Improved Intra Prediction Algorithm for HEVC with Conventional and Convolutional Neural Network Approach

[1]Swamy T N, [2]Ramesha K
[1.] Department of Electronics and Communication Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, India
[2.] Department of Electronics and Instrumentation Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, India
swamyohm@gmail.com

Article Info
Page Number: 272– 280
Publication Issue:
Vol 71 No. 2 (2022)

Abstract
Video compression has evolved since last decade and has extended its employability to many applications such as high definition televisions, video conferencing etc. High definition videos have nearly two times of bit rate than standard definition videos. Hence, video compression is an important approach to obtain information with lesser data size. The paper proposes an improved approach for intra-prediction in HEVC with modified resolution enhancement approach and intra prediction is considered for all 35 modes. The obtained results shows that the similarity between input and output is nearly same and PSNR average is around 42. Further, a deep learning approach has also been incorporated and coding unit depth is predicted for intra prediction. The developed model has provided 85% accuracy maintaining the similar PSNR.

Article History
***Article Received**: 30 December 2021*
***Revised**: 30 January 2022*
***Accepted**: 15 March 2022*
***Publication**: 09 April 2022*

Keywords—Intraprediction, Conventional, Convolution, Neural Networks, Deep Learning, PSNR, MSE.

I. INTRODUCTION

Video based applications exploits about 70% of the internet traffic and popularization of the digital TV/HDTV broad-casting, multimedia services and also video conferencing have raised the need of an efficient video compression technique. The primary goal of a compression technique is to reduce the data requirement, maintaining the same quality of video. The video coding standards have evolved from H.261 to H.265. The Standardization of algorithms on video coding holds the greater expectation of large markets for video communication equipment that increased commercial interest in development of an image and video coding standardization. Currently Advanced Video Coding (AVC) or H.264 is widely utilized video coding technique [1]. High Efficiency Video Coding (HEVC) has proved its significance by providing video quality same as H.264/AVC at 50% lesser bit rate. HEVC standard has flexible structure comprising three units namely Prediction Unit (PU), Coding Unit (CU), and Transform Unit (TU). The HEVC standard considers adaptive Coding Tree Unit (CTU) which has non-overlapping CU, and also PU for the flexible coding segmentation. For the purpose of intra coding, the CU assist two PU partitioning types, namely 2N*2N and N*N. HEVC is also an ideal choice for efficient video scaling applications. The Paper is organized as fallows. Section I consists of Introduction, Section II gives information on existing works, Section III describes intraprediction in HEVC, Section IV explains proposed intraprediction, Section V discuss about the results obtained and also gives analysis and finally VI sections gives about the overall work conclusions.

II. LITERATURE SURVEY

Carlos Silva Cardenas et. al., [3] have observed that Fractional Motion Estimation (FME) is a major part of CODEC since it remains active for nearly 40% to 60% of the total processing time. They have also claimed that Parallel Hardware Architecture employing the Sum of Absolute Differences (SAD) to find good matching of pixels considering minute interpolated pixels which improves the performance of the design. Anass Mansouri et. al., [4] have given the intra-prediction hardware architecture in HEVC. Their approach has attained high speed by adopting parallelism in architecture and pipeline method in processing. Satoshi Goto et. al., [5] gives Row-Based-Miss Information-Compression architecture to reduce the FIFOs size and Conflict Check methods to check pipeline risks.

Y Chen et. al., [6] has informed that the intra prediction phenomenon is reached considering 35 modes with higher CTU. The CTU opted using Rate Distortion Optimization (RDO) and Rough Mode Decision (RMD) mode which is a time consuming method. Along with the compression of video, it is also very important to maintain the resolution of the video. Any degradation of the resolution leads to loss of the information and also brings down the performance of the application. Hence, the proposed intra-prediction approach also focuses on the methods of enhancing and maintaining image resolution.

