [15].** These deep learning models outperform traditional techniques by leveraging large datasets and learning complex patterns that are difficult to capture using conventional statistical methods **(Liang & Wang, 2019) [16].**

2.2 Comparison of Existing Approaches

Traditional image fusion methods rely on mathematical transformations and statistical principles to integrate multi-source images. PCA reduces redundancy by transforming correlated spectral bands into uncorrelated components, but it often leads to information loss **(Yang et al., 2019) [17].** Wavelet Transform decomposes images into frequency components, providing multi-resolution analysis, yet it introduces artefacts when reconstructing fused images **(Kawulok et al., 2019) [18].** Similarly, IHS transformation is effective in improving spatial resolution but suffers from spectral distortion, limiting its application in high-precision remote sensing tasks **(Zhang et al., 2019) [19].**

Deep learning-based approaches overcome these challenges by automatically learning hierarchical features, enabling more robust fusion. CNNs have been widely used for super-resolution single images, where they enhance spatial details while preserving spectral consistency **(Cheng et al., 2020) [20].** Studies have demonstrated how well deep networks operate to extract contextual information and enhance classification accuracy in remote sensing applications **(Liu et al., 2020) [21].** Generative models, such as GANs and Image fusion has been further enhanced using transformer-based designs, which capture complex spatial patterns and producing high-resolution outputs with minimal artefacts **(Zhang et al., 2021) [22].**

Despite their advantages, deep learning models present challenges related to computational complexity and data dependency. Training deep networks requires large datasets and

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
significant processing power, which may limit their real-time applicability in remote sensing **(Kaur et al., 2021) [23].** Additionally, the inability to be interpreted in AI-driven fusion models remains a concern, as decision-making processes in neural networks are often considered "black-box" operations **(Jiang et al., 2021) [24].** Research efforts are currently focused on optimizing network architectures, incorporating attention mechanisms, and developing hybrid models that integrate traditional and deep learning-based techniques to enhance fusion performance **(Qiao et al., 2021) [25].**

Recent advancements in super-resolution techniques have further strengthened image fusion methodologies. Multi-frame super-resolution approaches have been explored to improve temporal fusion, allowing for enhanced detail extraction from low-resolution sequential images **(Li et al., 2021) [26].** Studies have highlighted the effectiveness of spatial-temporal fusion in uses like mapping cities and monitoring the environment **(Shan et al., 2022) [27].** The combination of deep learning and multi-frame image processing has resulted in significant improvements in data reconstruction and feature preservation, making it a promising direction for future research **(Zhou et al., 2022) [28].**

While deep learning continues to redefine the landscape of image fusion, ongoing research aims to tackle important issues including interpretability, computational economy, and flexibility in a variety of remote sensing datasets. Future advancements in neural architecture design, self-supervised learning, and quantum computing are expected to further refine deep learning-based image fusion techniques, making them more robust and scalable for real-world applications **(Wang et al., 2022) [29].** Additionally, new methodologies incorporating vast-receptive-field attention and adaptive reconstruction networks are being explored to enhance image resolution **(Lepcha et al., 2023) [30].** Recent studies have also demonstrated improvements in super-resolution of medical and geospatial images using GAN-based architectures **(Guerreiro et al., 2023) [31], (Gharibi & Faramarzi, 2023) [32].** Multi-frame spatio-temporal super-resolution techniques are proving beneficial for high-resolution image analysis in environmental applications **(Rikhsivoev et al., 2023) [33].**

3. METHODOLOGY

The research methodology has been described below in figure1. To approach image fusion and object detection using deep learning models, including other algorithms, the task begins with gathering and preprocessing of data. Using deep models begins with data gathering and pre-processing. Helicopter roundabout, diamond tennis court, basketball court, baseball, harbor bridge, ground track field, plane, ship storage tank, and large and small vehicles Data from distant photographs of a soccer field and a swimming pool are gathered and arranged for analysis and detection. Pre-processing procedures, including data rescaling, normalization, and data augmentation, are required before implementing deep learning models to make sure the input data is prepared for training. Because DL model can learn sequential patterns and understand temporal dependencies—for image fusion and object detection two crucial steps are important. In order to image fusion deep model Generative Adversarial Network (GAN). Two neural networks—a discriminator and a generator—that are trained simultaneously via adversarial training make up a GAN, a type of deep learning

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
model. The performance of the GAN model is assessed using the test data after it has been trained using the training data. Moreover, object detection is done using the innovative deep learning model referred to as the Region-based Convolutional Neural Network (R-CNN). Lastly, once the models have been trained, it is critical to assess their performance. Several measures, including as Accuracy, Structural Similarity Index (SSIM), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), are used to assess the models' accuracy and robustness. Below is a thorough explanation of every step.

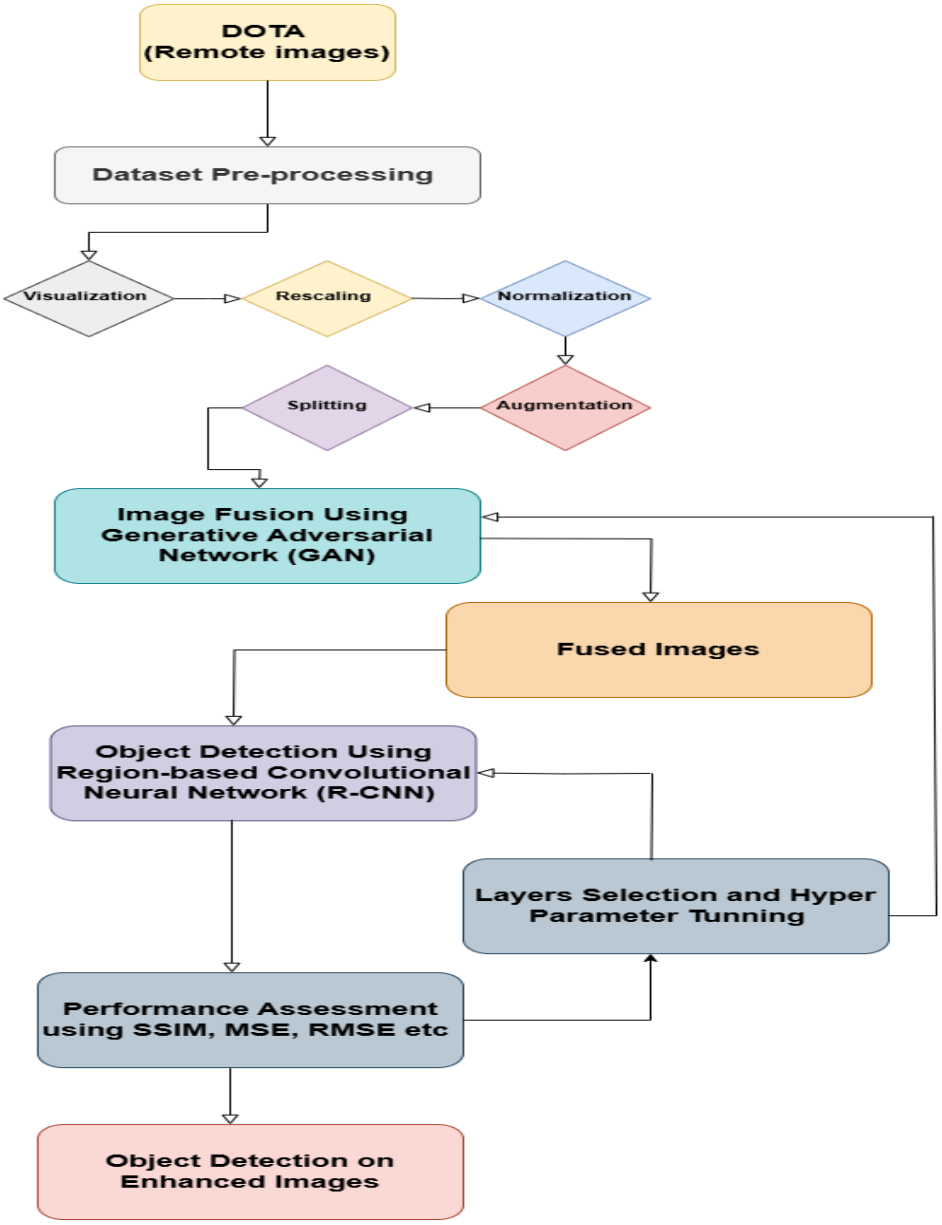


Figure 1: Proposed research methodology

3.1 Dataset description

A vast collection of remote sensing images created to facilitate object detection and analysis in aerial imagery is known as the DOTA. The full form of DOTA is Dataset for Object Detection in Aerial Images dataset. DOTA Image data was gathered from multiple sources,

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
such as the GF-2, JL-1, and Google Earth satellites, and was jointly donated by CycloMedia B.V. and the China Center for Satellite Data and Applications for Resources. The dataset offers a varied collection of visual data for research because it includes both RGB and greyscale images. The panchromatic band of GF-2 and JL-1 satellite photos, which normally offer high-resolution black-and-white imagery useful for in-depth spatial analysis, is where the greyscale images are taken from.

