IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Moving Object Detection in Improved Attention U-Net Enhanced HEVC Decoded Video

[1]G Sheeba, [2]M Maheswari

[1]Assistant Professor, [2]Professor

[1]Department of Electronics and Communication Engineering,

[1]Government College of Engineering Srirangam, Trichy, Tamil Nadu – 620012, India

Abstract: The consumer demand for video applications is rapidly increasing in the present digital market. This necessitates the compression of the video streams and the HEVC (High Efficiency Video Coding) standard significantly reduces network traffic and bandwidth needs. The coarse quantization that results in many artifacts is the main cause of the compressed video's quality loss. The goal of this work is to improve the quality of HEVC compressed decoded video using an Improved Attention U-Net architecture and to detect moving objects in the improved HEVC decoded video using a modified Mask R-CNN object detection approach. Deep learning based Denoising model can be used as a preprocessing technique in prior to Improved Attention U-Net GAN architecture. In the improved video, moving objects are found using the Mask R-CNN model. Using a post-processing technique like Non-Maximum Suppression (NMS), duplicate detections can be eliminated and the model's accuracy can be increased. The proposed method of quality enhancement demonstrated a maximum improvement of 7.71 dB for PSNR and a maximum improvement of 6.01% for SSIM, when compared to SRGAN (Super Resolution Generative Adversarial Network), MLDCNN (Multi Layered Deep Convolutional Neural Network) and Deep Learning with Super Interpolation – Laplacian Filter (DLSI-LF) methods. In comparison to Faster R-CNN and Single Shot multibox Detection (SSD) approaches, the proposed object detection method achieves a maximum precision of 0.9812, a maximum recall of 0.9789, a maximum F-measure of 0.9800 and a maximum accuracy of 0.9984 with a maximum improvement of 8.84%.

Index Terms – HEVC, Video Quality Enhancement, Moving Object Detection.

I. Introduction

The need for video is continuously rising in daily life, yet video resolution is rising and data storage is declining at the same time [1]. Video files are compressed while retaining quality with the help of video encoding. The High Efficiency Video Coding (HEVC) standard saved roughly 50% of the code rate with the same quality in comparison to the H.264 standard. Compression artifacts cause compressed videos to lose quality. In this circumstance, it is crucial to increase the quality of compressed videos [2]. One method to increase perceived video quality while maintaining a high compression rate is to filter the reconstructed frames to lessen the impact of various artifacts [3]. The standard of image and video quality is rising, as are consumer expectations. So, the quality of HEVC-compressed video must be improved [4]. HEVC offers more flexible intra- and inter-prediction coding than earlier standards, leading to better performance [5].

In order to remove redundant information, HEVC makes use of spatial and temporal correlations to increase coding efficiency at the encoder side [6]. The blurring, blocking, and ringing effects are just a few examples of the many abnormalities that emerge from the compressed video's quality loss. Certain restoration techniques can enhance the decoded video's quality, increasing the decoder's coding efficiency [7].

Convolutional Neural Networks (CNNs) have been used to successfully boost the visual quality of decoded frames. To enhance the quality of HEVC videos, several researchers have created sophisticated post processing filters for the HEVC decoder [8].

The issue of creating a High-Resolution (HR) frame from many Low-Resolution (LR) frames is referred to as Super Resolution reconstruction. One LR image must be used to rebuild the HR image, which is known as SISR (Single Image Super Resolution) [9]. Moreover, Generative Adversarial Networks (GANs) were applied to enhance the quality of the HEVC-compressed videos. As GANs are trained with adversarial loss, they can produce high-quality details [10]. A GAN's objective is to generate trustworthy images. SR-GAN (Super Resolution GAN), ESR-GAN (Enhanced SR GAN), and SR-UNET are deep learning methods for image improvement based on GAN.

Although GAN-based SR techniques deliver accurate results, GAN training is unstable. The U-Net architecture overcame these limitations to improve resolution. A U-Net network is a neural network built on an encoder-decoder architecture and intended for image segmentation and quality enhancement.

Detecting objects is intended to extract the classification and position data of a specific object from complicated scenarios. The two groups of deep learning object detection models are regression/classification-based approaches and region-proposal methods. Models for object identification that typically use regression include YOLO and SSD. The frame significantly increases detection speed in the regression-based method, but detection accuracy is still poor.

In Region-Convolution Neural Network (R-CNN), a region proposal technique, the pool layer of the Region of Interest (RoI) receives the feature map and its bounding box as inputs. The region proposal network, which offers end-to-end training, speeds up network training and improves the precision of regional extraction. The accuracy of the detection was increased by the addition of the Mask creation task. Mask R-CNN can learn rich features with the integrated Feature Pyramid Network (FPN) [11]. YOLO is slower than a Single Shot Detector (SSD) for several categories, which is equally accurate to a Faster R-CNN [12].

This research aims to utilize an Improved Attention U-Net architecture to enhance the quality of HEVC compressed video and utilize the Mask R-CNN for detecting moving objects in the enhanced HEVC decoded video.

II. RELATED WORK

Ma, D., Zhang et al. [13] suggested CVEGAN, unique GAN architecture, as a means of improving the quality of compressed video. Xiang, C. et al., [14] developed a new deep generative method for hiding video errors. Andrei, S.S et al. [15] presented SUPERVEGAN (Super Resolution Video Enhancement GAN) method for enhancing low bitrate streams' video while also eradicating compression artifacts from low bitrate streams. Wang, H., et al. [16] proposed GAN to derive realistic images for upscaling factors from super-resolution photographs. Li, Z., et al. [17] suggested the Residual-Attention UNet++ architecture to evaluate three different medical image datasets. Shalini, R. et al. [18] suggested an improved lightweight Attention Gate (AG)-equipped U-Net model to segment OD (Optical Disc) pictures. The segmentation accuracy of the neural network trained with the binary focus loss function was improved. Li, J et al. [19] proposed MF U-Net (Multi-scale Fusion U-Net) to address the problem of boundary pixel blurring and multi-scale variance in breast lesions. Feng, Z. et al. [20] suggested using a Stacked Reversed U-shape network (SRUNet) to extract various multi-scale characteristics by gradually improving up-and-down sampling of the feature maps. Wang, L. et al. [21] introduced a Context Based Prediction Enhancement (CBPE) model to increase coding effectiveness and reduce the energy of HEVC prediction residuals. Lin, J., et al. [22] suggested a CNN-based approach for block-level down/up-sampling-based video coding. Li, T. et al. [23] devised a technique for streamlining intra-mode HEVC, in which a deep convolutional neural network (CNN) model is trained to predict CTU partition rather than RDO. Zhang, G. et al. [24] presented a deep learning approach for CU partitioning in HEVC intracoding. Wu, M. et al. [25] suggested a cascade-trained Mask R-CNN-based object detection technique. Kiruthiga, G. et al [26] proposed a Convolutional Neural Network - Probabilistic Neural Network technique for identifying objects in security footage. Charouh, Z. et al. [27] proposed a CNN-based Deep Learning and Background Subtraction Moving Vehicle Detection Framework. Giraldo, J.H., et al [28] proposed a technique based on GCNN (Graph Convolutional Neural Network). L. Pang and K. Wong et al. [29] developed a technique for HEVC video that automatically detects moving objects. Chen, L. et al. [30] presented a quick object recognition technique for the HEVC intra compressed domain.

