# Quality Optimization of Live Surveillance Video With Crime and Anomaly Identification

Deepika Ajalkar[1], Masira Kulkarni[2], Akshay Popale[3], Abhishek Mahindrakar[4], Ashish Meshram[5], Prabal Khillarkar[6]

{dipikaus@gmail.com[1], masirashaikh96@gmail.com[2], akshaypopale@gmail.com[3], theabhishek3040@gmail.com[4], ashishmeshram229@gmail.com[5], prabalannakhillarkar@gmail.com[6]}

Department of Data Science, G. H. Raisoni College of Engineering and Management, Wagholi, Pune, 412207, Maharashtra, India.[1,2,3,4,5,6]

**Abstract.** Generative adversarial networks (GANs) and Super-Resolution Convolutional Neural Networks (SRCNNs) are AI models enhancing real-time CCTV footage. Extensive experiments on real-world datasets confirm their superiority over traditional upscaling methods. This tech not only improves visual clarity but aids in identifying suspects, license plates, and evidence. Generative AI also excels in crime scene identification by automatically spotting objects, vehicles, and individuals. This boosts law enforcement efficiency. However, ethical concerns, like privacy and biases, demand careful consideration when implementing AI surveillance. In summary, generative AI enhances CCTV footage, aiding crime investigation, but ethical issues must be addressed for responsible deployment. Index Terms—CCTV footage upscaling, deep learning, real-time processing, crime scene identification, object detection, convolutional neural network, surveillance systems.

**Keywords:** Video Quality Enhancement, Surveillance Video, Real-time Crime Detection, Anomaly Detection, Object Detection, CCTv Footage Optimization.

## 1 Introduction

CCTV cameras are frequently utilized for security and surveillance tasks, giving important footage for criminal investigation and crime prevention. However, one of the significant challenges in CCTV surveillance is dealing with low-resolution video streams, which often hinder the accurate identification of individuals and events. To address this limitation, researchers and developers have turned to the power of generative AI, particularly employing Generative Adversarial Networks (GANs). GANs are deep learning models that can generate realistic data samples, such as images or videos, by learning from existing data. Leveraging GANs, it becomes possible to upscale and enhance low-resolution CCTV footage in real-time, thereby improving the quality and clarity of the video streams.

ICISML 2024, January 05-06, Pune, India

DOI 10.4108/eai.5-1-2024.2341878

### 1.1 Generative Adversarial Networks (GANs)

A GAN is made up of two neural networks, the discriminator and generator that are engaged in a zero-sum game of competition. The viewer attempts to discern between the genuine image and the created image when the generator creates a high-resolution version of the low-resolution image. Competition between these networks leads to the emergence of valuable and valuable products.

### 1.2 Real-time Video Processing

An important part of the project is to ensure the efficiency of the scale-up process. Real-time video production requires efficient processing and hardware to perform calculations and provide smooth, real-time output.

### 1.3 Crime Scene Identification

The aim of the project is to create a crime report as well as progress. Using computer vision and object detection tools, the system will be able to detect and identify specific objects or events in CCTV footage that may indicate a crimor behavior.

### 1.4 Overview of the Dataset

The DIV2K dataset, which has 800 excellent 2K resolution photos, is a benchmark dataset for image super-resolution. It is to assess algorithms that improve the reso lution of images. With a wide range of scenes, objects, and textures, the dataset is diverse. It is frequently divided into training and validation sets and contains color images. Algorithm performance is evaluated using metrics like PSNR and SSIM. DIV2K is used by researchers to develop and test super-resolution image techniques in both academic and competitive contexts. For the most recent information on the dataset, always consult the official documentation.