The RGB images, on the other hand, which were provided by CycloMedia and Google Earth, offer rich colour information that improves object classification based on spectral properties. The lossless 'png' file format is used to store all of the images, guaranteeing high-quality data preservation for reliable object detection and remote sensing applications.

Table 1: DOTA dataset description

Attribute	Discription
Link for the dataset	DOTA https://captain-whu.github.io/DOTA/dataset.html
Sources	**CycloMedia B.V.:** Provided information gathered using its high-definition imaging equipment. **Centre for Resources in China Application and Satellite Data:** GF-2, a Chinese high-resolution optical Earth observation satellite, supplied the satellite data. **JL-1:** A constellation of commercial remote sensing satellites from China. **Google Earth:** Provided richly detailed RGB visuals, frequently from aerial and satellite imagery.
Composition	**RGB Images:** Sourced from CycloMedia and Google Earth. Give colour information that is essential for jobs that call for spectral-based differentiation. **Greyscale images:** The panchromatic band of the GF-2 and JL-1 satellites is the source of greyscale images. They are useful for detailed object detection since they often have a higher resolution than multispectral bands.
Format	The DOTA dataset's images are all saved in the 'png' format, which guarantees lossless compression and constant processing quality.
Characteristics	**High Diversity** Uses data from multiple sources, including satellites and aircraft platforms. **Multiple-spectral Nature** Provides supplementary information by combining RGB and

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

	greyscale images. **Difficult Variability** The scale, direction, and resolution of objects of interest vary.

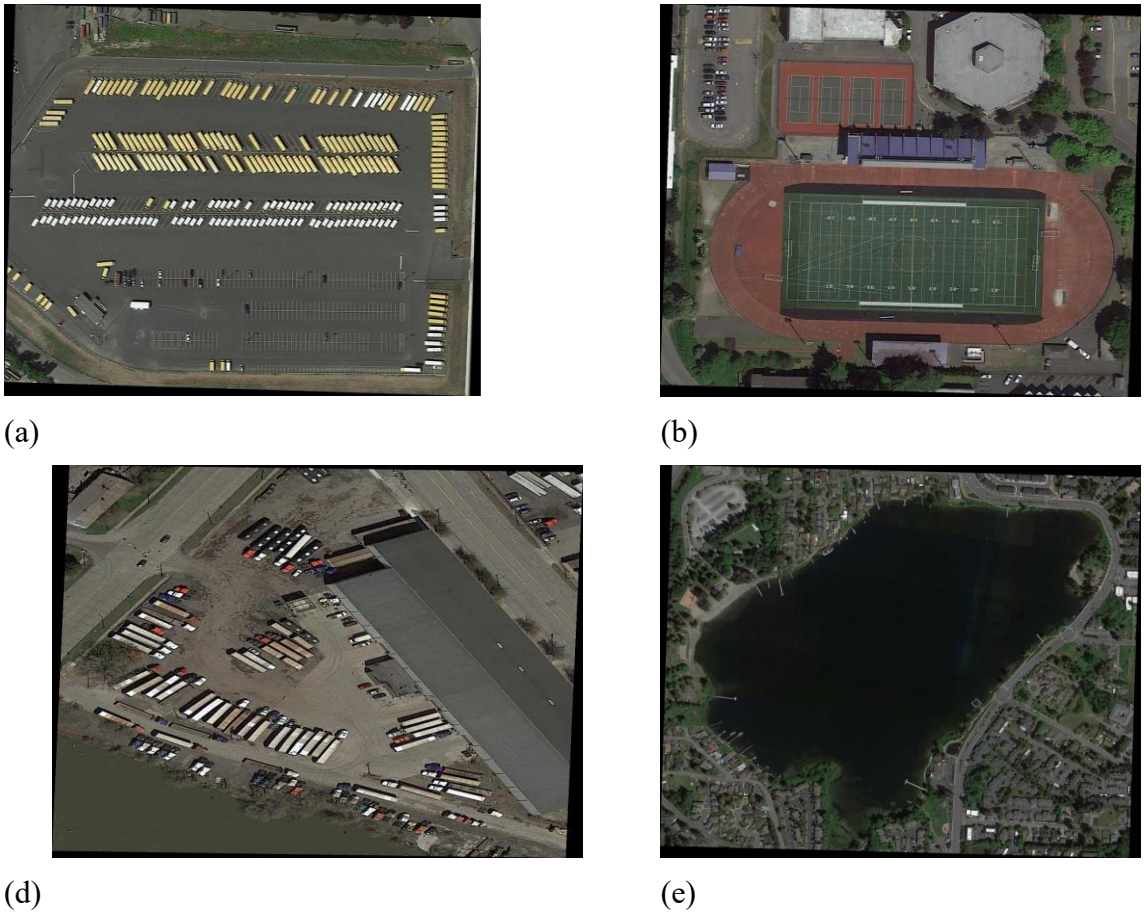
In DOTA dataset 1,793,658 object instances have in 18 distinct categories, all tagged with orientated bounding box annotations (OBB),. 11,268 aerial photos were used to gather these annotations.

Table 2: Class wise description in DOTA

Class Name	Number of Instances	Description
Plane	X instances	Airplanes in various orientations and sizes.
Ship	X instances	Different types of ships and boats.
Storage Tank	X instances	Circular or rectangular storage tanks.
Baseball Diamond	X instances	Baseball fields with distinct patterns.
Tennis Court	X instances	Tennis courts with visible markings.
Basketball Court	X instances	Basketball courts with hoops and boundaries.
Ground Track Field	X instances	Ground tracks, including running tracks.
Harbor	X instances	Harbors with boats, ships, and docks.
Bridge	X instances	Bridges of various lengths and shapes.
Large Vehicle	X instances	Trucks, buses, and similar large vehicles.
Small Vehicle	X instances	Cars and smaller vehicles.
Helicopter	X instances	Helicopters in different positions.
Roundabout	X instances	Circular intersections and traffic circles.
Soccer Ball Field	X instances	Soccer fields with goals and markings.
Swimming Pool	X instances	Residential or public swimming pools.
Container Crane	X instances	Cranes used in shipping container handling.

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)



(a) (b)

(d) (e)

Figure 1: Sample images

3.1.2 Dataset rescaling

Dataset rescaling is essential for preparing data for machine learning algorithms in image fusion and object detection projects, particularly when working with remote sensing image. The original size of the images wes varying i.e. Original size of P0008.jpg: (572, 618) Original size of P0011.jpg: (684, 763) Original size of P0005.jpg: (442, 489) etc. and it has been rescaled to 100*100. By standardizing the input, rescaling image of various sizes to a set size (such 100x100 pixels) helps maintain consistency and enhances the model's performance.

Further processing larger images in our work take a lot of memory and processing power, hence for training and inference times by downsizing all images to 100x100 pixels, which lowers the number of pixels the model must process. In real-time object detection jobs, when processing speed is crucial, this is extremely crucial.

3.1.3 Normalization

In deep learning, normalization is an essential step, especially when using Convolutional Neural Networks (CNNs) for image handling tasks like object detection and picture fusion.

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
The two formats that are typically used for remote sensing images are RGB (Red, Green, Blue) and GS (Greyscale). In these images, every pixel has an intensity value that corresponds to its colour or brightness. The values of the Red, Green, and Blue colour channels in RGB images normally range from 0 to 255. This also holds true for GS pictures, where the intensity of the pixels ranges from 0 (black) to 255 (white). However, input data in a smaller numerical range, usually [0, 1] or [-1, 1], tends to perform better in neural networks.

3.1.4 Data augmentation

Image augmentation techniques can greatly improve a Convolutional Neural Network's (CNN) performance and generalization capabilities in task like image fusion and object detection tasks, especially when working with remote sensing images. In remote sensing, where objects can appear at different angles, the model learns to detect objects from multiple orientations with the aid of the provided strategies, such as rotation range of 15, which permits random rotations between -15° and 15°.

A number of techniques, such as rotation, shifting, shearing, and zooming, were used to augment the data that already existed. By giving numerical stability and increasing the model's efficiency, 1./255 normalizes pixel values, scaling them between 0 and 1, guaranteeing quicker convergence during training. In order to help the model become invariant to small distortions, the shear range: 0.2 adds a shear transformation that slightly distorts images to represent various perspectives. In order to assist the model generalize to objects of varied scales-a crucial feature when recognizing items at different distances in satellite photos Zoom_range: 0.2 randomly zooms in or out of images. Horizontal_flip makes the model invariant to left-right orientation by flipping images horizontally, adding randomization. This is particularly helpful for object detection jobs where objects may appear in reverse orientations. Fill_mode: Nearest preserves image integrity by filling in any gaps left by operations like rotation or zooming with the closest pixel values. The image is shifted by up to 10% horizontally and vertically by width_shift_range: 0.1 and height_shift_range: 0.1.