Pengfei et al., [7] defines image acquisition model by means of lower bit depth scalar quantizer arrays in order to rebuild images with larger bit depths using inter-pixel correlations in image local regions and improved image quality has been attained. Lizhi Wang et al., [8] have proposed a faster video acquisition method based on Nyquist criteria and also on sampling algorithm compression. The hyperspectral video obtained by integration of overall required frequency band with the help of Nyquist sampling, built at temporal domain at lower frame rate. Alvaro et. al.,[9] have provided a screen resolution scheme for satellite images for both line spreading function and full width at a half maximum for spatial resolution matrix. Liang Peng et. al., [10] gives the work that is carried out to improve the image acquisition capacity using the Complex Programmable Logic Device, the CCD Camera, the Decoding chip and memory.

Daisuke Sugimura et. al., [11] give solutions to the noise and blurring concepts in image acquisition models. H. Huang et. al., [11] provides an improved performance HEVC algorithm using Convolution Neural Networks (CNN) technique. Here 9 angular modes uses CNN method for prediction with an improvement in average bit rate. Schiopu H. Huang et. al., [12] presents a deep neural model using multi-resolution design using CNN which provides improvement of 5% in average bitrate compared to HEVC standard. V. Sanchez et. al., [13] presents Hybrid Lossless intra prediction technique using Neural Networks to improve the prediction accuracy by 0.56 dB and 7.01 dB Peak Signal to Noise Ratio (PSNR). Mai Xu et. al., [14] present earlier terminated hierarchy based method of intraprediction to predict CU partition to reduce complexity in HEVC. Raz Birman et. al., [15] gives deep neural networks method in order to consider prediction and results has great reduction in mean square error. Thorsten et. al., [16] has given the encoder control using deep learning method to reduce the loss in efficiency of coding by 0.52% compared to standard method. Damian Ruiz et. al., [17] provides fast partition architecture used for detection using texture orientation algorithms providing better time saving of 67% compared to standard format of HEVC.

III. INTRAPREDICTION IN HEVC

Major steps involved in Intra Prediction Process includes reference samples array construction, sample prediction and final samples processing. HEVC intraprediction classification consist of Discrete Cosine (DC), Planar and Angular Prediction. DC prediction and planar prediction method are required to calculate the smoothness of the image. The structures and directional edges are predicted using Angular method.

A. *HEVC Coding Technique*

Quadtree based partitioning for HEVC as shown in Figure 1. This technique is more flexible, also efficiency of the coding depends on video resolution as well. It has 4 levels of coding namely, CTU, CU, PU, lastly TU. CTU converts larger frames to dimensions like 4 by 4, 16 by 16, 32 by 32, 64 by 64 as shown in Figure 2.

B. *DC Mode*

It is required in prediction of natural blocks. Probably it is helpful when the images have minute changes in the neighbor pixels values. It is achieved considering 1 and 2 equations.

$$\mathrm{Ppred}(\mathrm{a,b})\text{is the Predicted Pixel sample with}$$
$$(\mathrm{a,b}) = 0 \ldots\ldots\ldots\ldots . Nt - 1$$
$$\mathrm{Pref}(\mathrm{a,b})\text{is the neighboring Pixel sample with}$$
$$(\mathrm{a,b}) = 0 \ldots\ldots\ldots\ldots . Nt * 2 - 1$$

$$dc\_Value = (\sum_{a=0}^{Nt-1} P_{ref}(a,-1) + \sum_{x=0}^{Nt-1} P_{ref}(-1,b)) \gg (\log_2(N_t) + 1) \quad (1)$$

where Nt is the trasform block size

$$P_{pred}(0,0) = (P_{ref}(-1,0) + 2 * dc\_Value + P_{ref}(0,-1) + 2) \gg 2$$
$$P_{pred}(a,0) = (P_{ref}(a,-1) + 3 * dc\_Value + 2) \gg 2$$
$$P_{pred}(0,b) = (P_{ref}(-1,b) + 3 * \; dc\_Value + 2) \gg 2 \quad (2)$$

C. Planar Mode

It will remove any dis continuities while prediction. It is calculated by considering the horizontal and vertical linear prediction samples. The Calculation of Planar mode is shown in Equation 3.

