Enlighten GAN for Super-Resolution Images from Surveillance Car

Pallavi Adke[1], Ajay Kumar Kushwaha[2](✉), Pratik Kshirsagar[1], Mayur Hadawale[1], and Prajwal Gaikwad[1]

[1] Pimpri Chinchwad College of Engineering and Research, Pune, India
[2] Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India
akkushwaha@bvucoep.edu.in

Abstract. Law enforcement and security officers utilize surveillance cars to monitor and investigate suspicious activities, including traffic violations and neighbourhood security. This paper delves into the comprehensive utilization of surveillance cars in modern society. By designing and developing an upgraded system, we aim to address ethical and legal concerns surrounding surveillance practices. The primary objective is to obtain super-resolution images from captured footage using Generative Adversarial Networks (GANs) for image enhancement. GANs, a type of Neural Network, enable the creation of high-quality images from existing low-resolution data. This study explores the application of GANs to enhance image quality in the context of surveillance cars. The proposed system leverages GANs to generate high-resolution images from the low-resolution ones captured by the surveillance car. Additionally, we provide a comprehensive review of the ongoing research and advancements in this field. Existing surveillance systems often output low-resolution images, but through the implementation of Enlighten GAN, we can achieve high-resolution results. The innovative integration of GAN technology empowers the surveillance car system to respond quickly to known situations with improved image clarity, enabling effective monitoring and investigation. This paper contributes to the advancement of surveillance capabilities by providing valuable insights into the potential of GANs for enhancing image quality and super-resolution in surveillance applications.

Keywords: Generative Adversarial Network (GAN) · Surveillance Car · Super Resolution Generative Adversarial Networks (SRGAN)

1 Introduction

Recently, in the world of machine learning, generative adversarial networks (GANs) have become a very exciting innovation. GANs are called generative models because they generate new data instances, which in turn resemble the training data used for training purposes. For example, GANs are used to generate images that resemble human faces that don't have any real identity or don't belong to a real person. GANs come under unsupervised learning that can automatically discover and learn about the patterns and

P. Pareek et al. (Eds.): IC4S 2023, LNICST 536, pp. 433–445, 2024.
https://doi.org/10.1007/978-3-031-48888-7_37

regularities in the testing data involved so that it can also generate the output, which is new examples that match the original dataset. GANs have a very unique and smart way of training the training dataset of the generator model by framing the issue as a supervised learning problem with two sub-models: generator and discriminator. The generator is used to create or generate fake data that resembles the original one, and the discriminator model tries to classify or separate the fake and the original data. Both the models are trained simultaneously until the discriminator is unable to tell the difference between real and fake images; hence, we can understand that the generator is trained and is generating plausible images enough to fool the discriminator. Now, when trying to understand what super-resolution is, which is based on the idea that images with low resolution (noisy) or containing some disturbances are used to generate HR images, it creates original images with HR when it gives it images in LR. Now achieve super-resolution in the images that it obtains from the surveillance car. With the ESPCAM32 camera module, the surveillance car can take photographs and videos, and the output can be displayed on any screen via the hotspot connection to the ESP. With the help of a Bluetooth module and an Android device, the car can be controlled. This surveillance car's purpose is to capture videos and images from closed areas, where normally CCTV cameras don't look good or suffer from black zones. It is a mobile device, hence does not have any black zones, and can be very handy in the surveillance of any suspicious activity in the area. Now, with the help of GAN, it can construct HR images from the output of the surveillance car. CCTV output can also be processed using GAN to get highly resolved images.

2 Related Work

Ronneberger et al. [1] have developed a more advanced architecture called the "fully convolutional network" to improve the accuracy of image segmentation with very few training images. They added a continuous layer to the regular contract network and replaced the pooling operator with an upsampling operator to increase the resolution of the output. Goodfellow et al. [2] discussed an adversarial learning framework for training generative and discriminative models, where both models are multi-layer perceptrons. This approach, called adversarial nets, can be trained using back propagation and dropout algorithms, and sampling from the generated models can be done using only forward propagation. Yuntan et al. [3] applied adversarial learning to semi-supervised segmentation training and achieved performance comparable to fully supervised training with only half of the labelled data. Kulkarni et al. [4–6] discussed advanced techniques for non-linear filters, diffusion processes, wavelet denoising methods, thresholding techniques, and filters for fractional arithmetic. Zhang et al. [7] try to exploit the recently introduced conditional generative adversarial networks (CGAN) excellent generative modelling capabilities by applying an extra constraint that the de-rained image must be indistinguishable from its equivalent ground truth clean image. Xiang et al. [8] presented a deep neural network architecture for single-image rain removal called feature-supervised generative adversarial network (FS-GAN). Vaishnave et al. [9] identified gaps in the literature in order to spur the development of new data-driven algorithms and presented quantitative metrics for evaluating scene classification in satellite imagery. Zhang et al.