Table 3: Details of dataset augmentation technique

	Technique		**Value**
	Rotation_range	---	15
	Rescale	---	1./255
	Shear_range	---	0.2
	Zoom_range	---	0.2
	Horizontal_flip	---	True
	Fill_mode	---	Nearest
	Width_shift_range	---	0.1

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

	Height_shift_range	---	0.1

3.1.5 Dataset splitting

An important factor in the training process is the ratio at which the dataset is divided. A 4:1 ratio has been employed in the current study, which means that 20% of the dataset is put aside for testing and 80% for training. This is a standard procedure since it gives the model a sizable enough dataset to learn from while yet ensuring that the testing data is fairly evaluated. While the testing set acts as a standard to evaluate how effectively the model generalizes to novel, unseen examples, a larger percentage of training information aids in the model's learning. The purpose of testing set is to evaluate the model's efficacy and accuracy, while the training set aids in the development of its internal parameters (weights and biases).

3.1.6 Image fusion using GAN

Particularly for remote sensing applications, Generative Adversarial Networks (GANs) have demonstrated significant promise in improving image fusion methods. Images from various sensors, including infrared, multispectral, and panchromatic pictures, are frequently recorded in remote sensing with diverse spectral information and resolutions. Conventional picture fusion techniques can fall short in effectively preserving spectral and spatial features. A reliable answer to this problem is provided by GANs, which are composed of a discriminator and a generator. A fused image is produced by the generator by fusing the rich spectrum information from multispectral data with the high-resolution spatial details from panchromatic images. The architectural layout of the self-developed GAN model is displayed in Figure 4.

Here in the model, output layers (OL), convolutional layers (CL) and max-pooling layers (MP) are present. Several layers make up the Generator Architecture of a Generative Adversarial Network (GAN) used for image fusion. Each layer has unique parameters that aid in learning and produce high-quality fused images. A Dense layer is the first step in the process, which turns a latent input vector into a high-dimensional vector (of size 16384). Batch Normalization, which normalizes the activations to stabilize training, and ReLU activation, which adds non-linearity to capture intricate patterns, come next. After that, the output is transformed into a 3D tensor (4x4x1024) that is used in convolutional layers for image processing.

With a huge number of parameters (13,107,200) that correspond to the up sampling process, the initial Conv2DTranspose layer reduces the feature channels from 1024 to 512 while up sampling the image to an 8x8 resolution. Learning capacity is improved and stable activations are maintained by further Batch Normalization and ReLU activation layers. A second normalization and activation layer comes after the second Conv2DTranspose layer further up samples the resolution to 16x16 and decreases the feature channels from 512 to 256. Conv2DTranspose reduces the quantity of featured channels from 256 to 128 by up sampling the image to 32x32.

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
The final Conv2DTranspose layer creates the RGB image that is intended for image fusion by up sampling the image to 64x64 and reducing the feature channels to 3. The last activation layer scales the pixel values to the appropriate range, frequently with the use of a sigmoid or tanh. The ReLU activations enable the generator to learn intricate, non-linear correlations, while the Batch Normalization layers aid in preventing over fitting and accelerating convergence across the architecture. Over 18 million parameters make up the complete generator model, most of which are trainable. The generator may gradually get better and create realistic, high-resolution fused images thanks to its structure.

On the other hand discriminator architecture in a Generative Adversarial Network (GAN) for Image Fusion is to distinguish between generated and actual images. It starts with a Conv2D layer that creates a 32x32 feature map by the use of 64 filters to the picture that was entered. The kernel size and number of filters used during the convolution determine how many parameters (1,728) are in this layer. Unlike regular ReLU, the LeakyReLU the activation function introduces non-linearity, allowing the model to recognize increasingly complex patterns without resulting in dead neurones.

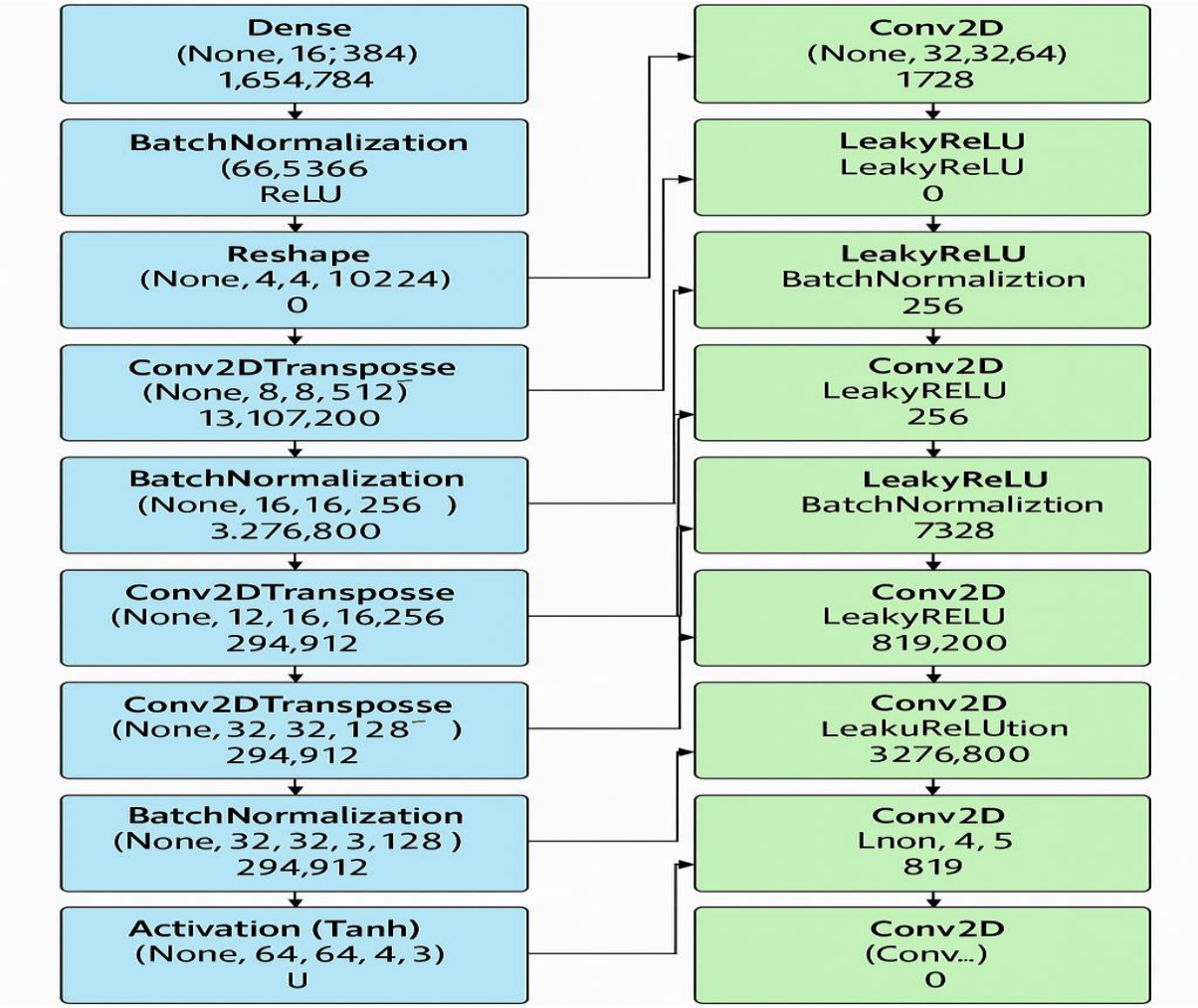


Figure 3: Proposed GAN model's architecture

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Conv2D_1, the following layer, increases the feature map depth to 128 while further reducing the image size to 16x16. By ensuring that the activations in the subsequent layer have a stable distribution, the Batch Normalization layer lowers the internal covariate shift and expedites training. With minor negative slope activations, another LeakyReLU maintains the learning potential. The model goes on to include a Conv2D_2 layer, which raises the feature map depth to 256 and decreases the resolution to 8x8. These activations are once more normalized via batch normalization. LeakyReLU_2, the following layer, keeps adding non-linearity.

The feature maps' depth is increased to 512 by the Conv2D_3 layer, which further lowers the resolution to 4x4. After reapplying batch normalization to stabilize the activations, a LeakyReLU_3 is used to preserve the complex feature flow. Lastly, the Conv2D_4 layer represents the discriminator's binary output (genuine or fake) by reducing the feature map to a 1x1 size with a single filter. The final classification output, which is either 0 (fake) or 1 (real), is obtained by applying a sigmoid activation after the Flatten layer flattens this feature map into a one-dimensional vector.

3.1.7 Object detection using RCNN

The object detection method known as Region-based Convolutional Neural Network (RCNN) employs a two-step procedure. Initially, it uses selective search to produce region ideas from an input image. A convolutional neural network (CNN) is then utilized to categorize and improve the bounding boxes based on these region proposals. Predicted object labels and the matching bounding box locations for identified objects make up the result. For efficient object detection, this method combines deep learning's benefits with region proposal networks

For the present work RCNN model that has been used is pre-trained model. The model has composed of the following layers.