## 2 Literature Survey

**Table 1:** Literature Survey.

| **Sr. No.** | **Published Year [References]** | **Research Paper Name** | **Description** |
|---|---|---|---|

| | | | |
|---|---|---|---|
| 1 | 2017 | Abnormal event detection in videos using generative adversarial nets | Detect anomalous events in videos using generative adversarial networks. This article addresses the problem of detecting anomalous events in crowded scenes. We propose to learn an internal representation of scene normality using a generative adversarial net (GAN) trained using regular frames and corresponding optical flow images. Since our GAN is trained only on normal data, it cannot generate abnormal events. During testing, real data is compared with both the appearance and behavior representations reconstructed by his GAN, and abnormal regions are detected by calculating local differences. |
| 2 | 2018 | Crime Scene Prediction by Detecting Threatening Objects Using Convolutional Neural Network | Predicting crime scenes without human intervention could have a huge impact on computer vision. In this paper, we present CNN (Convolutional Neural Network) in the use of detecting knives, blood and guns from images. Detecting these threatening objects from an image can allow us to predict whether a crime will occur and where the image was taken. |
| 3 | 2022 | Anomaly Detection in Video Surveillance using Deep Learning | Effective detection of anomalies in surveillance video under various conditions solves an important challenge in PC vision. This work proposes a background inference method using deep learning techniques of residual neural networks equipped to recognize different action objects of different sizes per segment. foreground segment per pixel. The proposed algorithm accepts input as a reference source and a target frame, which needs to be temporally adjusted, and generates instructions to segment the same spatial target. |

| | | | |
|---|---|---|---|
| 4 | 2020 | Generative Adversarial Network for Image Super-Resolution Combining Texture Loss | With the rise of the Internet and the rise of information technology, people are getting more and more information every day. The main way people get information is through images, videos and audio. Studies have shown that people get 60%∼80% of their information through vision, so seeing images is an important way to get information. But the quality of images can be limited by hardware stuff like imaging systems and how much bandwidth they use during image transmission. |
| 5 | 2020 | Low light image enhancement with adaptive sigmoid transfer function | Low-light image enhancement algorithms aim to produce visually pleasing images and to extract valuable information for computer vision applications. The task of improving image quality in low light conditions is a challenge. Current quality improvement methods are undeniably hindering visual aesthetics and have the major drawback of high computational complexity and lower efficiency. |

# 3 Methods

## 3.1 Video Quality Optimization using GAN

### 3.1.1 Introduction to GANs

Deep learning methods like neural networks are used in generative adversarial networks, a modeling methodology. In order to build or create new models that can be inferred from the original data, generative modeling is an unsupervised activity in machine learning that entails finding and learning patterns in the input data.

By posing the problem as a supervised learning task with two models - training power models that produce new models and attempting to categorize examples as authentic (based on their names) or false (generated) - GANs are an intelligent approach to train models. In a zero-sum game where two examples are trained together, the generator produces consistently high-quality samples until the example's discrimination is erroneous in half of the cases.

GANs are an exciting and rapidly changing field that holds the promise of models capable of producing realistic models across a wide range of problems, especially in imageto-image translation, such as translating summer photos into winter or daytime photos into night. and create realistic images of objects, scenes, and people that even humans can't tell are fake.

#### 3.1.2 GAN - Super Resolution GAN (SRGAN)

Image enhancement is a dynamic field that is moving away from old statistical techniques like bicubic interpolation and toward machine learning (ML) algorithms. A machine learning technique called super-resolution GAN (SRGAN) can upsample images to extremely high resolution.

To create high resolution images (up-to four times the original resolution), SRGAN uses deep neural networks in conjunction with GAN opponents. These develop more accurate and better-performing solutions, and frequently greater MOS (Points of Opposition) scores. High-resolution photos are first used to train SRGAN. sent to the generator after being down sampled into images of poor resolution. Then, the generator tries to produce the picture at a high resolution. The resulting results are compared using the discriminator. original high-resolution images combined with a super-resolution images. The GAN then transitions from splitter to splitter and to the generator once more.