$$P_{pred}(a,b) = \left((N_t - 1 - a) * P_{ref}(N_t,-1) + (N_t - 1 - b) * P_{ref}(a,-1) + b(b+1) * P_{ref}(-1,N_t) + N_t\right) \gg (\log_2(N_t) + 1) \quad (3)$$

D. Angular Prediction

High Frequency components and videos of complex texture using this method of prediction. It consists of 33 angular prediction modes and designed in different direction to provide lower complexity content. Equations 4 to 6 gives the calculations of angular prediction. Angular Intra Prediction for HEVC from mode 2 to mode 34 are shown in Figure 3.

$$dx = ((a+1) * Intrapredictedangle \gg 5 \quad (4)$$
$$iFa = ((a+1) * Intrapredictedangle \quad (5)$$
$$perdSamp\,[a][b] = \left((32 - iFa) * ref(b + idx + 1) + iFa * ref\{y + idx + 2\} + 16\right) \gg 5 \quad (6)$$

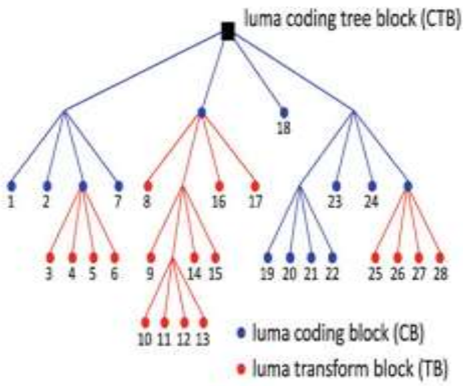

Fig. 1 The Quadtree Structure.

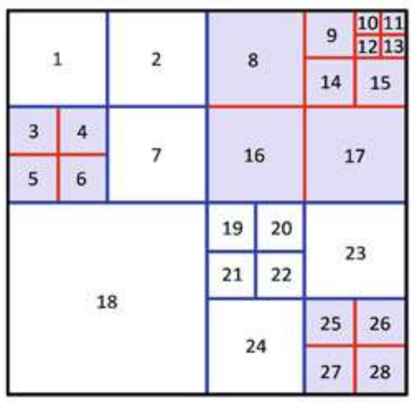

Fig. 2 64 x 64 Pixels Partition into CTU.

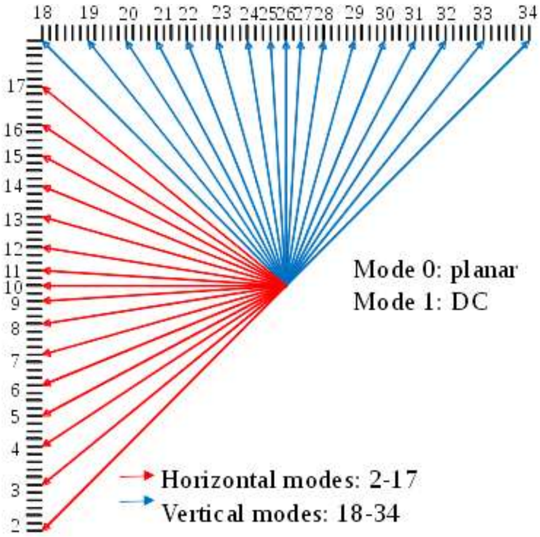

Fig. 3: HEVC Angular Intra Prediction from mode 2 to 34.

IV. PRAPOSED INTRAPREDICTION APPROACH FOR HEVC

In this paper two approaches have been discussed; one approach uses the conventional approach with improved image processing techniques and second approach comprises Convolutional Neural Network (CNN) approach with improved image processing techniques.

A. Improved Image Processing Techniques for Intra Prediction

The existing intra prediction approaches mainly focuses on improving the internal intra prediction flow. It has been extracted from the literature survey that improving the input quality of the intra prediction block improves the overall intra prediction output. Figure 4 shows the approach of intra-prediction where an additional image resolution enhancement approach has been included. It has been observed that a best resolution image provides better intra prediction output.