It is clear from the review of related works on improving the quality of HEVC decoded video, existing systems like Deep CNNs offer excellent accuracy, they also need more computational resources and require careful attention to prevent overfitting. The current approach, GAN, provides output of excellent quality, but it is difficult to train. U-Net architecture overcomes these shortcomings due to its high precision and low memory usage.

An Improved Attention U-Net architecture is used in the proposed system in order to enhance the quality of HEVC-decoded video and achieve better Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Method (SSIM) compared to existing methods like Super Resolution Generative Adversarial Network (SRGAN), Multi Layered Deep Convolutional Neural Networks (MLDCNN), and Deep Learning with Super Interpolation with Laplacian Filter (DLSI-LF) methods.

Object detection methods currently in use, such as R-CNN, Fast R-CNN, and Faster R-CNN, suffer from a number of limitations, including being computationally expensive, requiring a lot of memory, relying on region proposals. The Mask R-CNN approach fixes these issues. The Mask R-CNN method is used in the object detection of the enhanced HEVC decoded video in order to achieve better results in terms of precision, recall, F-measure and accuracy compared to current methods like Faster Region Convolutional Neural Networks (Faster R-CNN) and Single Shot multibox Detection (SSD) methods.

III. PROPOSED METHOD

The proposed framework for HEVC compressed video quality enhancement and moving object detection in enhanced HEVC decoded video is shown in Figure 1. The decoded HEVC video is passed through an Improved Attention U-Net architecture to improve the video quality, and an improved video output is produced. Prior to U-Net architecture as Generator in GAN for HEVC decoded video quality enhancement, advanced deep learning based denoising model is used as preprocessing method to improve the video quality. In the Improved Attention U-Net GAN enhanced HEVC decoded video, moving objects are detected using Mask R-CNN model. The Mask R-CNN approach can be altered in certain ways to enhance its performance. Duplicate detections can be eliminated and the model's accuracy can be increased by using post processing techniques like Non-Maximum Suppression (NMS).

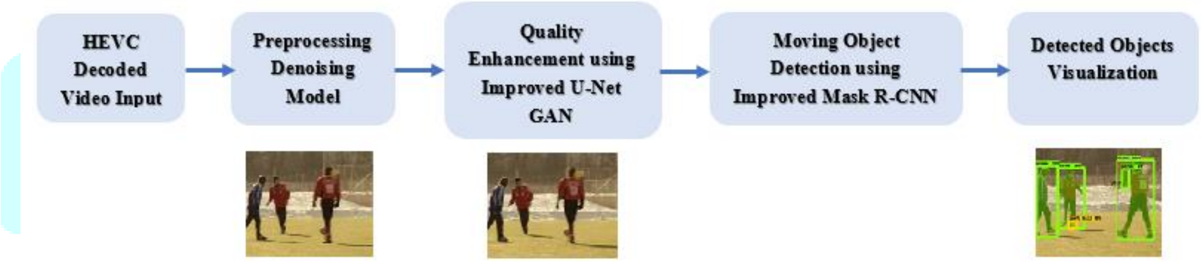

HEVC Decoded Video Input
Preprocessing Denoising Model
Quality Enhancement using Improved U-Net GAN
Moving Object Detection using Improved Mask R-CNN
Detected Objects Visualization

Figure 1. Proposed Methodology for Moving Object Detection in enhanced HEVC decoded video

3.1 Preprocessing Denoising Model

Video denoising methods can be used to reduce the amount of noise in the video frames. CNNs are ideal for preprocessing the videos, since they are built to automatically learn hierarchical data representations. The CNN architecture for denoising as shown in Figure 2. comprises convolutional layers, activation functions, pooling layers for downsampling, and deconvolutional layers for upsampling. Convolutional layers make up the first four layers of the algorithm and extract features from the input video frame; deconvolutional layers make up the next two layers and up-sample the denoised frame; and convolutional layers and activation layers make up the final two layers and produce the denoised output.