[10] suggested a GPS integration and semantic estimation method for decreasing per-pixel noise in region-like tasks that is simple to integrate into various backbone GAN designs. Devabalan [11] discussed the challenge of processing and analyzing remote sensing data quickly, and Isola [12] suggested conditional adversarial networks as a promising approach for many frame-to-frame translation tasks. Laxman [13] proposed a multi-scale gradient-based U-Net (MSG U-Net) for high-resolution frame-to-frame transformation. Park et al. [14] discuss the application of a loss function in the training of a fused network to reduce GAN-generated artefacts, restore fine details, and maintain colour components. The SRGAN method employs spectral normalization techniques to ensure Lipschitz continuity and mitigate the vanishing gradient issue. RaGAN, used in ESR-GAN, predicts a relative decision instead of a definitive one. Wassertein GAN, or WGAN, utilises IPM metrics modules for divergence minimization and provides a nonactivation value as output, which can be taken as a high Lipschitz metric. Gradient clipping can be used to modify weight parameters in one batch, making it an effective technique for GANs as well.

3 Methodology

GANs (Generative Adversarial Networks) consist of two deep networks: one is a discriminator network and the other is a generator network. Some parameters and Variables are as.

D = Discriminator.
G = Generator
θ_d = ParametersofDiscriminators
θ_g = ParametersofGenerator
$P_z(z)$ = InputNoiseDistribution
$P_{data}(x)$ = Originaldatadistribution
$P_g(x)$ = Generateddistribution

The discriminator is given both real and phony images, and it attempts to distinguish between the two. Returns a chance of the image being "true" in the range of 0 and 1. The generator tries to deceive the discriminator into believing the bogus images it produces are real images. Educating a GAN.

The binary entropy loss can be given as –

$$\log\big(\hat{y}, y\big) = [y.log\hat{y} + (1-y).\log\big(1-\hat{y}\big)] \tag{1}$$

where y = Original Data
$\hat{y}$ = Reconstructed Data
Discriminator Loss:

$$L^{(D)} = \max[\log(D(x)) + \log(1 - D(G(z)))] \tag{2}$$

Generator Loss:

$$L = min_G max_D[\log(D(x)) + \log(1 - D(G(z)))] \tag{3}$$

3.1 Training a GAN

Part 1: While the Generator is not in use, the discriminator is trained. The network is only propagated forward; no back propagation is carried out. The discriminator is tested to determine if it can accurately identify them as real after being trained on real data for n epochs. Also, the Discriminator is trained on the fictitious generated data from the Generator at this phase to determine if it can correctly identify them as fictitious. Discriminator training is shown in Fig. 1.

Part 2: While the Discriminator isn't in use, the Generator is being taught. After the Discriminator has been trained using the Generator's fabricated data, we may utilize the predictions to train the Generator and improve the Discriminator's prior state. Generator training is shown in Fig. 2.