In the proposed approach, 4k video applied as input with frame rate of 35. Video is then converted into frames (Hereafter Images) for further processing. The images are then pre-processed following color conversions, image re-sizing and image transformation. Image in YC_bC_r format is converted to RGB and further converted to grayscale for processing. The image is re-sized to the ratio of 1920:1280 to simplify the processing without losing any information. The obtained image is treated as matrix and then transformed to prepare for quantization; image transform is then followed by scaling. The image with different quantization parameters has been partitioned into different sized block sets such as 4x4, 8x8, 16x16 and 32x32 matrix.

The reference samples have been mined from upper row and left-hand column neighboring of the particular block. For the scenario such as corner blocks, where the reference block is unavailable, the closet reference block has been considered. The image is the inverse quantized and inverse scaled; before image resolution enhancement. The obtained output is fed to image resolution enhancement and PSNR and MSE parameters have been calculated for individual sample with the help of screen resolution factors namely two, four, eight and sixteen. The analysis with respect to mentioned parameters is done to select appropriate image bearing in mind dissimilar form factors for specific image formats. The selected best image is then up-sampled and forwarded for intra-prediction.

During intra prediction, order selection is required and for the testing purpose, 64x64 order is chosen. Then CU and PU is determined using CTU. The prediction modes targeted are DC mode, planar mode and 33 angular mode, with smoothening approach of neighborhood pixel approach (Top and Left). The predicted frames are then reordered and predicted output is obtained. The intra prediction flow adopted is discussed in Figure 5. As per the literature, the lesser block set results in lesser complexity, ease in transform and quantization coefficient computation, hence the maximum block set of 32x32 has been considered. To test the image with different quantization parameters, image has been partitioned into different sized block sets. The reference samples have been mined from upper row and left column neighboring of the selected block. For the scenario such as corner blocks, where the reference block is unavailable, the closet reference block has been considered. As mentioned earlier, seven different modes have been processed, by considering different set of reference samples providing prediction block for each mode.

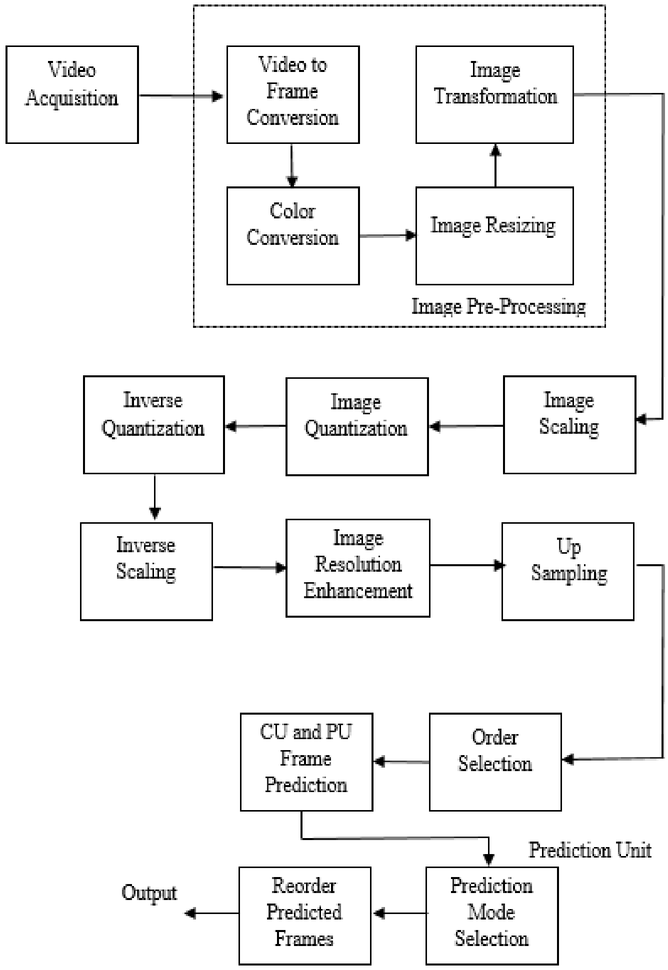

Fig. 4 Improved Intra Prediction Approach.