Convolutional layers: Feature maps are extracted from input photos by convolutional layers.

Region Proposal Network (RPN): It suggests potential areas or object-containing bounding boxes.

ROI polling: Proposals are transformed into fixed-size feature maps via ROI pooling.

Fully Connected Layers: For precise object detection, bounding boxes are refined and regions are classified.

3.1.8 Model’s training parameters

The various learning parameters of the model has been given in table 4

Table 4: Training parameters of Model

Parameters	**Value**
Input Shape	Variable
Dataset Division	Random

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Model	Sequential
Convolutional Layers	Variable
Activation Function for Hidden Layers	Sigmoid
Activation Function for Convolutional Layers Layer	Relu
Loss Function	Sparse categorical cross entropy
Optimizer	Adam
Metrics	Mean_squared_error
Validation Check	6
Epochs	10-100

The modest learning rate was chosen since significant changes to the deep filters utilizing convolution were not desired. It was discovered that the learning process collapsed when training was conducted with losses that erupted to extremely high levels and could not recover. The maximum value that might facilitate the model's positive learning was used to determine the batch size. This is due to the fact that larger batch sizes are known to provide a more accurate gradient calculation for the amount of weight update, as was previously indicated.

The total number of epochs during model training was determined by using endpoints for every epoch. When over fitting was detected, the endpoint of the prior epoch was used. First, a 75% training set and a 25% validation set were created from the dataset.

As stated in the training/test splits, one hundred and five present was used for model training and twenty-five percent for testing. An enormous amount of parameters were used to instruct the model. This illustrates how the amounts of parameters and the model's depth have changed the computation time. The model's competency from the perspective of its architectural design is demonstrated by the extremely low memory and processing costs needed to train it with this sensible number of parameters—roughly 12MB.

3.1.9 System configuration

Training was conducted using the method of back-propagation on a small batch size of 32 and epochs of 10-100. To increase the precision, several parameters, including decay, dropout function, and learning rate, were set to 1e-6, 0.5, and 1e-3, respectively. A PC outfitted with an Intel®Core (TM) i5-6200U, 8 GB of RAM, a CPU operating at 2.30 GHz, and Python 3.6 software with Keras 2.2.4 and a Tensorflow 1.12.0 backend library was used for the various experiments.

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Additionally, the tests were carried out on a T4 GPU running Google Colab for up to 12 hours straight. Depending on the quantity of GPU traffic at the time, the training time changed with each epoch.

Table 5: Virtual hardware details

	Hardware		**Details**
	Platform	---	Google Colab
	GPU	---	T4
	Programming version	---	Python 3
	RAM	---	14 GB
	ROM	---	64GB

4. Results

When analysing implemented algorithms, the image data for images is taken into account. A range of images indicating that various criteria have been used, and algorithms are put into practice and assessed. The results if the same is presented below.

4.1 GAN's results' analysis

Using remote images data, GAN learning-based multi item detection is used and results has been shown in this paper. For image fusion the results' analysis of GAN model is presented below.

5.2.1 Epoch wise analysis

For each epoch while training on DOTA dataset images, the GAN's performance is seen below. in Table 7

Table 4: GAN analysis for epoch

Epoch 1/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **21s** 280ms/step - disc_loss: 0.5698 - gen_loss: 7.8367
Epoch 2/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 221ms/step - disc_loss: 0.4392 - gen_loss: 5.9216
Epoch 3/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 229ms/step - disc_loss: 0.3782 - gen_loss: 6.0829
Epoch 4/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 171ms/step - disc_loss: 0.3578

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- gen_loss: 6.1182
Epoch 5/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 219ms/step - disc_loss: 0.3598 - gen_loss: 6.5320
Epoch 6/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 176ms/step - disc_loss: 0.3496 - gen_loss: 6.6073
Epoch 7/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 174ms/step - disc_loss: 0.3437 - gen_loss: 6.5039
Epoch 8/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 179ms/step - disc_loss: 0.3493 - gen_loss: 6.6457
Epoch 9/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 175ms/step - disc_loss: 0.3537 - gen_loss: 6.9110
Epoch 10/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 203ms/step - disc_loss: 0.3425 - gen_loss: 6.9123
Epoch 11/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 315ms/step - disc_loss: 0.3459 - gen_loss: 7.4318
Epoch 12/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 221ms/step - disc_loss: 0.3397 - gen_loss: 7.2793
Epoch 13/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 171ms/step - disc_loss: 0.3443 - gen_loss: 7.0175
Epoch 14/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 178ms/step - disc_loss: 0.3358 - gen_loss: 7.4002
Epoch 15/80

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 175ms/step - disc_loss: 0.3322 - gen_loss: 7.3112
Epoch 16/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 184ms/step - disc_loss: 0.3343 - gen_loss: 7.2842
Epoch 17/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 173ms/step - disc_loss: 0.3326 - gen_loss: 7.4142
Epoch 18/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 184ms/step - disc_loss: 0.3332 - gen_loss: 7.6676
Epoch 19/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 322ms/step - disc_loss: 0.3329 - gen_loss: 7.6582
Epoch 20/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 220ms/step - disc_loss: 0.3300 - gen_loss: 8.0341
Epoch 21/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 172ms/step - disc_loss: 0.3328 - gen_loss: 7.5399
Epoch 22/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 178ms/step - disc_loss: 0.3414 - gen_loss: 7.9710
Epoch 23/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 171ms/step - disc_loss: 0.3335 - gen_loss: 7.8657
Epoch 24/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 177ms/step - disc_loss: 0.3358 - gen_loss: 7.7594
Epoch 25/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 178ms/step - disc_loss: 0.3333 - gen_loss: 7.8904

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Epoch 26/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 198ms/step - disc_loss: 0.3322 - gen_loss: 7.8236
Epoch 27/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 243ms/step - disc_loss: 0.3322 - gen_loss: 7.9611
Epoch 28/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 172ms/step - disc_loss: 0.3362 - gen_loss: 7.8540
Epoch 29/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 273ms/step - disc_loss: 0.3311 - gen_loss: 8.1988
Epoch 30/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 181ms/step - disc_loss: 0.3300 - gen_loss: 7.9911
Epoch 31/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 176ms/step - disc_loss: 0.3298 - gen_loss: 7.8841
Epoch 32/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 172ms/step - disc_loss: 0.3306 - gen_loss: 8.1533
Epoch 33/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 236ms/step - disc_loss: 0.3301 - gen_loss: 8.0752
Epoch 34/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 236ms/step - disc_loss: 0.3327 - gen_loss: 8.2038
Epoch 35/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 186ms/step - disc_loss: 0.3320 - gen_loss: 8.3917
Epoch 36/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 176ms/step - disc_loss: 0.3302

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- gen_loss: 8.2708
Epoch 37/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 176ms/step - disc_loss: 0.3296 - gen_loss: 8.2240
Epoch 38/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 180ms/step - disc_loss: 0.3290 - gen_loss: 8.3127
Epoch 39/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 204ms/step - disc_loss: 0.3304 - gen_loss: 8.3660
Epoch 40/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 223ms/step - disc_loss: 0.3281 - gen_loss: 8.3691
Epoch 41/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 310ms/step - disc_loss: 0.3314 - gen_loss: 8.3203
Epoch 42/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 184ms/step - disc_loss: 0.3328 - gen_loss: 8.3906
Epoch 43/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 177ms/step - disc_loss: 0.3316 - gen_loss: 8.4284
Epoch 44/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 175ms/step - disc_loss: 0.3280 - gen_loss: 8.3138
Epoch 45/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 181ms/step - disc_loss: 0.3274 - gen_loss: 8.5134
Epoch 46/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 241ms/step - disc_loss: 0.3285 - gen_loss: 8.1792
Epoch 47/80

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 233ms/step - disc_loss: 0.3283 - gen_loss: 8.2897
Epoch 48/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 179ms/step - disc_loss: 0.3287 - gen_loss: 8.3743
Epoch 49/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 185ms/step - disc_loss: 0.3272 - gen_loss: 8.3563
Epoch 50/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 181ms/step - disc_loss: 0.3279 - gen_loss: 8.4620
Epoch 51/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 178ms/step - disc_loss: 0.3299 - gen_loss: 8.4360
Epoch 52/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 177ms/step - disc_loss: 0.3327 - gen_loss: 8.4390
Epoch 53/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 234ms/step - disc_loss: 0.3394 - gen_loss: 8.2684
Epoch 54/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 234ms/step - disc_loss: 0.3375 - gen_loss: 8.3501
Epoch 55/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 179ms/step - disc_loss: 0.3311 - gen_loss: 8.5664
Epoch 56/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 337ms/step - disc_loss: 0.3296 - gen_loss: 8.5567
Epoch 57/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 185ms/step - disc_loss: 0.3288 - gen_loss: 8.6372