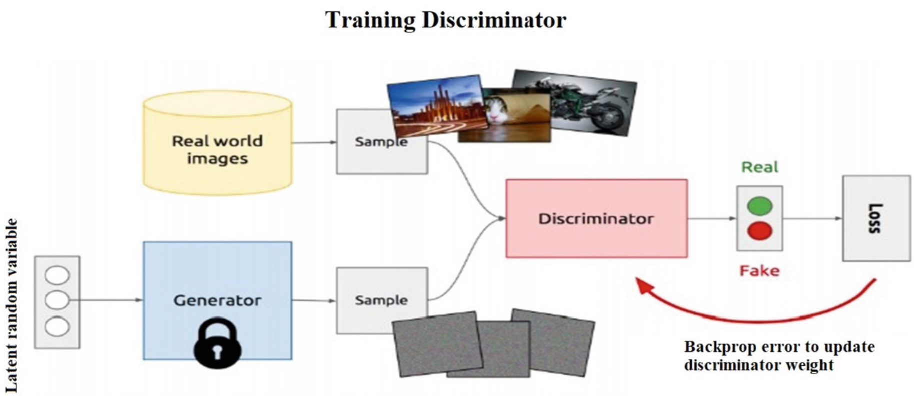

Fig. 1. Training Discriminator

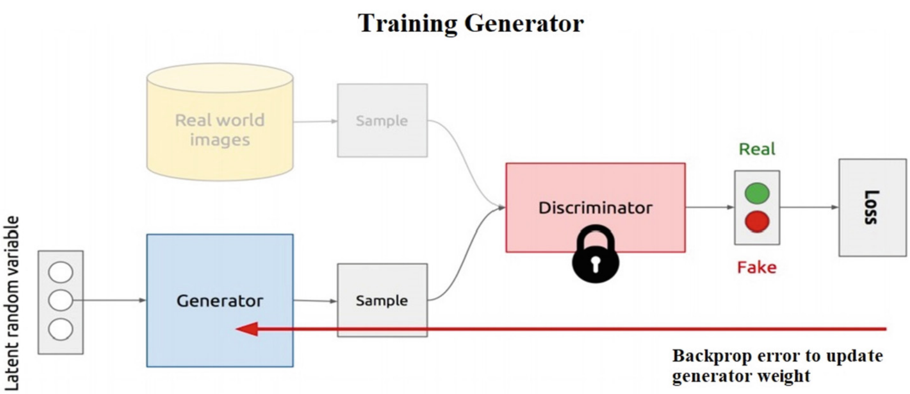

Fig. 2. Training Generator

3.2 SRGAN (Super Resolution Generative Adversarial Networks)

As part of an ongoing effort, one of the generative adversarial networks applications: SRGAN has been put into practice. By upscaling an image from a specified low-resolution to a comparable high-resolution image with greater visual quality, a process known as "super-resolution" is used. High-resolution photographs offer better reconstructed details of the settings and individual objects, or enhanced details as they may appear at any contemporary crime scene. An example of SRGAN is shown below in Fig. 3.

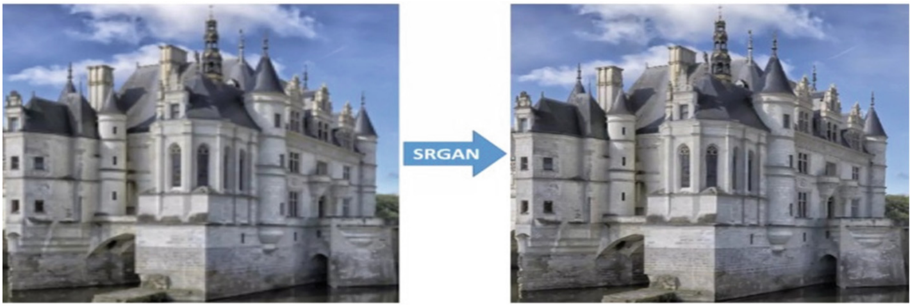

Fig. 3. Example of SRGAN

3.3 SRGAN Architecture

The generator architecture, which is used to produce excellent super-resolution images, is essentially a fully convolutional SRResNet model. To ensure that the overall architecture adapts appropriately to the quality of the photos, the discriminator model, which also serves as an image classifier, has been added. The images produced as a result are significantly better. Generator architecture is shown in Fig. 4. Given that the neural network can select the best suitable value on its own in this situation, parametric ReLU is preferred.