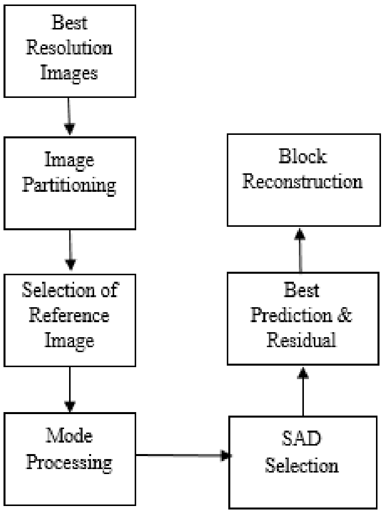

Fig. 5 Intra-prediction Process Flow.

Sum of Absolute Differences (SAD) for each mode has been computed and then compared to obtain the lowest SAD to select the appropriate mode. In case of two or more modes with equal SAD, the last processed mode has been considered. The best SAD for a block gives information about the best prediction block. The original block and best prediction block is subtracted to obtain the best residual block. The obtained residual-block has been transformed and quantized based on quantization parameters. To avoid any errors during decoding, the block has been reconstructed by performing de-quantization and inverse transform. The reconstructed block is used as the reference of the next block to be predicted.

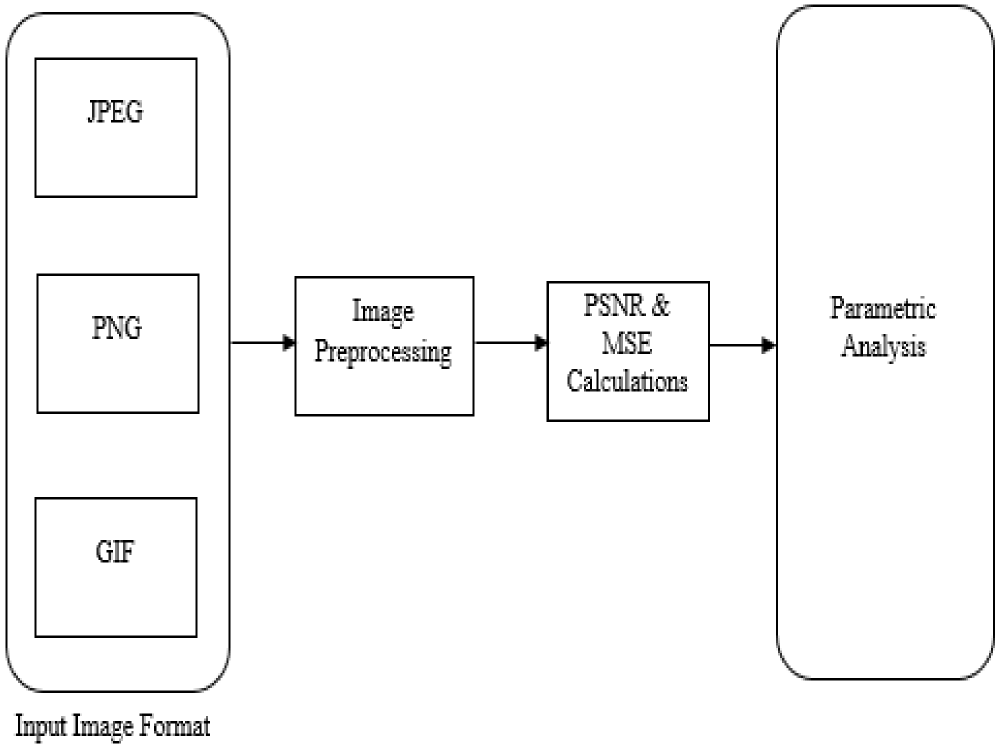

Fig. 6 Resolution Enhancement Approach for Intra-Prediction. [18]

The model shown in Figure 6, supports image formats namely, JPEG, PNG and GIF. The image frames are pre-processed by re-sizing, interpolation and filtering the noise in the pre-processing unit. Image parameters such as PSNR and Mean Square Error (MSE) have been calculated for every sampled image considering 2, 4, 8 and 16 as screen resolution factors. Based on the different form factors parameter analysis have been carried out and best suitable image has been selected for intra prediction.