#### 3.1.3 Dataset

The model was trained using the DIV2K dataset, which includes excellent photos at DIVerse 2K resolution. A list of altered files with various genres of videos, including 10 to 15 second cartoons, movies, sports, drama, animation, and music, was made for practical purposes. Compare films with various pixel counts - 144p, 240p, and 360p, for example - to actual photos.

The go-to resource for article-based image explanations is the Flickr30k collection. In this post, we introduce the Flickr30k Site, which links quotations from several articles of the same organization to the same image and associates them with 276 thousand books. The site supports Flickr30k's 158 thousand articles with 244 thousand common citation chains. For automatic picture recognition to advance further and for the language to be easily understood, this clarification is required. They enable us to create new measures in the area of picture information. We offer a fundamental framework for this job that integrates bias selection, color separation, product detection, and graphic design. Despite the fact that our kernel compares to the finest state-of-the-art models, we discover that the latter's results are difficult to apply to advancements in water count, such as inverted imaging, highlighting the shortcomings of the existing approach. more is required. research.

#### 3.1.4 GAN architecture for improving video quality

Refer fig. 1

### 3.2 Deep Learning Crime Anomaly Detection

#### 3.2.1 The foundational theories and concepts of anomaly detection

a Normal vs. Anomalous Data: Data that is normal vs. abnormal Anomalies are departures from this behavior, and anomaly detection occurs when the majority of the data in the data set is "normal" or follows a pattern. It is possible to take advantage of inconsistencies for a variety of reasons, such as data points that are distant from the center, unforeseen spikes or dips in time series data, or uncommon events that happen seldom.

b Unsupervised Learning: A record identifying which elements are normal and which are missing is not typically required while performing an unaudited task like anomaly detection. To find weaknesses based on product features, it instead uses statistics and machine learning.

c Statistical Methods: Statistical techniques are frequently employed to identify uncertainty. The ”normal” range of the data is determined using a variety of statistical metrics, including mean, variance, and standard deviation. Other information content is regarded as being untrustworthy.

d Threshold-Based Detection: Setting a threshold is one of the simplest approaches to find vulnerabilities. Data material that exceeds this limit will be labeled suspicious. Finding the ideal beginning place can be challenging and frequently needs talent.

e Time Series Anomaly Detection: In order to investigate parameters in time series data, it is necessary to spot odd patterns or trends across time. Applications like those for fraud detection, network monitoring, and preventing future problems. Anomalies in time series can be found using techniques like moving average and exponential sweeping and seasonal breakdown.

f Machine Learning Model Evaluation: Due to the lack of error codes, it might be challenging to assess the effectiveness of a machine learning model for mistake detection. Examples of common evaluation metrics include precision, recall, F1 score, and receiver operating characteristic (ROC) curve.

g Online vs. Offline Anomaly Detection: Online anomaly detection uses current data, while offline anomaly detection uses historical data. Applications that call for fast action, like network security, depend on real-time detection.

h False Positives and False Negatives: Anomaly detection systems frequently strike a balance between negative (no anomaly) and negative (normal data incorrectly categorized as abnormal). The decision is based on the particular application and its results.

### 3.2.2 Constructing a Dataset to Look for Crime Anomalies

a UCF-Crime: There are 128 hours of photos in the UCF-Crime dataset, which is a fairly sizable dataset. consists over 1,900 uncensored, full-length security footage from 13 horrifying situations, involving torture, arrests, arson, agony, accidents, crashes, explosions, fights, robberies, gunshots, robberies, and store explosions. Torture and theft. These flaws were chosen because they have a big effect on public safety. There are two things you can do with this knowledge. First, all abnormalities in one group and all actions in the other are included in the anomaly identification process. Next, list 13 uncommon jobs.

b XD-Violence: XD-Violence is a sizable database for researching violence in motion pictures.

c VFP290K: A new, extensive research document on fallen people called VFP290K: Vision-Based Fallen Persons (VFP290K) contains pictures of fallen people that have been gathered. in a variety of circumstances all across the world. 294,714 men were taken out of 178 recordings, spanning 131 situations in 49 different locales, for VFP290K.