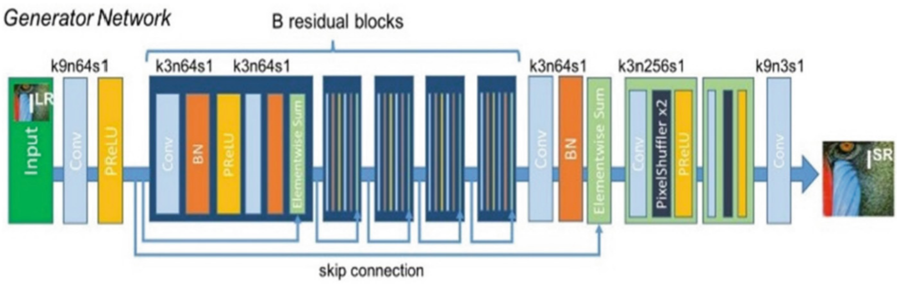

Fig. 4. SRGAN Generator Architecture

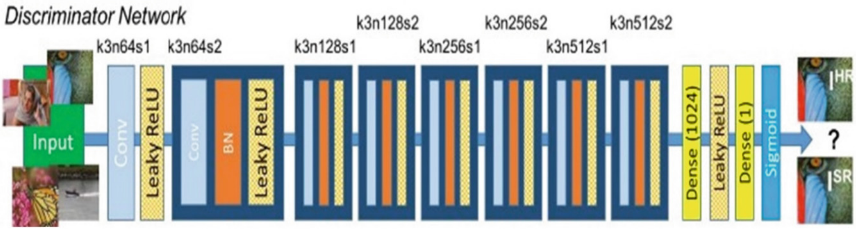

Fig. 5. SRGAN Discriminator Architecture

The discriminator architecture in the figure above functions to distinguish between genuine images and super-resolution images. The adversarial min-max problem is addressed by the discriminator model that is built. Discriminator architecture is shown above in Fig. 5.

A large number of residual blocks are used in the following layer of the feed-forward fully coevolutionary super-resolution reconstruction model (SRResNet) model. Following a batch normalization layer, a Parametric ReLU activation function, a convolutional layer with batch normalization, another convolutional layer with batch normalization, and a final elementwise sum technique, each residual block has a convolutional layer with 33 kernels and 64 feature maps. The feed-forward output and the skip connection output are both used by the elementwise sum method to produce the final output. After 4x upsampling the convolutional layer in this model architecture, super-resolution images are created using the pixel shuffler. The categorization action is carried out via the sigmoid activation function.

3.4 SRGAN Model Training

To build the SRGAN model and train it as needed for this project, we will use the VS Code platform.

Bringing in the necessary libraries Implementing all the crucial libraries needed to complete the ensuing tasks is the initial step in starting our SRGAN project. We have set up each and every necessary library.

Perceptual Loss:
Loss function for perception the effectiveness of our generator network depends on how our perceptual loss function, 1 SR, is defined and design a loss function that evaluates a solution with regard to perceptually meaningful properties, even if l SR is frequently represented based on the MSE. The weighted sum of a content loss (l SR X) and an adversarial loss component is how it define the perceptual loss as follows in Eq. 4:

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR} \tag{4}$$

where $l_X^{SR} = ContentLoss$, $10^{-3} l_{Gen}^{SR} = AdversarialLoss$

3.5 Preparing the Dataset

The datasets are in the form of individual zip files. These files contain training and validation files for both low-resolution and high-resolution images. Once the download is complete, it extracts them accordingly. For low-resolution images, the size is 96 pixels by 96 pixels, and for high-resolution images, it is 396 pixels by 396 pixels. For training purposes, the same pixel size is required. The low-resolution image data set must also be equal to the high-resolution image data set. This means if it takes 500 LR images, it must use 500 h images; the number should be the same. Training the SRGAN Model After successfully setting up the libraries and collecting the datasets, it will construct the SRGAN architecture and start to train the model [15–17].

Sample outputs of the proposed trained Model are shown in Fig. 6 and 7.