B. Convolutional Neural Network Approach for Intra Prediction

The CNNIP model has been employed to predict CU depths and based on the predicted CTU, partitioning has been carried out, which further does intra prediction. The CNN model has been developed with an assumption to implement on FPGA, thus the model should be simple model and maintain the performance as obtained by the conventional approach developed earlier. In HEVC intra-prediction, for each 64x64 CTU, the encoder takes more time for identifying the best CU depths (16x16 matrix). Hence to minimize this time CNN model has been employed which identifies the best CU depth and thus predicts CTU. HEVC-CU-depths dataset has been utilized for training this model. YUV videos are used with encoding information for database.

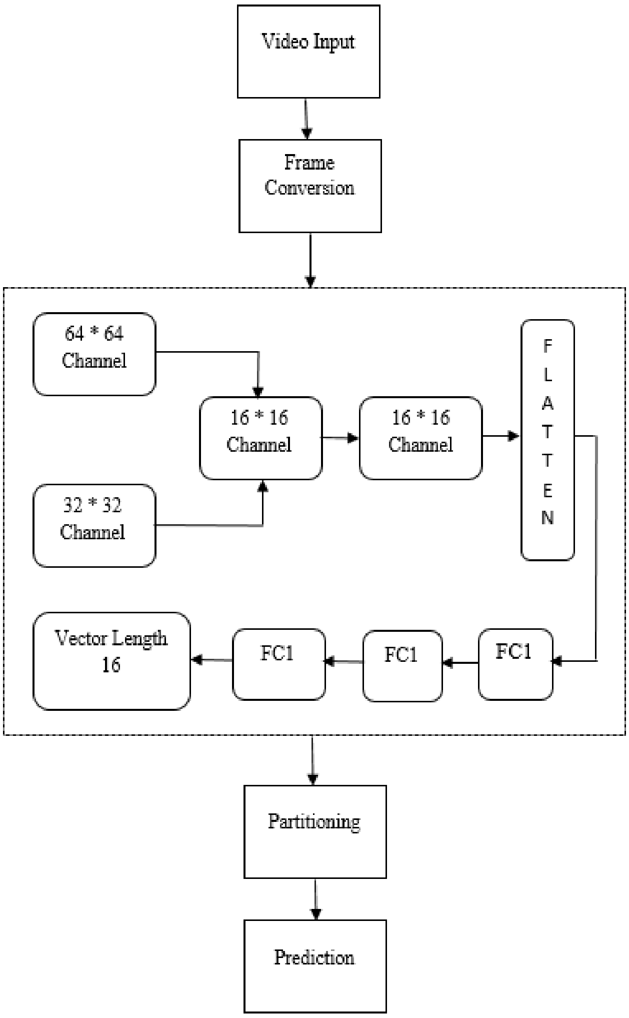

Fig. 7 Proposed CNN model for Intra Prediction (CNNIP).

Further the database has been processed by converting videos into image files with corresponding serialized encoding information. Total 890 images have been created and divided as 60%, 20%, 20% for train, test and validation respectively. The developed model has been tuned with the hyper parameters as batch size, epochs and learning rate as 128, 50 and 0.001 respectively. Focused functionality of model has been to obtain a 64x64 CTU, by predicting a 16x16 matrix which represents its depth decision to obtain output vector of length 16. A 64x64 CTU means to 16 labels and if it has to divide into four 32x32 CUs, then each 32x32 CU corresponds to only four labels, which reduces complexity of task. The depths considered are 0, 1, 2 or 3. 0^{th} depth signifies the 64 x 64 CU and encoded as it is. 1^{st} depth wants that the 64 x 64 CU will be further split into four 32 x 32 CUs and then encoded. The proposed model uses both the 64 x 64 CU and the current 32 x 32 CU as input, and outputs 4 labels signifying depths.