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Epoch 58/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 183ms/step - disc_loss: 0.3290 - gen_loss: 8.4993
Epoch 59/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 179ms/step - disc_loss: 0.3273 - gen_loss: 8.6959
Epoch 60/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 216ms/step - disc_loss: 0.3281 - gen_loss: 8.6131
Epoch 61/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 231ms/step - disc_loss: 0.3283 - gen_loss: 8.5181
Epoch 62/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 182ms/step - disc_loss: 0.3288 - gen_loss: 8.7258
Epoch 63/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 182ms/step - disc_loss: 0.3301 - gen_loss: 8.7140
Epoch 64/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 182ms/step - disc_loss: 0.3278 - gen_loss: 8.4519
Epoch 65/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 178ms/step - disc_loss: 0.3268 - gen_loss: 8.6329
Epoch 66/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 196ms/step - disc_loss: 0.3274 - gen_loss: 8.5957
Epoch 67/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 232ms/step - disc_loss: 0.3267 - gen_loss: 8.8518
Epoch 68/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 204ms/step - disc_loss: 0.3292

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- gen_loss: 8.6197
Epoch 69/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 178ms/step - disc_loss: 0.3295 - gen_loss: 8.7582
Epoch 70/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 181ms/step - disc_loss: 0.3289 - gen_loss: 8.7226
Epoch 71/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 177ms/step - disc_loss: 0.3273 - gen_loss: 8.6805
Epoch 72/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 178ms/step - disc_loss: 0.3303 - gen_loss: 8.6670
Epoch 73/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **1s** 178ms/step - disc_loss: 0.3290 - gen_loss: 8.6443
Epoch 74/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 232ms/step - disc_loss: 0.3318 - gen_loss: 8.8800
Epoch 75/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **4s** 460ms/step - disc_loss: 0.3297 - gen_loss: 8.5932
Epoch 76/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 184ms/step - disc_loss: 0.3291 - gen_loss: 8.9299
Epoch 77/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 185ms/step - disc_loss: 0.3298 - gen_loss: 8.7037
Epoch 78/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **3s** 186ms/step - disc_loss: 0.3280 - gen_loss: 8.8530
Epoch 79/80

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 180ms/step - disc_loss: 0.3287 - gen_loss: 8.6693
Epoch 80/80
9/9 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **2s** 233ms/step - disc_loss: 0.3305 - gen_loss: 8.5558

A pair of neural networks, the generator and discriminator are both trained concurrently in a competitive environment to form Generative Adversarial Networks (GANs). While the discriminator decides if the incoming data is produced or real, the generator seeks to generate data that closely matches actual data. For efficient training, it is essential to comprehend the loss functions of these two networks due to their intricate dynamics. Here loss function has been calculated for both. As per table 5.1 the performance of model is promising.

4.2 Loss curve analysis

The GAN model's loss plot indicates that the discriminator and generator were balanced during training. It can differentiate between authentic and fraudulent data, as demonstrated by the discriminator loss. In the meantime, the generator loss is relatively low and steady.

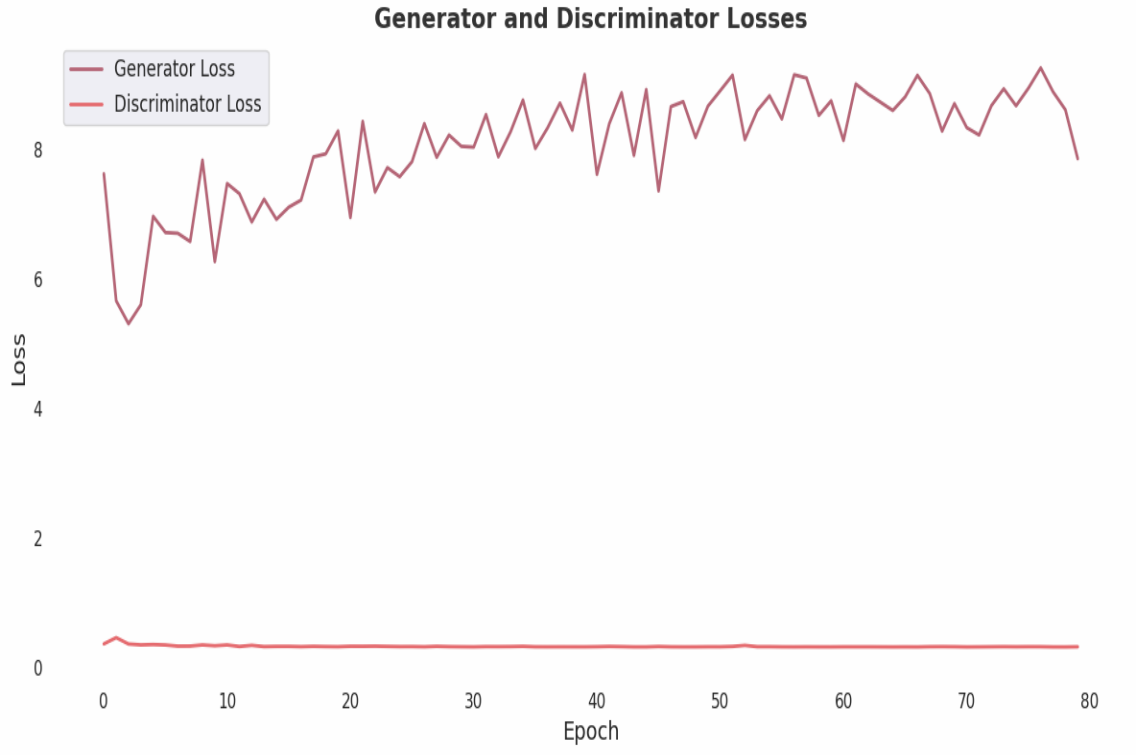


Figure 4: Loss curve for GAN

4.3 RCNN results analysis

For object detection, RCNN models perform better. As reliable multiple object detection is necessary for a variety of remote image applications, hence deep learning-based detection association model is used in this chapter to detect.

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
The detection model for multiple objects in aerial image that is based on deep learning is work on the following methodology.

1. Object labeling
2. Detector training using labeled frames
3. RCNN model training using visual characteristics and bounding box locations

Three different loss graphs that each emphasize a different facet of the model's learning process have been created for the model's performance assessment according to the described process.

Box loss figure 5, also known as frame loss, is represented by the first graph. It quantifies the discrepancy between the ground truth boxes and expected bounding boxes. This loss penalizes any misalignment or size problems and evaluates how well the model locates objects inside the image. A steady decrease in this loss suggests that the model is improving its ability to recognize and precisely frame things.

The object loss figure 6, which measures the model's accuracy in forecasting the existence of objects inside the identified bounding boxes, is depicted in the second graph. Reduced values indicate a better capacity to distinguish between regions with and without items.

The class loss figure 7 is shown by the third graph, which focuses on object categorization. It assesses how well the model predicts the categories of the objects that are observed. A steady reduction in this loss indicates that the model's accuracy in classifying things has increased.

These graphs show how the model has improved over time in terms of object localization, detection, and classification accuracy, offering a thorough overview of the training process.

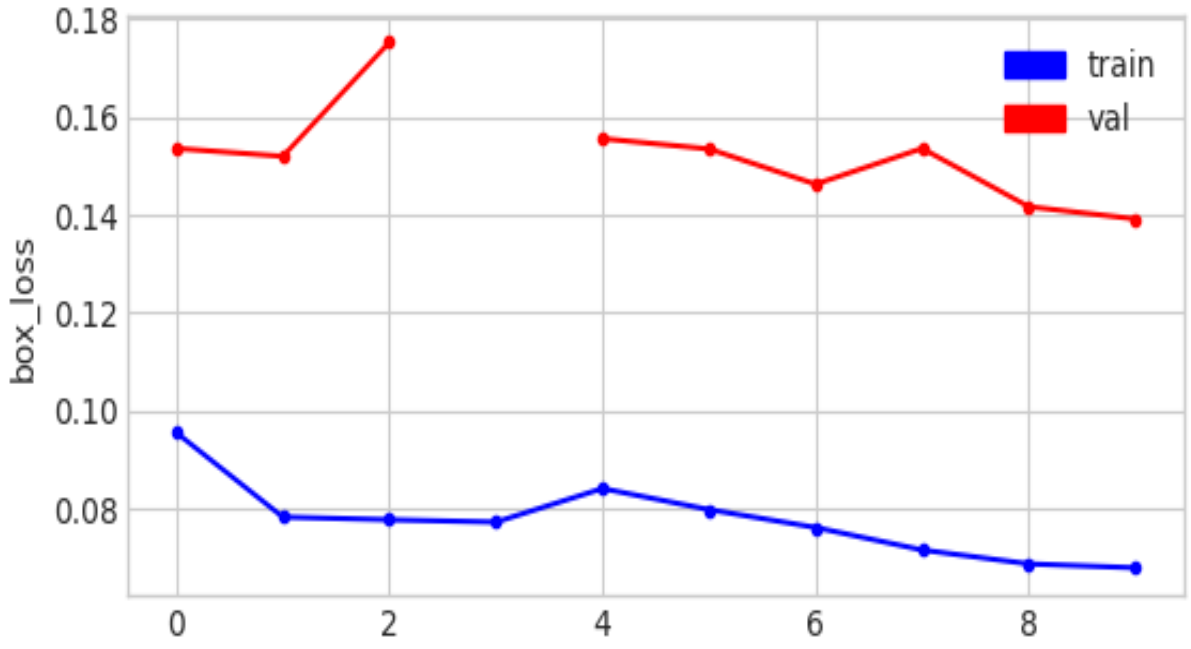


Figure 5: Training/ validation loss graph

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
The box loss during the training phase is seen by the blue line (train loss). The blue line's downward trend indicates that the model is improving at locating the bounding boxes around things as it is being trained. The loss indicates efficient optimisation during training because it begins relatively high at the beginning (epoch 0) and steadily decreases as the epochs go by. The model has learnt effectively from the training data by the latter epochs, as evidenced by the training loss stabilising at a low value.