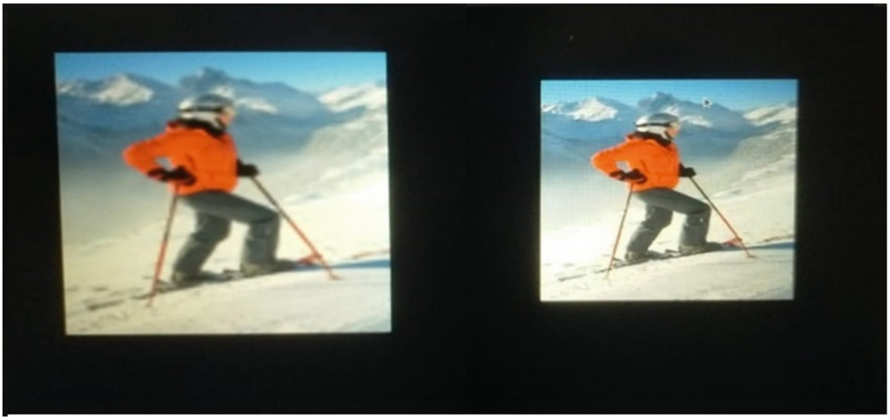

Fig. 6. Model training Result 1

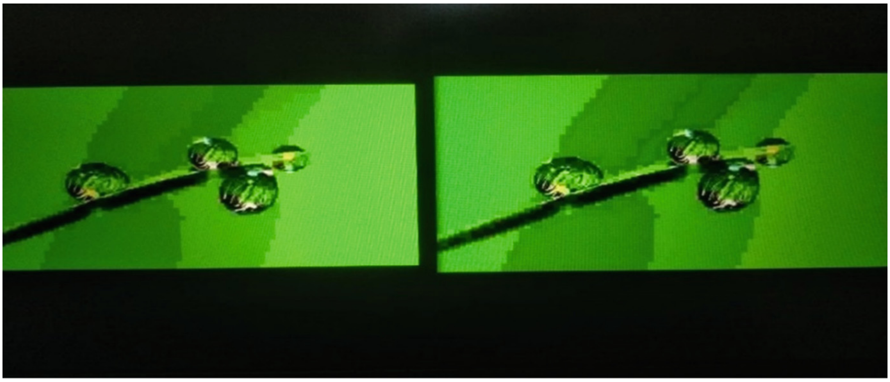

Fig. 7. Model training Result 2

4 Dataset

It uses two types of datasets: one that was obtained from the surveillance car and created, and another from Image Pair (Microsoft).

1. **Manual Dataset:** It is obtained as an output from around 1000 images from our surveillance car. These images were resolved and obtained as HR and LR images for the training of the model so that the model becomes familiar with the output of our car. It also obtained testing data from the car that was used for the input of the GAN to generate its highly resolved reconstruction. The images were captured in all formats to ensure proper training for the model.
2. **Image Pair Dataset:** Realistic Super Resolution refers to generating HR images from some or all the other low-quality images of the same scenery. The problem here is that low-resolution images completely erase the HR visual frequency information in the image.

Many deep and machine learning methods have been put into use or proposed to demonstrate the development of a model that is used to recover lost characteristics of a particular image in the generated one. Hence, using this strategy, deep and machine learning methods have been able to solve super resolution problems.

Data is very important to machine learning, especially when the DL algorithms are data-driven, where data drives the whole algorithm. Now, to solve the problem, gather and form the information and acquire and generate data, which may be just as important to the solution as solving the issue. This research aims to present a new way of collecting real data using various novel data acquisition techniques. Super resolution, noise deprecation, and quality improvement techniques can use this as an input. Click the same image with the same scenery with a low-resolution camera and a high-resolution camera. It took around 11000 images that are low and high-resolution as training data. And it can use testing data to convert those images into SR images.

5 Result Analysis

The performance of the model can be evaluated by using following parameters:

1. **SNR:** The square root of the quantity of photons in the image's brightest region is used to calculate the signal-to noise ratio (SNR) of a digital microscopic image. The ideal range is from 20 to 40.
2. **PSNR:** This ratio is often used to compare the original and compressed images. The ideal range is 30–50 dB.
3. **Jaccard's index:** A statistic used to assess the similarities and diversity of samples is the Jaccard index. It is commonly referred to as the Jaccard similarity coefficient.

Image compression 1 and image compression 2 are shown in Figs. 8 and 9, respectively. The result of the analysis is mentioned in Table 1.