### 3.2.3 Machine Learning Anomaly Detection Algorithms

Machine learning has a wide range of defect detection algorithms that can identify patterns and distinguish between expected actions. Algorithms for machine learning are crucial for finding anomalies. The following is a list of a few well-liked anomaly detection algorithms:

a Isolation Forest: The classification forest approach divides conflicts in unsupervised learning using individual trees. The split values that fall within the feature's range are then randomly chosen after a feature has been chosen at random. Anomalies can be found more rapidly since fewer segments are needed to tell them apart from other data.

b One-Class Support Vector Machines (SVM): A common method of anomaly detection called One Class SVM seeks to find hyperplanes that cover the majority of the data while reducing the inclusion of anomalies. The model is constructed from training data, and testing data is categorized as normal or abnormal based on how closely it resembles the model.

c Autoencoders: Autoencoders are neural network models that produce the input data. Large-scale reconstruction issues are thought to be negligible because autoencoders are trained on reliable data for error detection. Autoencoders can capture intricate and erratic patterns, making them valuable for detecting anomalies.

d Gaussian Mixture Models (GMM): GMM makes the assumption that data are produced from a continuous Gaussian distribution. GMM assesses parameters using data with low probability. When the data typically has more than one Gaussian distribution, this technique performs well.

e Local Outlier Factor (LOF): The local variance of the data in relation to its neighbors is determined by the LOF algorithm. Data points that have a lower village density than their surrounding villages are seen as suspicious. LOF is very helpful when trying to find discrepancies in datasets.

f Support Vector Data Description (SVDD): SVDD is a variant of SVM that encapsulates frequencies by surrounding the data with a hypersphere. The hypersphere excludes data points that are deemed questionable.

g Random Forests: By training a decision tree forest on reliable data, random forests can also be utilized for blind search. Low average numbers of forest trees or low average voter turnout are regarded as dubious data elements.

### 3.3 Integration of GAN and Deep Learning

#### 3.3.1 Combining video quality improvement and anomaly identification

a Ensemble Learning: Using ensemble learning to combine anomaly detection and video quality improvement: In ensemble learning, various models are combined to enhance overall performance. There are various methods, including:
Voting: Combine predictions from many models and select the one that receives the most support (for classification) or the most votes (for regression).
Stacking: Present a metamodel that takes input from the outputs of other models.
Bagging and Boosting: To merge various base models, use approaches like Random Forest (bagging) or AdaBoost (boosting). When you have multiple models, each with a particular set of attributes, a combination works effectively.

b Sequential Models: Models can be combined in a sequential manner, with one model's result being used as the input for another. This is frequently used in architectures in the pipeline style.

c Model Fusion: Model fusion is the process of combining the outcomes of various models of certain processes, typically in a neural network, to produce a cohesive representation. Connecting, adding, or using other suitable functions can accomplish this.

d Multi-Modal Learning: When processing information from various sources or formats (such as text and images), you can segregate the information into distinct formats for each type of data before combining it using the fusion method. This is typical in apps like captioning, where you may make captions by fusing together text and photos.

e Knowledge Transfer: Use one model as a feature extractor for previously learnt features and a second model that has been optimized for a particular purpose. The weight of the previous model is frequently utilized as a starting point for training a new working model in transfer learning.

f Attention Mechanisms: Follow the tracking process, which enables one model to concentrate on a certain input based on the output of another model, to use attention mechanisms. This is specifically true for system-dependent operations like machine translation.

g Recurrent Integration: You can utilize neural networks (RNN) or changes therein to combine data from various samples across time if you have a data set.

h Custom Architectures: To integrate different models in a way that best meets your objectives, you may need to design a custom neural network architecture depending on the details of your particular challenge.

i Transferring knowledge and perfecting it: Use the weights from another model, which could be from a different variable or function, to fine-tune your current model. If you just have a small amount of data to work with, this approach is particularly helpful.

j Design a neural network architecture that mixes many neural networks into shared or dispersed networks. This is known as a hybrid model. Different learning modes may be available for each sub-network.