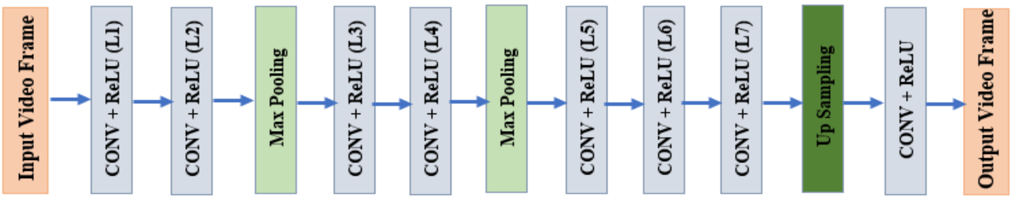


Figure 2. Deep learning based Denoising model

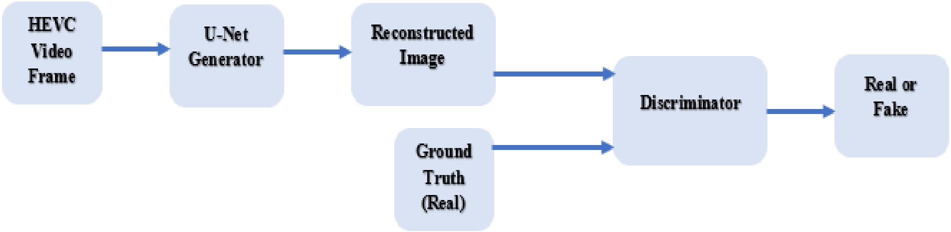


Figure 3. Block Diagram of U-Net GAN Architecture for Video Quality Enhancement

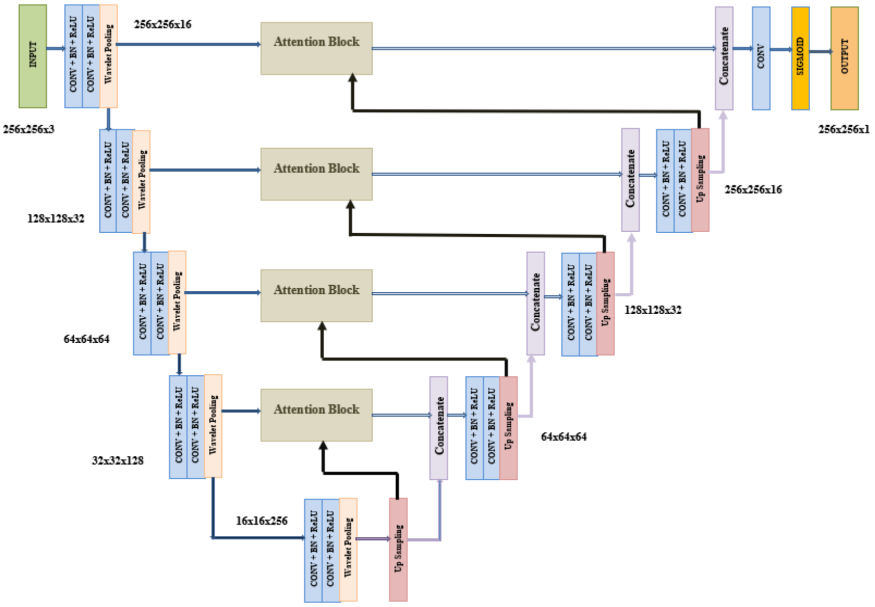


Figure 4. Improved Attention U-Net Architecture

3.2 Quality Enhancement using an Improved Attention U-Net GAN Architecture

The Generator and Discriminator are the two components that make up Generative Adversarial Networks (GANs). As seen in Figure 3, GANs have been applied to the elimination of compression artifacts from HEVC decoded video. The Improved Attention U-Net design is used to create the generator network in order to enhance the performance of the GAN, as illustrated in Figure 4. The HEVC decoded video is improved using the Improved Attention U-Net architecture. The contracting path, which has convolution layers, and the expanding path, which has up-convolution (deconvolution) layers, make up the Improved Attention U-Net. The skip connections between the expanding and contracting paths are another feature of the U-Net. The network receives the input image through the paths of contraction and expansion. Spatial information is decreased while feature information is boosted during the contraction. The decoder takes the opposite path, gradually raising the spatial resolution and compressing the channel dimension to change the encoded image. Channel-wise concatenation of features at the same depth level establishes a direct link between the encoder and the decoder.

The proposed network architecture consists of convolution layers, wavelet pooling levels, inverse wavelet-based up-sampling layers, and a sigmoid layer at the end. Each encoder and decoder block used two convolution blocks, which were 3x3 2D convolutional layers. Batch Normalization and the activation function (ReLU) are used for the encoder and decoder units along each layer of the convolutional network. To complete the merging of information, Attention Blocks are used. To scale the output of each of these networks, the activation function was sigmoid. The network was made deeper by increasing the number of layers from 16 to 256. Wavelet pooling can reduce the size of the feature map by using wavelets. The image is compressed using the Haar wavelet pooling technique as it requires fewest calculations.

Attention Block is implemented as depicted in Figure 5. The attention coefficients obtained in the Attention Block were used to scale low-level features on the encoder path. The expansion path's up sampling features and the encoder path's low-level features are the two inputs that the attention block accepts. Both inputs are convolved and batch normalized before being combined. The first activation function, ReLU, is applied to the additional signal before Convolution and Batch Normalization. The second activation function, Sigmoid and Resample, is then applied to the resulting signal to produce the attention coefficient ‘a’. The output is then produced by multiplying the encoder feature by the coefficient, pixel per pixel. In the suggested architecture, this attention block is used in between the decoder and encoder blocks. The Discriminator is another deep network that learns to distinguish between actual and fake images, and its structure is depicted in Figure 6. The Discriminator consists of convolutional layers with a 3x3 stride 1 kernel. The number of channels in these convolution layers increases linearly from 16 to 256 as we go deeper into the architecture. Batch Normalization layers with leaky ReLU activations are present for each convolution layer. After the last convolution layer, a Sigmoid activation function is used to create the final improved image.

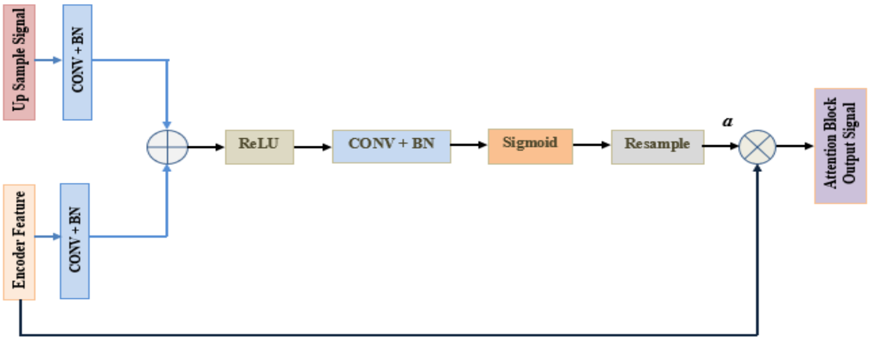


Figure 5. Implementation of Attention Block

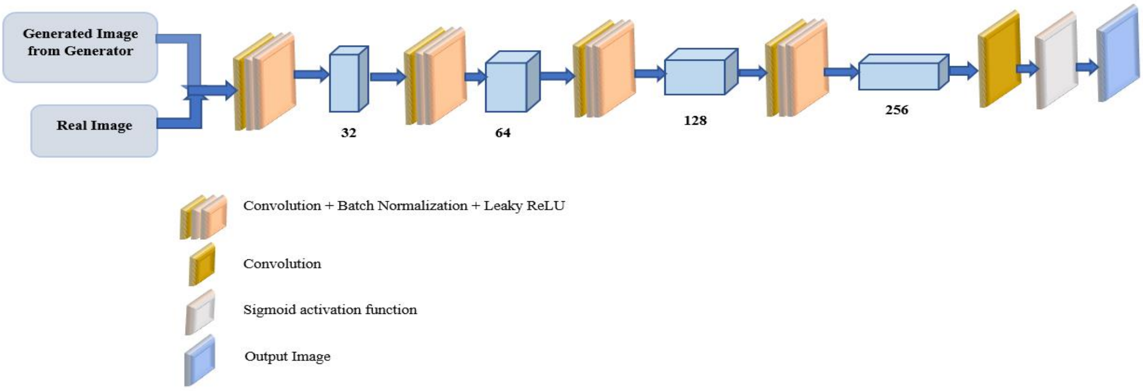

Generated Image from Generator
Real Image
32
64
128
256
Convolution + Batch Normalization + Leaky ReLU
Convolution
Sigmoid activation function
Output Image