Fig. 8. Image comparison 1

Fig. 9. Image comparison 2

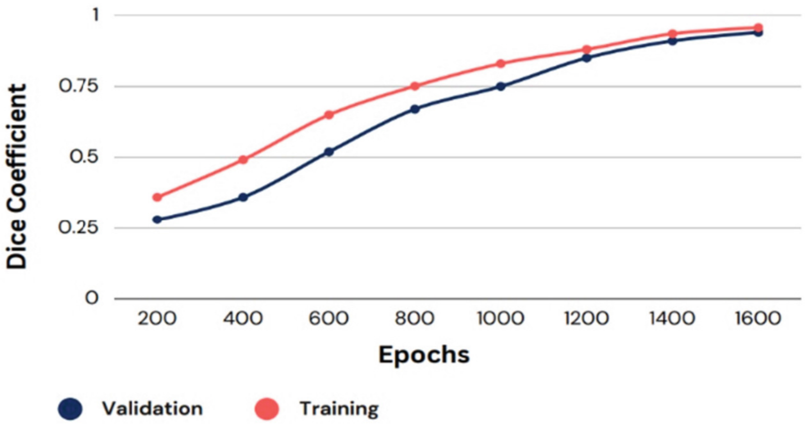

Fig. 10. Training accuracy verses Validation accuracy

Table 1. Result Analysis

	Image Pair 1	Image Pair 2	Image Pair 3	Image Pair 4	Image Pair 5
Jaccard's Index	0.9668	0.8929	0.9759	0.9742	0.9474
PSNR	44.20	43.08	44.57	42.06	44.75
SNR	25.20	19.76	25.39	24.65	26.84

6 Advantages

The advantages of the proposed system can be summarized as follows:

1 With the help of super resolution, it can identify the culprits and submit the results as evidence.
2 A surveillance car being mobile means it does not need to move, and the car can also go places where humans would not dare.
3 It can quickly identify trees, roads, bikers, people, and parked automobiles, and even compute the distance between various objects using GAN and machine.
4 GANs generate data similar to the original data.
5 GANs go into the details of the data and can be easily interpreted into different versions. This is useful when working with machine learning.
6 It can improve model generalization.

The hardware is shown in the Fig. 11.

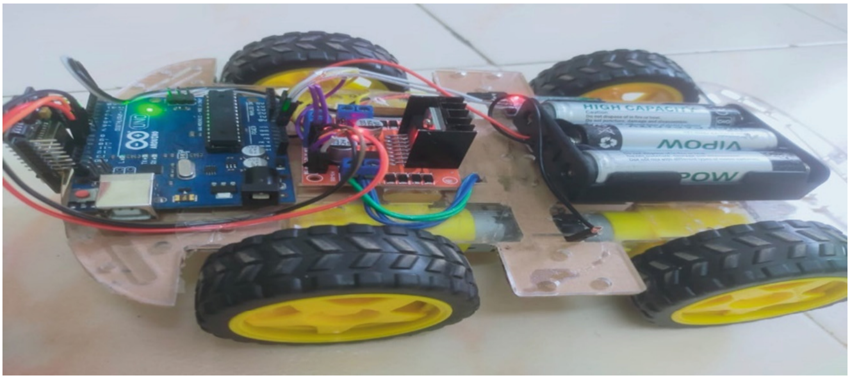

Fig. 11. Hardware model

7 Comparison

The advanced surveillance car represents a substantial leap forward from older surveillance methods. Integrating cutting-edge components like the ESP CAM 32, a Bluetooth module, and an Android application for control, our system showcases significant advancements. Particularly noteworthy is the incorporation of SRGAN (Super-Resolution Generative Adversarial Network) technology, which enhances image quality to unparalleled levels. This integration introduces several transformative features and improvements to the surveillance process.

1. **Real-time Remote Control:** Operators can precisely navigate the car using the Android app, responding swiftly to emerging situations.
2. **Cost-effectiveness:** The system's integration of readily available components reduces hardware costs without compromising image clarity.
3. **Adaptable and Versatile:** With a modular design and open-source components, the surveillance car can be upgraded with emerging technologies.
4. **Real-time Monitoring and Analysis:** The Android app provides instant access to live video feeds and data insights, optimizing responsiveness and efficiency.