The box loss on the validation dataset is shown by the red line (validation loss). Because of the disparity in data exposure, the validation loss is initially greater than the training loss, which is typical. The validation loss shows a declining tendency over time, with some minor changes, including occasional spikes. Though not as regularly as the training loss, the validation loss eventually decreases as well.

Further as per figure 8,the model is successfully gaining knowledge from the training data, as evidenced by the consistent decline in training loss. However, there may be some difficulties with generalisation or overfitting tendencies in particular epochs if the validation loss varies somewhat throughout epochs. The gap between training and loss of validation, especially in the early stages, illustrates the model's incapacity to make generalizations from training to unseen validation data. However, as training goes on, the gap gets a little smaller.

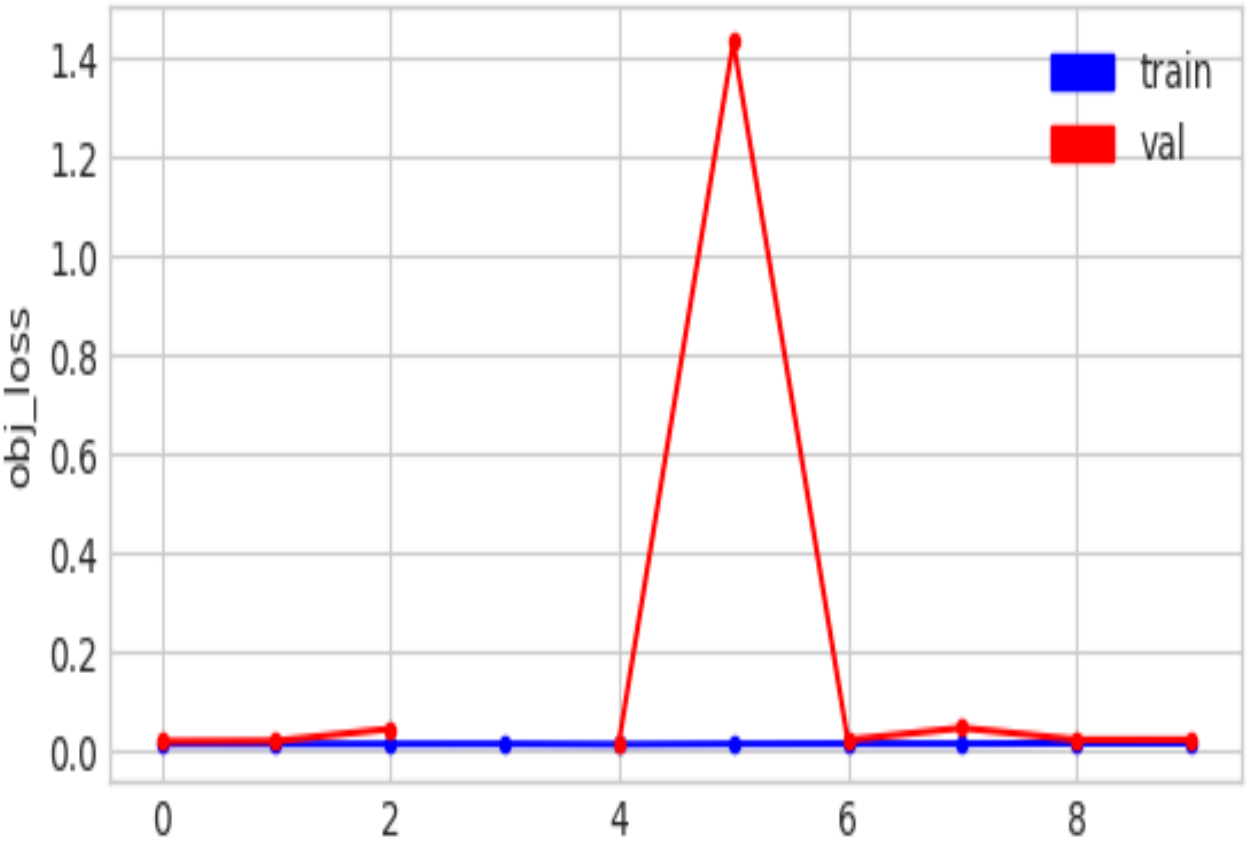


Figure 6: Training/ validation object loss graph

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
As per figure 9, throughout all epochs, the training loss stays near zero, demonstrating how well the model detects the existence of objects in the training data. The training loss's flat trend indicates that the model has probably converged early for this feature and has rapidly learnt to discriminate objects during training.

On the other hand, at epoch 4, the validation loss exhibits a notable uptick, peaking at approximately 1.4. This suggests that during this particular epoch, the model had trouble predicting the presence of objects on the validation dataset. The validation loss rapidly returns to almost zero after the surge, matching the training loss by subsequent epochs.

The graph shows that, save from a brief instability during validation at epoch 4, the model is generally successful in predicting the presence of objects. Although the spike might call for more research into the validation data or training procedure for that epoch, the minimal object loss overall suggests that the model is well-trained for object confidence prediction.

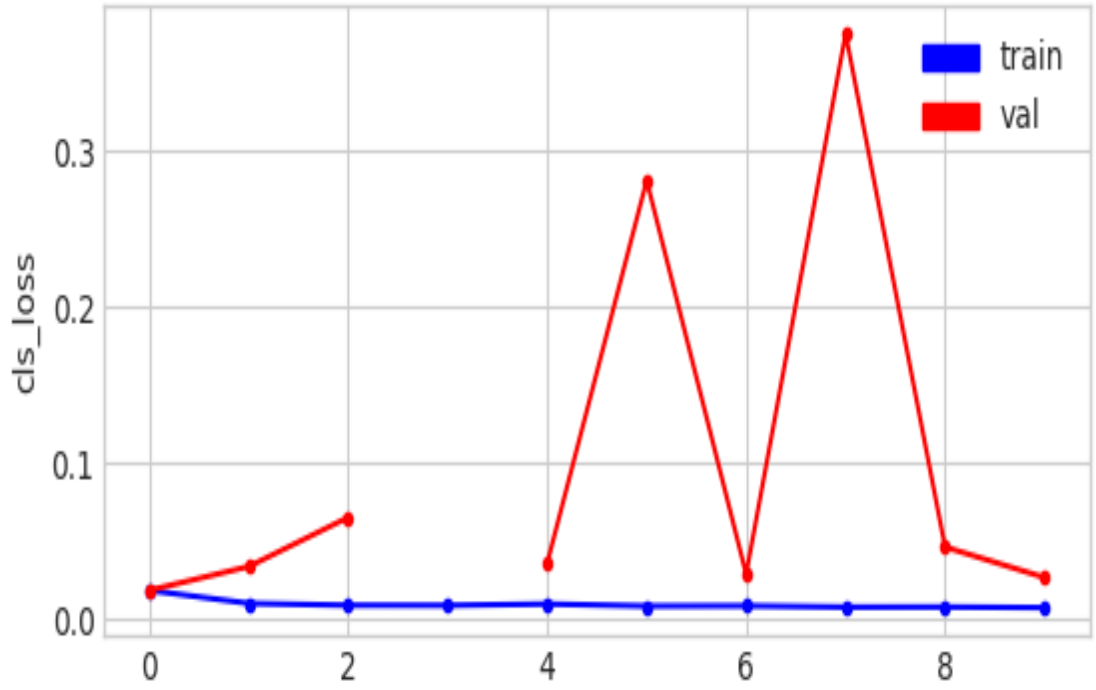


Figure 7: Training/ validation object's class loss graph

The model has successfully learnt to categorise objects in the training data when the training loss is continuously close to zero during all epochs. Minimal classification mistakes during training are suggested by the stability and trend that is close to zero. There are noticeable jumps in the validation loss at epochs 4 and 7, reaching about 0.3. These spikes imply that at those particular epochs, the model had trouble accurately classifying items in the validation dataset. After every spike, the loss decreases, suggesting that the model has recovered or adjusted to some extent.According to the graph, the model does a good job of categorising items in the training data, but it occasionally has trouble with the validation dataset.