### 3.3.2 Information on the implementation and system architecture

Refer fig. 3.

## 4 Related Work

Detection of irregularities. Anomaly detection is one of the most challenging and pervasive problems in computer vision. [1, 3, 4, 7, 8, 8, 11, 12, 18, 19, 20, 21]. For video surveillance applications, there have been several attempts to detect aggression or antagonism [6, 10, 11, 15]. In order to identify human hostility, Datta et al. advised looking at a person's movements and limb orientation. Kooij and others [10] searched surveillance footage for hostile activity using audio and visual data. Gao et al. reported violent flow characteristics in order to detect violence in crowd recordings. modern times Mohammadi et al. suggested a brand-new classification method based on behavior heuristics. [15] Videos that are both violent and gentle Earlier this year, Mohammadi et al. [15] suggested a brand-new method based on behavior heuristics for categorizing violent and non-violent videos. The authors of [7] suggested employing tracking to model people's typical motion and highlight divergence from that motion as an anomaly, in addition to differentiating between violent and non-violent behaviors. Due to the difficulties in locating reliable traces, there are a number of techniques to prevent tracking and identify patterns of global mobility. Histograms, topic modeling, motion patterns, and mixed dynamic texture models with social force are a few examples. Other examples include the Hidden Markov Model (HMM) on Local Spatial Temporal Volumes [11] and a context-driven method [21]. Taking into account the training examples that highlight typical behaviors and these tactics The distributions of common motion patterns can be found. and spot anomalies in patterns with low likelihood.Following the success of sparse approaches for representation and dictionary learning in a variety of computer vision problems, researchers in [13, 20] used sparse representation to learn. However, getting annotations for training is challenging and time-consuming, particularly for videos. Recently, [7] applied reconstruction loss to detect anomalies and deep learning-based autoecoders to learn the model of typical behaviors. Our method uses just sparsely labeled training data and takes into account both abnormal and normal behavior for anomaly detection.