8 Conclusion

In this paper, a cutting-edge surveillance system that combines the power of Enlighten GAN with a mobile robot car is described to address blind spots in existing CCTV and surveillance setups. The system is designed to venture into areas where traditional surveillance systems might be limited, ensuring comprehensive coverage and enhanced security. At the core of the solution lies a highly versatile robot car equipped with state-of-the-art technology that can be remotely operated for secure and efficient surveillance. The robot car's mobility allows it to access remote or hard-to-reach locations, enabling a broader scope of monitoring and investigation. To facilitate seamless control and communication with the robot car, it incorporated a robust Bluetooth module. This feature ensures reliable and real-time data exchange between the mobile surveillance unit and the controlling device, enhancing responsiveness and usability. The Android app complements the system, providing a user-friendly interface for operators to control and manage the robot car remotely. Through the app, users can initiate surveillance sessions, receive live video feeds, and access real-time data insights, bolstering the system's overall surveillance capabilities. In addition to the mobility and control aspects, the system addresses the challenge of low-resolution imagery commonly encountered in surveillance scenarios. It leverages the power of Enlighten GAN to achieve super-resolution images captured by the surveillance system or robot car. This sophisticated image enhancement technique enables us to obtain clearer, high-resolution images, thereby improving the accuracy of person and number plate recognition. By integrating the robot car, Bluetooth module, Android app, and super-resolution image processing, the surveillance system stands as a comprehensive and effective solution for tackling blind spots in surveillance coverage. The combination of advanced components ensures seamless operation, enhanced data quality, and increased surveillance efficiency, empowering law enforcement and security personnel with a powerful tool to maintain public safety and security effectively.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. Int. Conf. Med. Image Comput. Comput-Assis. Interv. Springer Int. Publishing **18**, 234–241 (2015)
2. Goodfellow, I., et al.: Generative adversarial nets. Adv. Neural. Inf. Process. Syst. **27**, 2672–2680 (2014)
3. Tan, Y., Wu, W., Tan, L., Peng, H., Qin, J.: Semi-supervised medical image segmentation based on generative adversarial network. J. of New Media, 4, 3, 155, (2022)
4. Kulkarni, P., Madathil, D.: A review on echocardiographic image speckle reduction filters. Biomed. Res. **29**(12), 2582–2589 (2018)
5. Kulkarni, P., Madathil, D.: A review of echocardiographic image segmentation techniques for left ventricular study. ARPN J. Eng. Appl. Sci. **13**(10), 3536–3541 (2018)
6. Kulkarni, P., Madathil, D.: Fully automatic segmentation of LV from echocardiography images and calculation of ejection fraction using deep learning. Int. J. Biomed. Eng. Technol. **40**(3), 241–261 (2022)
7. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. IEEE Trans. Circuits Syst. Video Technol. **30**, 3943–3956 (2019)
8. Xiang, P., Wang, L., Wu, F., Cheng, J., Zhou, M.: Single-image de-raining with feature-supervised generative adversarial network. IEEE Signal Process. Lett. **26**, 650–654 (2019)
9. Vaishnnave, M.P., Devi, K.S., Srinivasan, P.: A study on deep learning models for satellite imagery. Int. J. Appl. Eng. Res. **14**, 881–887 (2019)
10. Zhang, Y., Yin, Y., Zimmermann, R., Wang, G., Varadarajan, J., Ng, S.K.: An enhanced GAN model for automatic satellite-to-map image conversion. IEEE Access **8**, 176704–176716 (2020)
11. Devabalan, P.: Satellite image processing on a grid based computing environment. Int. J. Comput. Sci. mobile Comput., 3, 1039–1044, (2014)
12. Isola, P., Zhu, J. Y., Zhou, T., Efros, A. A.: Image-to-image translation with conditional adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1125–1134, (2017)

13. Kumarapu Laxman, Laxman, K., Dubey, S.R., Kalyan, B., Kojjarapu, S.R.V.: Efficient high-resolution image-to-image translation using multi-scale gradient U-Net. In: International Conference on Computer Vision and Image Processing, Springer International Publishing, (2021)
14. Park, J., Han, D.K., Ko, H.: Fusion of heterogeneous adversarial networks for single image dehazing. IEEE Trans. Image Process. **29**, 4721–4732 (2020)
15. Kulkarni, P., Madathil, D.: Echocardiography image segmentation using semi-automatic numerical optimisation method based on wavelet decomposition thresholding. Int. J. Imaging Syst. Technol. **31**(4), 2295–2304 (2021)
16. Leclerc, S., et al.: Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans. Med. Imaging **38**(9), 2198–2210 (2019)
17. Kushwaha, A.K., Khatavkar, S.M., Biradar, D.M., Chougule, P.A.: Depth estimation and navigation route planning for mobile robots based on stereo camera. Lect. Notes Inst. Comput. Sci. Soc. Inform. Telecommun. Eng. **472**, 180–191 (2023)