Figure 6. Structure of Discriminator

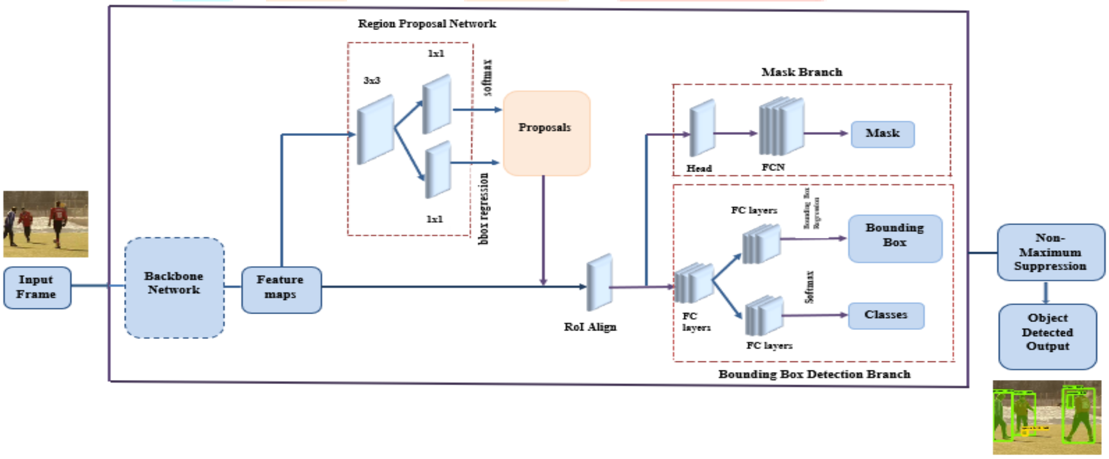


Figure 7. Moving Object Detection using Mask R-CNN

3.3 Moving Object Detection using Mask R-CNN

Mask R-CNN is a deep learning model for tasks like object detection and segmentation. Non-Maximum Suppression (NMS) is utilized as a post-processing technique to reduce duplicate detections and increase accuracy in order to enhance the performance of the Mask R-CNN method. The block diagram for Moving Object Detection using Mask R-CNN is shown in Figure 7. The input image first passes through the MobileNet backbone, which is made up of several convolutional layers and extracts features at various scales. The Feature Pyramid Network (FPN), which generates a variety of feature maps at various scales, receives the output feature maps from the MobileNet backbone. These feature maps are used by both the mask branch and the Region Proposal Network (RPN) to anticipate instance masks and propose object regions. The RPN proposes probable object regions in the image using the feature maps produced by the FPN. The RoI align layer uses the RPN's suggested regions to extract features from the FPN feature maps. It creates a fixed-size feature map for each proposal and then utilizes bilinear interpolation to align the features with the proposal boundaries. Fully Connected (FC) layers receive the characteristics from the RoI align layer. Moreover, the RoI align features provide input to a branch called mask that builds an instance mask for each proposal. The feature map's resolution is continuously increased using a sequence of convolutional and up-sampling layers in the mask branch, which are then triggered with a sigmoid function to produce a mask probability map.

3.4 Post Processing of Mask R-CNN object detection

In order to maintain only the most certain and non-overlapping detections, Non-Maximum Suppression (NMS), a post processing approach, is utilized in object detection models. Low-confidence predictions are filtered out using a defined confidence threshold. Bounding box predictions that fall short of the threshold in terms of confidence score are ignored. The remaining bounding box predictions are arranged in descending order according to their confidence scores.

Table 1. Specifications of Dataset

Video Sequence	Resolution	Number of Frames	Size
Container	386 x 288	150	44.55 MB
Husky	386 x 288	120	37.13 MB
Hall Monitor	386 x 288	150	44.55 MB
Soccer	386 x 288	150	44.55 MB
Ice	770 x 576	150	44.5 MB
Four People	1280 x 720	300	88.51 MB

IV. RESULTS AND DISCUSSION

The proposed work aims to detect moving objects in enhanced HEVC decoded video and the framework was verified on several test videos encoded with HM16.20 model. Low Delay P main configuration was selected and the quantization parameter of 37 was used. The remaining parameters are set to HM's default settings. The tests are performed on a computer running at 64-bit version of Windows 10 with a 2.10 GHz CPU and 8 GB of memory. The test video sequences were downloaded from https://media.xiph.org/video/derf/ to test the performance of the proposed model. The specifications of the test video sequences are as shown in Table 1 and some frames of these test video sequences are shown in Figure 8.

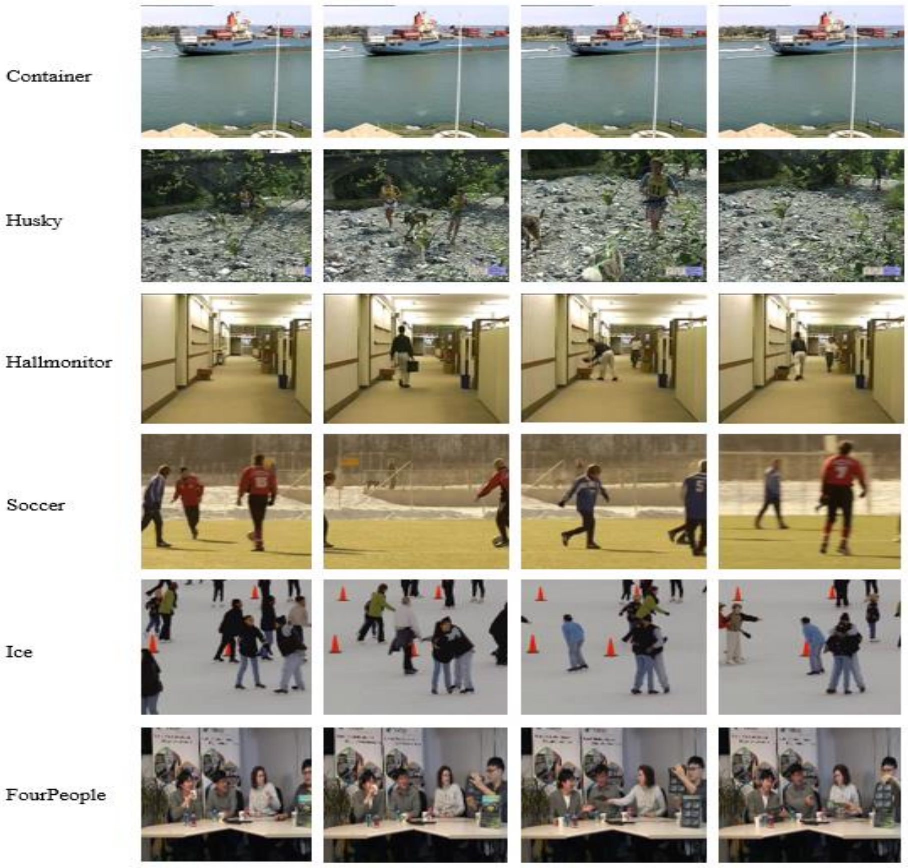


Figure 8. Sequence of Frames of Test Video Sequences

Performance metrics used to evaluate video quality enhancement include the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Method (SSIM). With rising PSNR and SSIM values, the similarity between the reconstructed images and the original images grows. The most used estimate of image quality measuring metric is Mean Square Error (MSE) and is represented by the equation (1).

$$MSE = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}[\,\hat{\mathrm{h}}(\mathrm{n},\mathrm{m}) - \mathrm{h}(\mathrm{n},\mathrm{m})\,]\textasciicircum 2 \qquad (1)$$

The PSNR is the most often used metric for assessing the effectiveness of reconstruction in lossy image compression codecs. The PSNR is represented by the equation (2)

$$PSNR = 20\log_{10} S - 10\log_{10} MSE \tag{2}$$

The maximum value in the image data is represented by S, which is 255 for the 8-bit unsigned integer data type. A quality evaluation criterion based on perception is the SSIM. By using SSIM, the perceived quality of pictures and videos is calculated. It calculates how similar the original and recovered images are to one another.