## 5 Figures

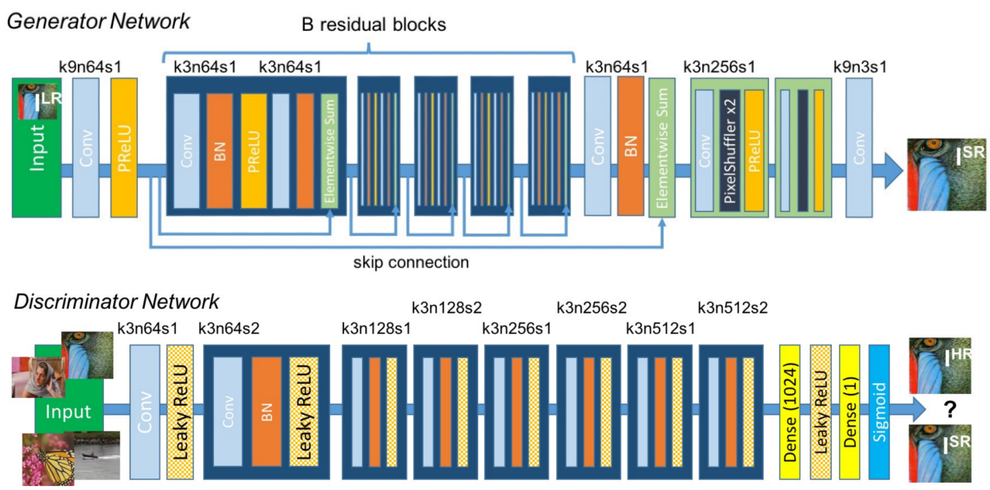


**Fig. 1.** Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

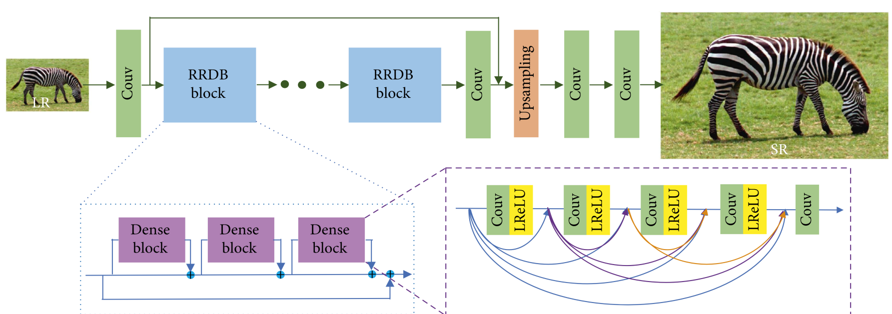


**Fig. 2.** SRGAN's Architecture.

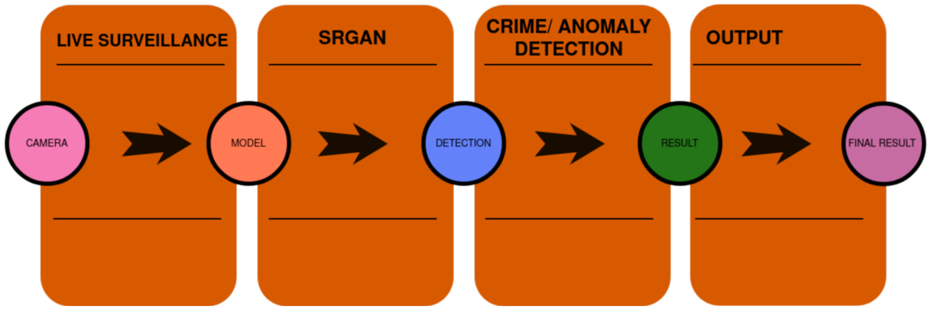


**Fig. 3.** System architecture and implementation details.

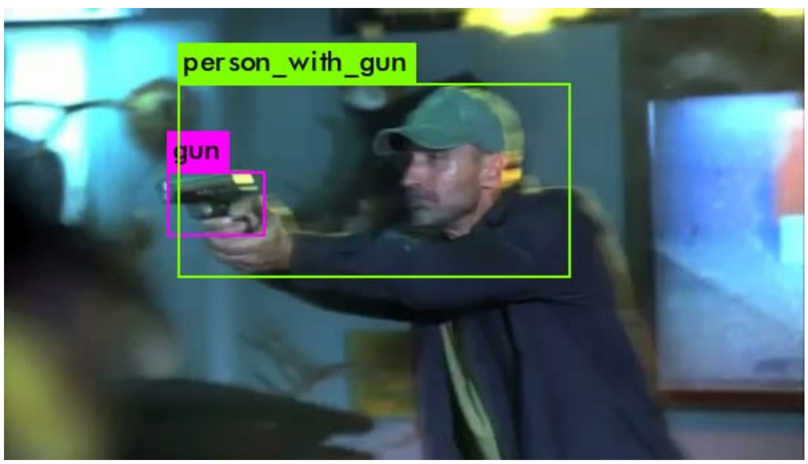


**Fig. 4.** Person with gun.


**Fig. 5.** Anomalies identified.

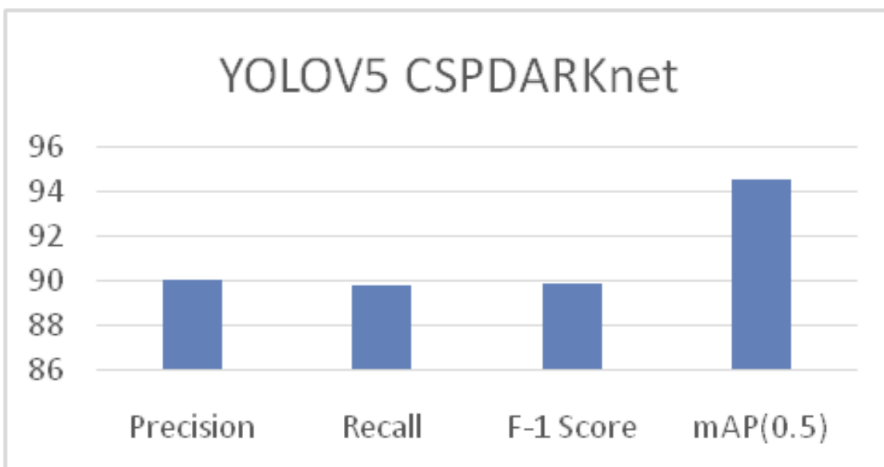


**Fig. 6.** YOLO V5 Performance Graph.

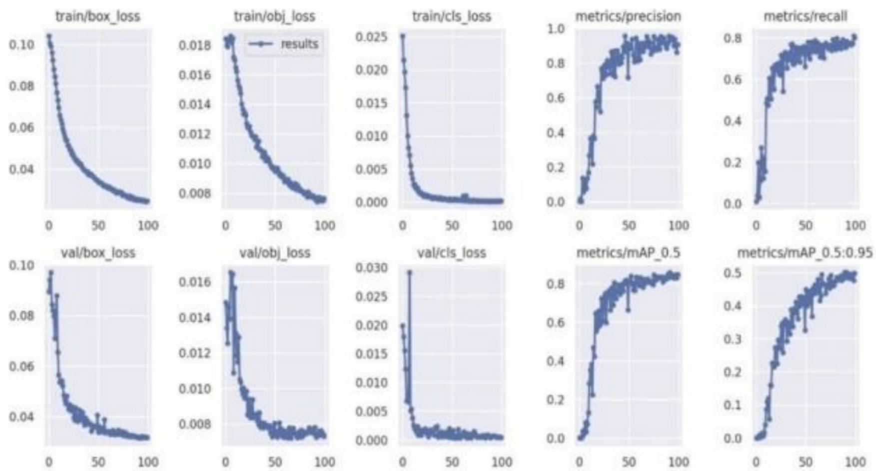


**Fig. 7.** YOLO V5 Training Loss.

# 6 Results

## 6.1 Video optimization

The image presented in section 5 (Fig. 2) shows the architecture of Super Resolution Adversarial Network and also the input and output results. We can see low resolution image image is up-scaled using the SRGAN algorithm. In our project we have used the SRGAN algorithm to optimize the quality of video. The video is converted into frames and each frame is pass through the SRGAN algorithm and later all up-scaled and enhanced image are combined to produce the final video output.

We then computed the Structural Similarity Index (SSIM) score and Peak signal-to-noise ratio (PSNR) for every pair of the original and matching upsampled frame.The average SSIM score is 0.678 and the average PSNR score is 24.24. These scores are not very high, but they show that SRGAN can be effectively used as a video up-sampler by collating the generated high-quality frames if trained correctly.

## 6.2 Anomalies identification

Three thousand images and 100 epochs were used to train the YOLO V5 model. Due to the Training model, two hours were spent. When we evaluate our model using 600 photos, of which 416 are categorized as anomalous(The frame contains guns; otherwise, our model is normal.) quickly determine if an anomaly is present or not. By lowering the number of categories and classes over our model to reduce the number of anomalies to just two whether or not we accelerated and refined our model to only identify anomalies. Because of After training and validation, we discovered YOLO V5. model produces good results. Following assessment of our model possessed an 89.8 % recall rate, a 90.1 % precision score, and 94.6 % is the mAP score.

Refer Fig. 6 & Fig. 7, In Fig. 7, The YOLOV5 model's box loss, objectness loss, and classification loss are represented in the first three columns. The leftmost row is for training, and the second row is for validation. These columns show how well the algorithm predicts the object. These demonstrate the accurate recognition of classes used during the training process, such as knives, guns, and billets. The model works well in an open environment, is appropriate for detecting guns and knives, and can be used to identify anomalies when a weapon is detected in the frame.