More research should be done on the validation loss spikes, perhaps by looking for abnormalities in the validation data or improving the model to increase its capacity for

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
generalisation. The approach appears promising overall, however it might require minor tweaks to more reliably handle classification problems.

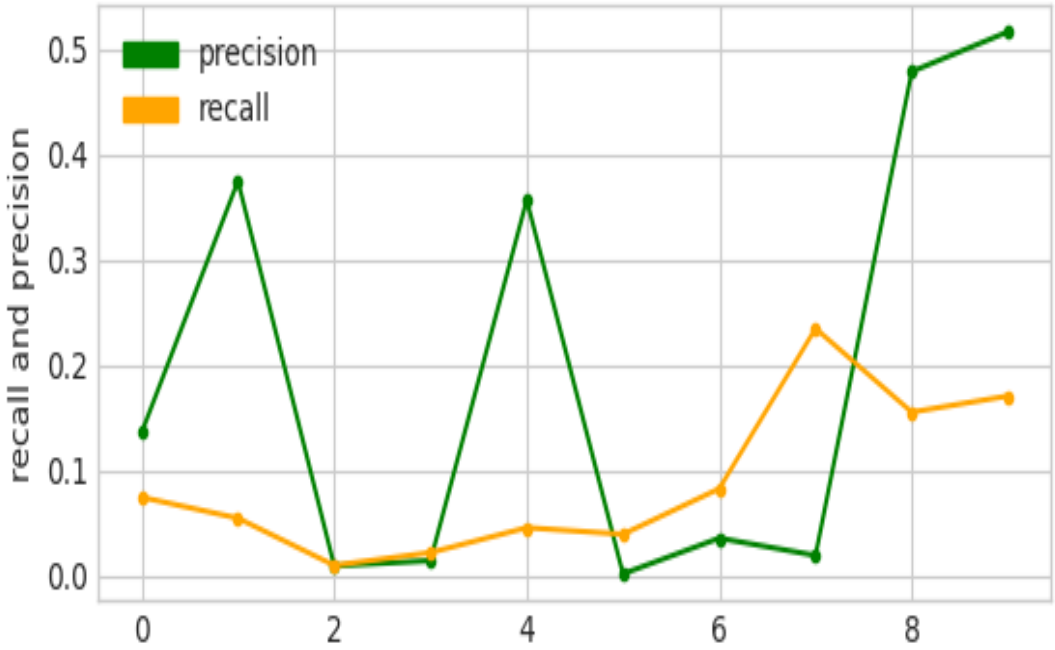


Figure 8: Precision/recall analysis

The reeults of precision and recall has been shown in figure 9 The precision and recall findings, two crucial measures for assessing model performance, particularly in classification tasks, are shown in Figure 11. While accuracy is the percentage of properly recognized favorable forecasts over all of the model's positive predictions, recall measures the proportion of real positives that were successfully detected.

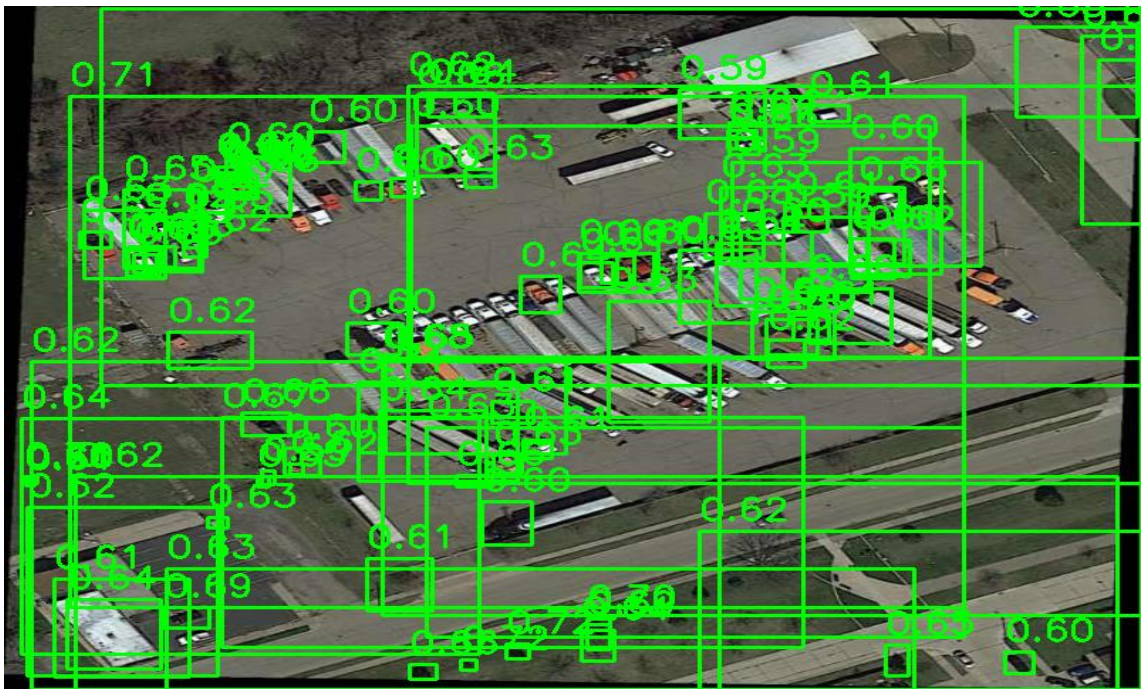
Both recall and accuracy are probably plotted over various thresholds or epochs in the graph. When the precision of the model is great, it can accurately predict a positive class and reduces false positives. Conversely, a high recall indicates that most true positive samples are correctly captured by the model, reducing false negatives.

A steady rise in recall and precision over time points to better model learning and optimisation. The model may be overly cautious, predicting fewer positives to prevent false positives, if precision is high but recall is low. On the other hand, a high recall but low precision means that while the model predicts a lot of positives, some of them are wrong.

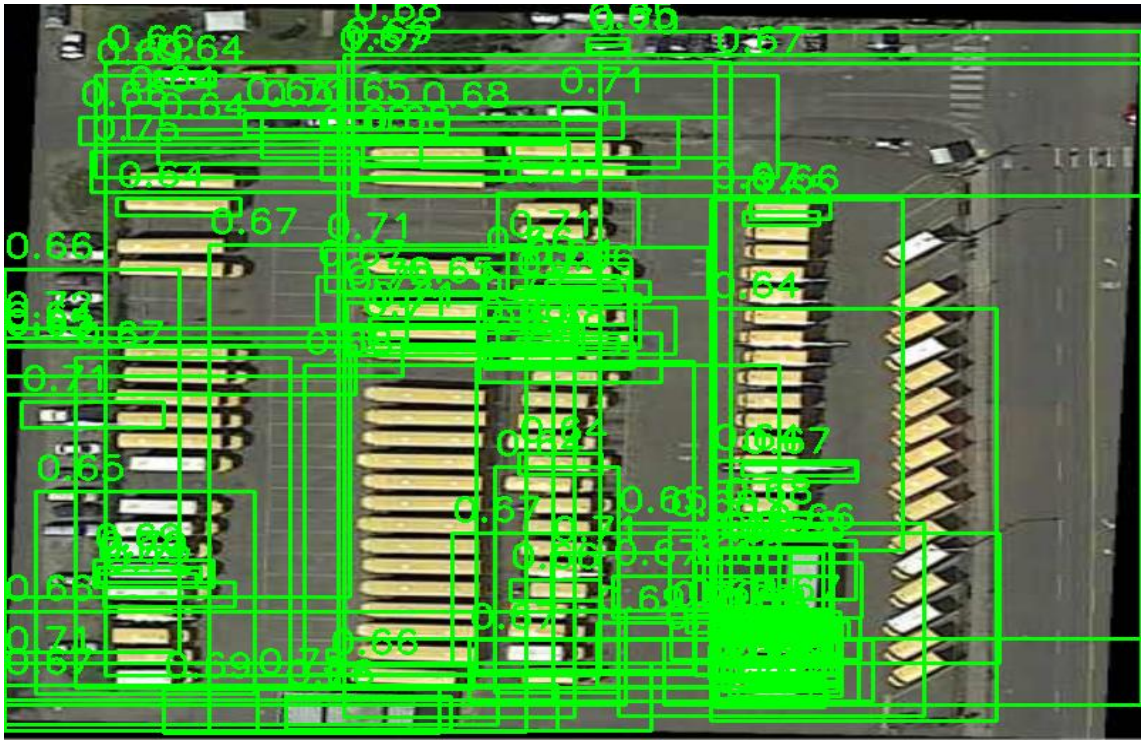
As per figure 9, the model attains a precision of over 0.5 and a recall of roughly 0.18, indicating an unbalanced trade-off between these two crucial metrics. More over half of the model's optimistic forecasts are true. when the precision parameter is higher than 0.5. This implies that the model is only mediocrely successful in preventing false positives and guaranteeing that the majority of its positive predictions are accurate.