Several metrics are used by object detectors to assess their performance. Recall compares the corrected predictions to the actual predictions, whereas precision shows the percentage of correct predictions.

$$Precision = \frac{True\ Positive}{All\ Predictions} \tag{3}$$

$$Recall = \frac{True\ Positive}{All\ Ground\ Truth} \tag{4}$$

$$F - measure = 2\ x\ \frac{(Recall\ x\ Precision)}{(Recall + Precision)} \tag{5}$$

The simplest logical performance metric is accuracy, which is just a basic ratio of accurately predicted observations to all observations.

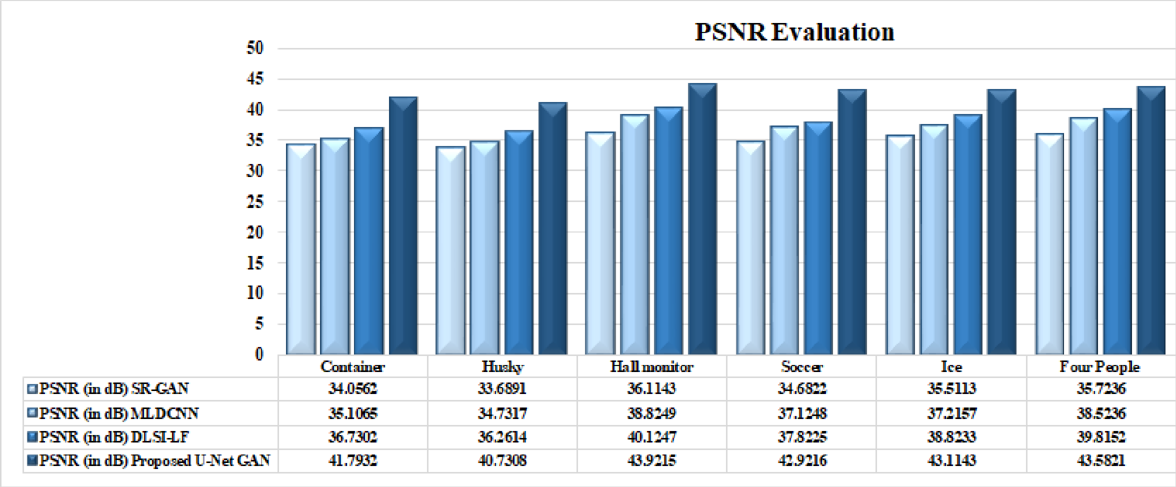
$$Accuracy = \frac{True\ Positive + True\ Negative}{All\ Observations} \tag{6}$$

Where

All Predictions = (True Positive + False Positive)
All Ground Truth = (True Positive + False Negative)
All Observations = (True Positive + False Positive + False Negative + True Negative)

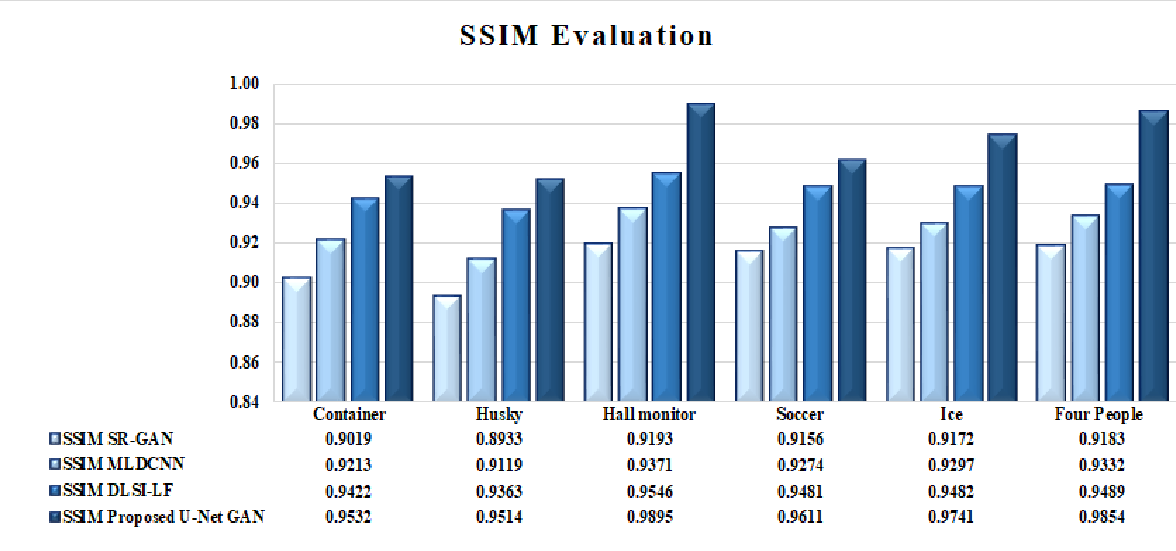
The HEVC-decoded video was enhanced using the Improved Attention U-Net GAN architecture. The PSNR and SSIM values of the proposed method were compared with those of existing methods such as SRGAN [31], MLDCNN [32], and DLSI-LF [33] approaches. The assessed outcomes are displayed in Figure 9 and 10. The results show that the suggested strategy for improving video quality reaches a maximum PSNR of 43.92 dB and a maximum SSIM value of 0.9895 In terms of PSNR evaluation, this shows a 7.71 dB improvement over the SRGAN method, a 5.76 dB improvement over the MLDCNN method, and a 4.41 dB over DLSI-LF method. In terms of SSIM evaluation, the proposed method shows an improvement of 6.01% compared to the SRGAN method, an improvement of 4.36% compared to the MLDCNN method and an improvement of 2.34% com-pared to the DLSI-LF method. Figure 11 displays the average performance evaluation of video quality enhancement in terms of PSNR and SSIM.

The improved HEVC decoded video was then applied to an object detection model using Mask R-CNN. The proposed approach is compared with other models such as Faster R-CNN [34] and SSD [35] models for various test video sequences in terms of precision, recall, F-measure and accuracy, as shown in Table 2. From the experimental findings, the suggested method of object detection attains a maximum precision of 0.9812 with an improvement of 7.07%, a maximum recall of 0.9789 with an improvement of 7.05%, a maximum F-measure of 0.9800 with an improvement of 7.04%, and a maximum accuracy of 0.9984 with an improvement of 8.84% when compared to object detection method without the video quality enhancement. Figure 12 shows the overall performance evaluation of the object detection models. The results of moving object detection in enhanced HEVC decoded video are shown in Figure 13.