## 7 Conclusion

For improving surveillance and law enforcement, generative AI has showed tremendous potential when used for real-time CCTV footage upscaling and crime scene identification. This cutting-edge system enhances low-resolution CCTV footage using deep learning algorithms, producing crisper imagery for suspect identification and crime prevention. Human operators' workloads are lightened by generative AI, freeing them up to concentrate on important tasks. Rapid event analysis is aided by the crime scene identification module, speeding up investigations. For future advancements, it is essential to develop AI models, algorithms, and dataset diversity in order to increase accuracy and reliability. Responsible deployment also requires addressing privacy laws and eliminating any biases. In conclusion, incorporating generative As technology develops, integrating AI into CCTV augmentation and crime detection has the potential to improve public safety and create safer neighborhoods.

## References


[1] B. Anti and B. Ommer. Video parsing for abnormality detection. In ICCV, 2011.

[2] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, Oliver Wang, Deep Video Deblurring, IEEE International Conference on Computer Vision and Pattern Recognition, Pages 237 – 246, 2017.

[3] Basharat, A. Gritai, and M. Shah. Learning object motion patterns for anomaly detection and improved object detection. In CVPR, 2008.

[4] X. Cui, Q. Liu, M. Gao, and D. N. Metaxas. Abnormal detection using interaction energy potentials. In CVPR, 2011.

[5] A. Datta, M. Shah, and N. Da Vitoria Lobo. Person-on-person violence detection in video data. In ICPR, 2002.

[6] Y. Gao, H. Liu, X. Sun, C. Wang, and Y. Liu. Violence detection using oriented violent flows. Image and Vision Computing, 2016.

[7] Cheng, H.; Liu, X.; Wang, H.; Fang, Y.; Wang, M.; Zhao, X. SecureAD: A secure video anomaly detection framework on convolutional neural network in edge computing environment. IEEE Trans. Cloud Comput. 2020, 10, 1413–1427.

[8] T. Hospedales, S. Gong, and T. Xiang. A markov clustering topic model for mining behaviour in video. In ICCV, 2009.

[9] Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In CVPR, 2014.

[10] J. Kooij, M. Liem, J. Krijnders, T. Andringa, and D. Gavrila. Multimodal human aggression detection. Computer Vision and Image Understanding, 2016.

[11] L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In CVPR, 2009.

[12] W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes. TPAMI, 2014.

[13] C. Lu, J. Shi, and J. Jia. Abnormal event detection at 150 fps in matlab. In ICCV, 2013.

[14] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In CVPR, 2009.

[15] S. Mohammadi, A. Perina, H. Kiani, and M. Vittorio. Angry crowds: Detecting violent events in videos. In ECCV, 2016.

[16] I. Saleemi, K. Shafique, and M. Shah. Probabilistic modeling of scene dynamics for applications in visual surveillance. TPAMI, 31(8):1472–1485, 2009.

[17] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. In ICCV, 2015.

[18] S. Wu, B. E. Moore, and M. Shah. Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In CVPR, 2010.

[19] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe. Learning deep representations of appearance and motion for anomalous event detection. In BMVC, 2015.

[20] B. Zhao, L. Fei-Fei, and E. P. Xing. Online detection of unusual events in videos via dynamic sparse coding. In CVPR, 2011.

[21] Y. Zhu, I. M. Nayak, and A. K. Roy-Chowdhury. Context aware activity recognition and anomaly detection in video. In IEEE Journal of Selected Topics in Signal Processing, 2013.

[22] Deepika Ajalkar, Anil Kumar, Ashish Sharma, Deepak A Vidhate, A Deepak, TR Vijaya Lakshmi. An Smart Intelligence Performance Analysis Using ANN Classifiers For Soil Color Texture Identification in International Journal of Intelligent Systems and Applications in Engineering, 2024.