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
The discrepancy between recall and precision indicates that the model prioritises precision over recall. An unbalanced dataset, improper threshold settings, or unduly cautious projections could be the cause of this.



(a)



(b)

Figure 9: Sample figure of object detetcion

Received: August 10 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5. Conclusion

A potent method to improve the spatial and spectral quality of remote sensing images is deep learning-based image fusion. These fusion techniques greatly enhance feature extraction, noise reduction, and structure preservation by combining several images sources, resulting in more precise and trustworthy image analysis. By contrasting their performance with non-fused images and examining their wider ramifications across several domains, this work investigates the efficacy of combining techniques based on deep learning for remote sensing applications.

For this work, an extensive collection of remote sensing images DOTA dataset has been used. This dataset is created to improve object detection and analysis in aerial data. The China Centre for Resources Satellite Data and Application and CycloMedia B.V. have contributed a large range of RGB and greyscale photographs to the dataset.

To accomplish image fusion, the study uses deep learning models, such as Transformers, Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs). To enhance the fusion results, these models make use of self-attention mechanisms and hierarchical feature learning. To determine whether fused images are better than non-fused ones, comparative analyses are carried out using both qualitative and quantitative measures, such as information retention, structural similarity, and visual clarity. The self-devised Generative Autoencoder (GAE) model is designed for unsupervised feature learning and image reconstruction, leveraging structured decoder architecture to transform latent representations into high-fidelity images. Before performing batch normalisation and ReLU activation to stabilise learning and add non-linearity, the model starts with a fully linked dense layer that enlarges the latent space into a high-dimensional vector of 16,384 units. Similarly, A convolutional neural network (CNN)-based architecture, the self-developed discriminator model is intended for binary classification in adversarial learning and is commonly employed in Generative Adversarial Networks (GANs).

The results of GAN depcited that over several epochs, the Generative Adversarial Network (GAN) showed increasing success in producing realistic samples during training. After the operation of GAN, RCNN model has been applied for object detetcion.

The findings show an imbalance between both parameters, with the model achieving a precision of more than 0.5 and a recall of roughly 0.18. A high precision number indicates that the model successfully reduces false positives, guaranteeing the accuracy of the majority of its positive predictions. A higher proportion of false negatives results from the model's inability to identify a sizable percentage of true positive cases, as indicated by the lower recall value.

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

REFERENCES

1. Deshmukh, M., & Bhosale, U. (2010). Image fusion and image quality assessment of fused images. *International Journal of Image Processing (IJIP)*, *4*(5), 484.
2. Serpico, S. B., Dellepiane, S., Boni, G., Moser, G., Angiati, E., & Rudari, R. (2012). Information extraction from remote sensing images for flood monitoring and damage evaluation. *Proceedings of the IEEE*, *100*(10), 2946-2970.
3. Fookes, C., Lin, F., Chandran, V., & Sridharan, S. (2012). Evaluation of image resolution and super-resolution on face recognition performance. *Journal of Visual Communication and Image Representation*, *23*(1), 75-93.
4. Kanaev, A. V., & Miller, C. W. (2013). Multi-frame super-resolution algorithm for complex motion patterns. *Optics express*, *21*(17), 19850-19866.
5. Kumar, M., & Singh, R. K. (2013, April). Digital image processing of remotely sensed satellite images for information extraction. In *Conference on Advances in Communication and Control Systems (CAC2S 2013)* (pp. 406-410). Atlantis Press.
6. Shao, P., Yang, G., Niu, X., Zhang, X., Zhan, F., & Tang, T. (2014). Information extraction of high-resolution remotely sensed image based on multiresolution segmentation. *Sustainability*, *6*(8), 5300-5310.
7. Wang, Z., Liu, D., Yang, J., Han, W., & Huang, T. (2015). Deep networks for image super-resolution with sparse prior. In *Proceedings of the IEEE international conference on computer vision* (pp. 370-378).
8. Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, *38*(2), 295-307.
9. Chen, X., Fang, T., Huo, H., & Li, D. (2015). Measuring the effectiveness of various features for thematic information extraction from very high resolution remote sensing imagery. *IEEE Transactions on geoscience and remote sensing*, *53*(9), 4837-4851.
10. Alparone, L., Aiazzi, B., Baronti, S., & Garzelli, A. (2015). *Remote sensing image fusion*. Crc Press.
11. Solanky, V., & Katiyar, S. K. (2016). Pixel-level image fusion techniques in remote sensing: a review. *Spatial Information Research*, *24*, 475-483.
12. Ma, C., Xia, W., Chen, F., Liu, J., Dai, Q., Jiang, L., ... & Liu, W. (2017). A content-based remote sensing image change information retrieval model. *ISPRS International Journal of Geo-Information*, *6*(10), 310.
13. Ha, V. K., Ren, J., Xu, X., Zhao, S., Xie, G., & Vargas, V. M. (2018). Deep learning based single image super-resolution: A survey. In *Advances in Brain Inspired Cognitive Systems: 9th International Conference, BICS 2018, Xi'an, China, July 7-8, 2018, Proceedings 9* (pp. 106-119). Springer International Publishing.
14. Zhang, S., Liang, G., Pan, S., & Zheng, L. (2018). A fast medical image super resolution method based on deep learning network. *IEEE Access*, *7*, 12319-12327.
15. Ravì, D., Szczotka, A. B., Shakir, D. I., Pereira, S. P., & Vercauteren, T. (2018). Effective deep learning training for single-image super-resolution in endomicroscopy exploiting video-

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
registration-based reconstruction. *International journal of computer assisted radiology and surgery*, *13*, 917-924.

16. Liang, S., & Wang, J. (Eds.). (2019). *Advanced remote sensing: terrestrial information extraction and applications*. Academic Press.
17. Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J. H., & Liao, Q. (2019). Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, *21*(12), 3106-3121.
18. Kawulok, M., Benecki, P., Piechaczek, S., Hrynczenko, K., Kostrzewa, D., & Nalepa, J. (2019). Deep learning for multiple-image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, *17*(6), 1062-1066.
19. Zhang, W., Liu, Y., Dong, C., & Qiao, Y. (2019). Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3096-3105).
20. Cheng, G., Xie, X., Han, J., Guo, L., & Xia, G. S. (2020). Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 3735-3756.
21. Liu, Y., Chang, M., & Xu, J. (2020). High-resolution remote sensing image information extraction and target recognition based on multiple information fusion. *IEEE access*, *8*, 121486-121500.
22. Zhang, H., Xu, H., Tian, X., Jiang, J., & Ma, J. (2021). Image fusion meets deep learning: A survey and perspective. *Information Fusion*, *76*, 323-336.
23. Kaur, H., Koundal, D., & Kadyan, V. (2021). Image fusion techniques: a survey. *Archives of computational methods in Engineering*, *28*(7), 4425-4447.
24. Jiang, J., Wang, C., Liu, X., & Ma, J. (2021). Deep learning-based face super-resolution: A survey. *ACM Computing Surveys (CSUR)*, *55*(1), 1-36.
25. Qiao, C., Li, D., Guo, Y., Liu, C., Jiang, T., Dai, Q., & Li, D. (2021). Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature methods*, *18*(2), 194-202.
26. Li, Y., Sixou, B., & Peyrin, F. (2021). A review of the deep learning methods for medical images super resolution problems. *Irbm*, *42*(2), 120-133.
27. Shan, L., Bai, X., Liu, C., Feng, Y., Liu, Y., & Qi, Y. (2022). Super-resolution reconstruction of digital rock CT images based on residual attention mechanism. *Advances in Geo-Energy Research*, *6*(2).
28. Zhou, L., Cai, H., Gu, J., Li, Z., Liu, Y., Chen, X., ... & Dong, C. (2022, October). Efficient image super-resolution using vast-receptive-field attention. In *European conference on computer vision* (pp. 256-272). Cham: Springer Nature Switzerland.
29. Wang, H., Wei, M., Cheng, R., Yu, Y., & Zhang, X. (2022). Residual deep attention mechanism and adaptive reconstruction network for single image super-resolution. *Applied Intelligence*, *52*(5), 5197-5211.

Received: August 10 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
30. Lepcha, D. C., Goyal, B., Dogra, A., & Goyal, V. (2023). Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Information Fusion*, *91*, 230-260.
31. Guerreiro, J., Tomás, P., Garcia, N., & Aidos, H. (2023). Super-resolution of magnetic resonance images using Generative Adversarial Networks. *Computerized Medical Imaging and Graphics*, 102280.
32. Gharibi, Z., & Faramarzi, S. (2023). Multi-frame spatio-temporal super-resolution. *Signal, Image and Video Processing*, *17*(8), 4415-4424.
33. Rikhsivoev, M., Yusupov, A., Begmatov, S., Arabboev, M., Nosirov, K., Saydiakbarov, S., ... & Khamidjonov, Z. (2023). Working Principles of Multi-Frame Image Super-Resolution. *Science and innovation*, *2*(Special Issue 3), 244-249.

Received: August 10 2025