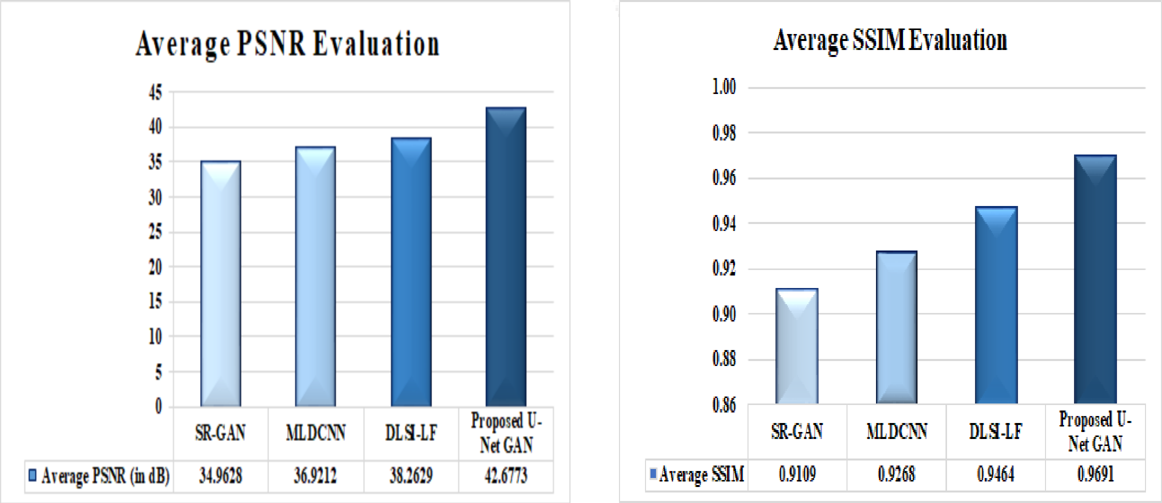
	Container	Husky	Hall monitor	Soccer	Ice	Four People
□PSNR (in dB) SR-GAN	34.0562	33.6891	36.1143	34.6822	35.5113	35.7236
□PSNR (in dB) MLDCNN	35.1065	34.7317	38.8249	37.1248	37.2157	38.5236
■PSNR (in dB) DLSI-LF	36.7302	36.2614	40.1247	37.8225	38.8233	39.8152
■PSNR (in dB) Proposed U-Net GAN	41.7932	40.7308	43.9215	42.9216	43.1143	43.5821

Figure 9. PSNR evaluation of Improved Attention U-Net architecture



	Container	Husky	Hall monitor	Soccer	Ice	Four People
□SSIM SR-GAN	0.9019	0.8933	0.9193	0.9156	0.9172	0.9183
□SSIM MLDCNN	0.9213	0.9119	0.9371	0.9274	0.9297	0.9332
■SSIM DLSI-LF	0.9422	0.9363	0.9546	0.9481	0.9482	0.9489
■SSIM Proposed U-Net GAN	0.9532	0.9514	0.9895	0.9611	0.9741	0.9854

Figure 10. SSIM evaluation of Improved Attention U-Net architecture



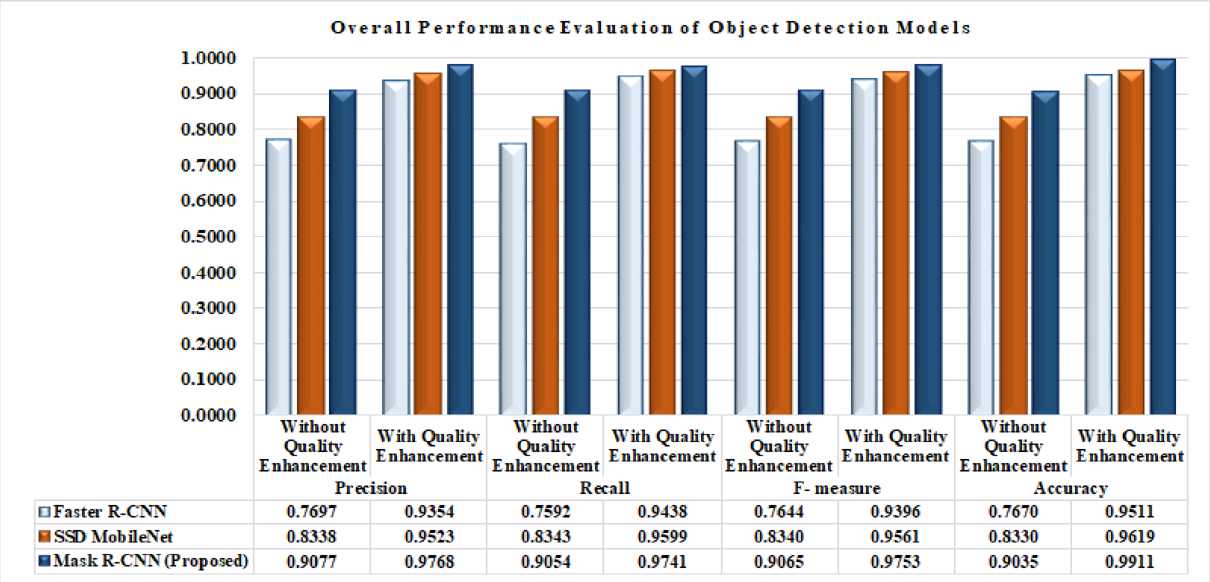
	SR-GAN	MLDCNN	DLSI-LF	Proposed U-Net GAN
□ Average PSNR (in dB)	34.9628	36.9212	38.2629	42.6773

	SR-GAN	MLDCNN	DLSI-LF	Proposed U-Net GAN
■ Average SSIM	0.9109	0.9268	0.9464	0.9691

Figure 11. Average PSNR and SSIM Evaluation for Video Quality Enhancement

Table 2. Performance Comparison of Object Detection Models

Object Detection Model	Video Sequence	Object detection without Quality Enhancement				Object Detection with Improved U-Net GAN Quality Enhancement			
		Precision	Recall	F-Measure	Accuracy	Precision	Recall	F-Measure	Accuracy
Faster R-CNN	Container	0.7638	0.7549	0.7593	0.7627	0.9308	0.9471	0.9389	0.9588
	Husky	0.7631	0.7465	0.7547	0.7543	0.9219	0.9382	0.9300	0.9499
	Hall Monitor	0.7766	0.7708	0.7737	0.7786	0.9387	0.9423	0.9405	0.9487
	Soccer	0.7687	0.7568	0.7627	0.7646	0.9386	0.9443	0.9414	0.9467
	Ice	0.7726	0.7586	0.7655	0.7664	0.9429	0.9462	0.9445	0.9469
	Four People	0.7733	0.7678	0.7705	0.7756	0.9395	0.9449	0.9422	0.9554
SSD MobileNet	Container	0.8286	0.8295	0.8290	0.8285	0.9542	0.9714	0.9627	0.9695
	Husky	0.8269	0.8294	0.8281	0.8277	0.9453	0.9625	0.9538	0.9606
	Hall Monitor	0.8388	0.8375	0.8381	0.8362	0.9521	0.9547	0.9534	0.9612
	Soccer	0.8348	0.8352	0.8350	0.8341	0.9585	0.9538	0.9561	0.9674
	Ice	0.8366	0.8365	0.8365	0.8358	0.9533	0.9532	0.9532	0.9504
	Four People	0.8373	0.8374	0.8373	0.8358	0.9506	0.9637	0.9571	0.9624
Mask R-CNN (Proposed)	Container	0.8989	0.9014	0.9001	0.9015	0.9745	0.9707	0.9742	0.9883
	Husky	0.8955	0.8985	0.8970	0.8990	0.9723	0.9698	0.9711	0.9835
	Hall Monitor	0.9138	0.9128	0.9133	0.9092	0.9812	0.9789	0.9800	0.9984
	Soccer	0.9114	0.9038	0.9076	0.9027	0.9763	0.9738	0.9749	0.9895
	Ice	0.9129	0.9050	0.9089	0.9037	0.9773	0.9754	0.9752	0.9933
	Four People	0.9136	0.9109	0.9122	0.9046	0.9792	0.9759	0.9761	0.9934



	Precision		Recall		F- measure		Accuracy	
	Without Quality Enhancement	With Quality Enhancement	Without Quality Enhancement	With Quality Enhancement	Without Quality Enhancement	With Quality Enhancement	Without Quality Enhancement	With Quality Enhancement
□ Faster R-CNN	0.7697	0.9354	0.7592	0.9438	0.7644	0.9396	0.7670	0.9511
■ SSD MobileNet	0.8338	0.9523	0.8343	0.9599	0.8340	0.9561	0.8330	0.9619
■ Mask R-CNN (Proposed)	0.9077	0.9768	0.9054	0.9741	0.9065	0.9753	0.9035	0.9911

Figure 12. Overall Performance Evaluation of Object Detection Models

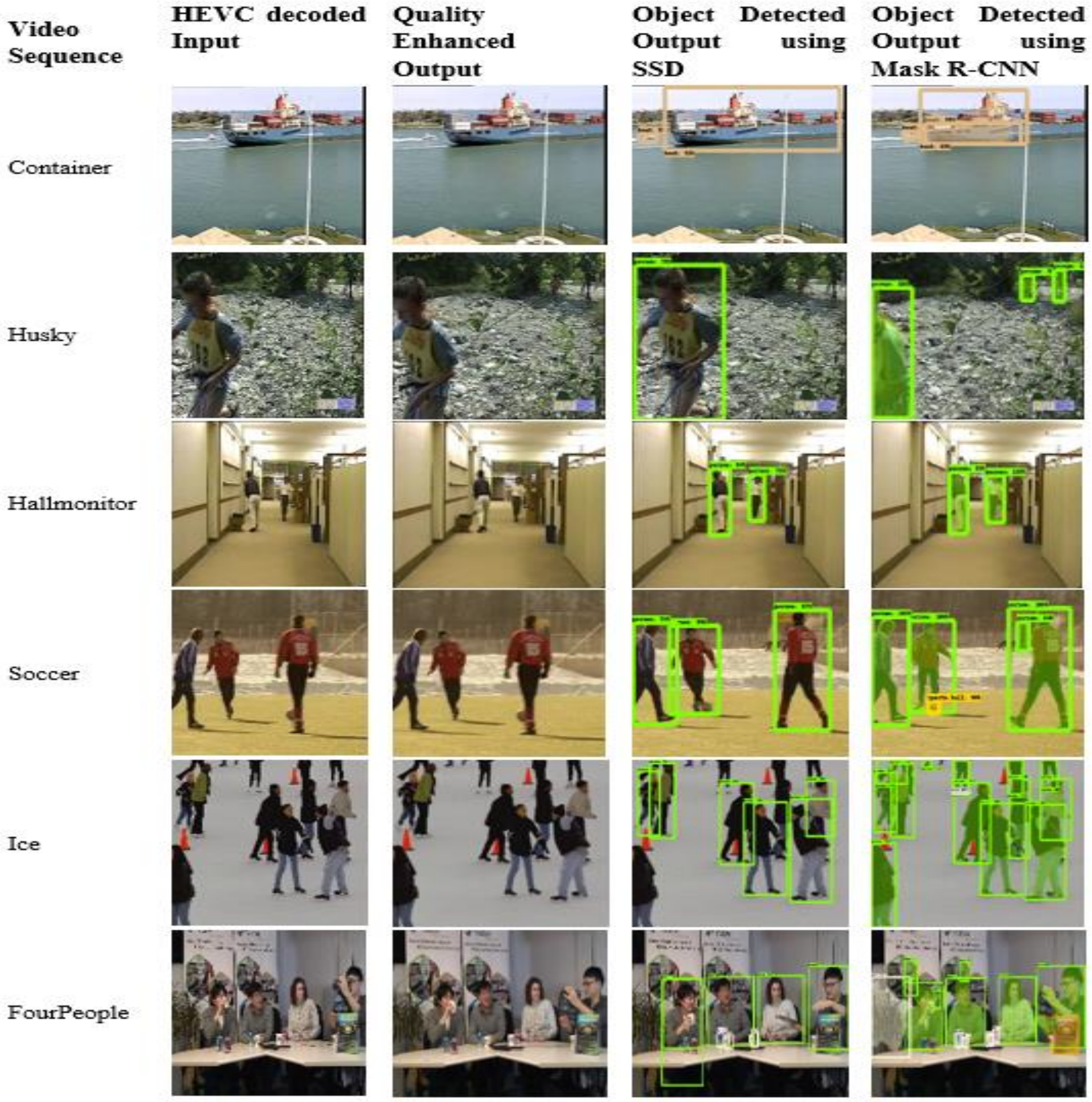


Figure 13. Outcomes of Moving Object Detection in Enhanced HEVC Decoded Video

V. CONCLUSION

In this work, an Improved Attention U-Net architecture is used to improve the quality of HEVC decoded video, and further moving objects are detected in the enhanced HEVC decoded video using the Mask R-CNN method. Comparing the proposed approach of quality enhancement to SRGAN, MLDCNN (and DLSI-LF methods, it showed a maximum improvement of 7.71 dB in terms of PSNR and a maximum improvement of 6.01% in terms of SSIM. When compared to object detection without quality enhancement, the suggested object detection method achieves maximum precision of 0.9812, a maximum recall of 0.9789, a maximum F-measure of 0.9800, and a maximum accuracy of 0.9984 with an improvement of 8.84%. Thus, the proposed Mask R-CNN method of object detection achieved the best detection accuracy. In future work, an option to choose a moving object detection method based on application requirements will be provided, so that when accuracy is important, the Mask R-CNN method will be used, and when detection speed is required, the SSD method will be used.

REFERENCES

[1] Huang, H., Schiopu, I. and Munteanu, A., 2020, September. Low-Complexity Angular Intra-Prediction Convolutional Neural Network for Lossless HEVC. In 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP) (pp. 1-6). IEEE.

[2] Zhang, S., Herranz, L., Mrak, M., Blanch, M.G., Wan, S. and Yang, F., 2022, May. DCNGAN: A Deformable Convolution-Based GAN with QP Adaptation for Perceptual Quality Enhancement of Compressed Video. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2035-2039). IEEE.

[3] Galteri, L., Seidenari, L., Bertini, M., Uricchio, T. and Del Bimbo, A., 2019, October. Fast video quality enhancement using gans. In Proceedings of the 27th ACM international conference on multimedia (pp. 1065-1067).

[4] Wang, T., He, J., Xiong, S., Karn, P. and He, X., 2020, August. Visual perception enhancement for HEVC compressed video using a generative adversarial network. In 2020 International Conference on UK-China Emerging Technologies (UCET) (pp. 1-4). IEEE.

[5] Wang, Z. and Li, F., 2021. Convolutional neural network based low complexity HEVC intra encoder. Multimedia Tools and Applications, 80(2), pp.2441-2460.

[6] Hsu, T.Y., Lu, Y.J., Hsieh, T.H. and Wang, C.C., 2021, November. An Efficient HEVC Intra Frame Coding Based on Deep Convolutional Neural Network. In 2021 IEEE/ACIS 22nd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 138-141). IEEE.

[7] Wang, T., Chen, M. and Chao, H., 2017, April. A novel deep learning-based method of improving coding efficiency from the decoder-end for HEVC. In 2017 data compression conference (DCC) (pp. 410-419). IEEE.

[8] Yang, R., Xu, M., Liu, T., Wang, Z. and Guan, Z., 2018. Enhancing quality for HEVC compressed videos. IEEE Transactions on Circuits and Systems for Video Technology, 29(7), pp.2039-2054.

[9] Tej, A.R., Halder, S.S., Shandeelya, A.P. and Pankajakshan, V., 2020, July. Enhancing perceptual loss with adversarial feature matching for super-resolution. In 2020 International joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.

[10] Zhang, S., Herranz, L., Mrak, M., Blanch, M.G., Wan, S. and Yang, F., 2022. PeQuENet: Perceptual Quality Enhancement of Compressed Video with Adaptation-and Attention-based Network. arXiv preprint arXiv:2206.07893.

[11] Wenjie, J. and Xiaoshu, L., 2019, November. Research on super-resolution reconstruction algorithm of remote sensing image based on generative adversarial networks. In 2019 IEEE 2nd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE) (pp. 438-441). IEEE.

[12] MathuraBai, B., Maddali, V.P., Devineni, C., Bhukya, I. and Bandari, S., 2022. Object Detetcion using SSD-MobileNet. International Research Journal of Engineering and Technology, 9, pp.2668-2771.

[13] Ma, D., Zhang, F. and Bull, D.R., 2024. CVEGAN: a perceptually-inspired gan for compressed video enhancement. Signal Processing: Image Communication, p.117127.

[14] Xiang, C., Xu, J., Yan, C., Peng, Q. and Wu, X., 2019, May. Generative adversarial networks based error concealment for low resolution video. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1827-1831). IEEE.

[15] Andrei, S.S., Shapovalova, N. and Mayol-Cuevas, W., 2021. SUPERVEGAN: Super resolution video enhancement GAN for perceptually improving low bitrate streams. IEEE Access, 9, pp.91160-91174.

[16] Wang, H., Wu, W., Su, Y., Duan, Y. and Wang, P., 2019, July. Image super-resolution using a improved generative adversarial network. In 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC) (pp. 312-315). IEEE.

[17] Li, Z., Zhang, H., Li, Z. and Ren, Z., 2022. Residual-attention UNet++: a nested residual-attention U-net for medical image segmentation. Applied Sciences, 12(14), p.7149.

[18] Shalini, R. and Gopi, V.P., 2022. Deep learning approaches based improved light weight U-Net with attention module for optic disc segmentation. Physical and Engineering Sciences in Medicine, 45(4), pp.1111-1122.

[19] Li, J., Cheng, L., Xia, T., Ni, H. and Li, J., 2021. Multi-scale fusion U-net for the segmentation of breast lesions. IEEE Access, 9, pp.137125-137139.

[20] Feng, Z., Lee, F. and Chen, Q., 2022. SRUNet: stacked reversed U-shape network for lightweight single image super-resolution. IEEE Access, 10, pp.60151-60162.

[21] Wang, L., Fiandrotti, A., Purica, A., Valenzise, G. and Cagnazzo, M., 2019, May. Enhancing HEVC spatial prediction by context-based learning. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4035-4039). IEEE.

[22] Lin, J., Liu, D., Yang, H., Li, H. and Wu, F., 2018. Convolutional neural network-based block up-sampling for HEVC. IEEE Transactions on Circuits and Systems for Video Technology, 29(12), pp.3701-3715.

[23] Li, T., Xu, M. and Deng, X., 2017, July. A deep convolutional neural network approach for complexity reduction on intra-mode HEVC. In 2017 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1255-1260). IEEE.

[24] Zhang, G., Xiong, L., Lian, X. and Zhou, W., 2019, August. A CNN-based coding unit partition in HEVC for video processing. In 2019 IEEE International Conference on Real-time Computing and Robotics (RCAR) (pp. 273-276). IEEE.

[25] Wu, M., Yue, H., Wang, J., Huang, Y., Liu, M., Jiang, Y., Ke, C. and Zeng, C., 2020. Object detection based on RGC mask R-CNN. IET Image Processing, 14(8), pp.1502-1508.

[26] Kiruthiga, G. and Yuvaraj, N., 2021. Improved object detection in video surveillance using deep convolutional neural network learning. International Journal for Modern Trends in Science and Technology, 7(11), pp.108-114.

[27] Charouh, Z., Ezzouhri, A., Ghogho, M. and Guennoun, Z., 2022. A resource-efficient CNN-based method for moving vehicle detection. Sensors, 22(3), p.1193.

[28] Giraldo, J.H., Javed, S., Werghi, N. and Bouwmans, T., 2021. Graph CNN for moving object detection in complex environments from unseen videos. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 225-233).

[29] Pang, L. and Wong, K., 2021, December. Moving Object Detection in HEVC Video. In 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) (pp. 1416-1421). IEEE.

[30] Chen, L., Sun, H., Katto, J., Zeng, X. and Fan, Y., 2021, August. Fast object detection in hevc intra compressed domain. In 2021 29th European Signal Processing Conference (EUSIPCO) (pp. 756-760). IEEE.

[31] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

[32] Kuanar, S., Conly, C. and Rao, K.R., 2018, June. Deep learning based HEVC in-loop filtering for decoder quality enhancement. In 2018 Picture Coding Symposium (PCS) (pp. 164-168). IEEE.

[33] Sheeba, G. and Maheswari, M., 2023. HEVC video quality enhancement using deep learning with super interpolation and laplacian filter. IETE Journal of Research, 69(11), pp.7979-7992.

[34] Gavrilescu, R., Zet, C., Foşalău, C., Skoczylas, M. and Cotovanu, D., 2018, October. Faster R-CNN: an approach to real-time object detection. In 2018 International Conference and Exposition on Electrical And Power Engineering (EPE) (pp. 0165-0168). IEEE.

[35] Saji, R.M. and Sobhana, N.V., 2021, February. Real Time Object Detection Using SSD For Bank Security. In IOP Conference Series: Materials Science and Engineering (Vol. 1070, No. 1, p. 012060). IOP Publishing.