V. Results and Discussions

The results of improved resolution enhancement approach as discussed in the earlier works [18, 19] are showing better PSNR for screen resolution factor 4. For validating the performance of intra-prediction, PSNR parameters have been chosen primarily. The same approach when adopted to the proposed intra-prediction, it has improved the PSNR and structural similarity values. Table 1 shows the obtained PSNR values for first 10 modes. Figure 8 and 9 shows the complete values of PSNR and SSIM respectively for all 35 modes. It has been observed that, after 14^{th} mode the PSNR and SSIM degrades and again after 25^{th} mode it regains its peak. The CNNIP model maintains the similar PSNR and SSIM values with reduction in the time required for processing for quicker CU depth prediction. The model has achieved 85% accuracy due to CCN approach incorporated in the work. The quality of the predicted output improved as the value of the PSNR increases.

TABLE I. PSNR FOR FIRST 10 MODES

Modes	PSNR
	CNNIP architecture (proposed)
1	**33.3447**
2	**46.0164**
3	**41.0157**
4	**42.0298**
5	**42.3653**
6	**42.6379**
7	**43.5977**
8	**43.6052**
9	**45.6946**
10	**45.7546**
11	**45.8041**

Fig. 8 PSNR for 35 modes of intra-prediction.

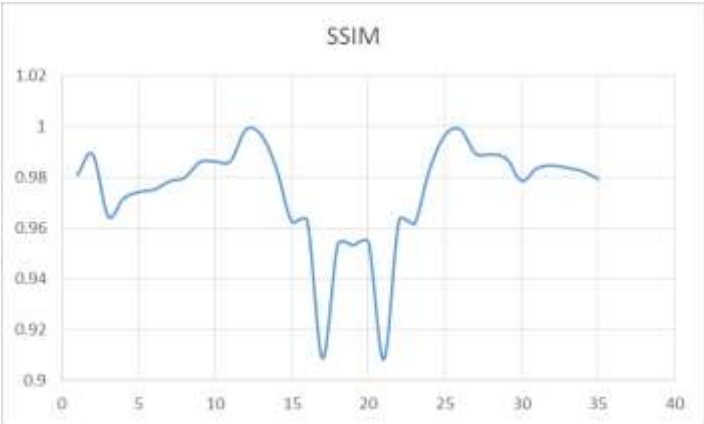

Fig. 9 SSIM for 35 modes of intra-prediction.

VI. CONCLUSION

Presently HEVC technique is the best approach used to achieve better video compression by retaining the quality of video. The CNNIP has been analyzed for 35 different modes using SSIM and PSNR parameters. The modified resolution enhancement improves the video quality as the average structural similarity index obtained is nearly 96% similarity to input data. PSNR obtained is well within the acceptable range of 30 to 50 dB.

REFERENCES

[1] G.J. Sullivan, J.R. Ohm and W.J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Transactions on Circuits Systems, Video Technology, vol. 22, issue 12, pp.1649–1668, 2012.

[2] M. Abadi, Tensorflow: large-scale machine learning on heterogeneous distributed systems (2016), https://arxiv.org/abs/1603.044672016.

[3] Jorge Soto León, Ernesto Villegas Castillo and Carlos Silva Cárdenas, "A High Parallel HEVC Fractional Motion Estimation Architecture", Proceedings of IEEE International Conference on ANDESCON, pp. 1-4, 2016.

[4] Abdessamad El-Ansari, Ali Ahaitouf and Anass Mansouri, "An Efficient Hardware Architecture of Intra Prediction in HEVC Standard", Proceedings of 11th International Symposium on Design & Test, pp. 319-322, 2016.

[5] Jianbin Zhou, Takeshi Yoshimura and Satoshi Goto, "VLSI Implementation of HEVC Motion Compensation with Distance Biased Direct Cache Mapping for 8K UHDTV Applications", IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, issue 2, pp. 380-393, 2017.

[6] Y. Chen and W. Jiang, "Gradient based Fast Mode Decision Algorithm for Intra Prediction in HEVC", Proceedings of 2nd International Conference on Consumer Electronics, Communications and Networks, pp. 1836–1840, 2012.

[7] Pengfei Wan, Jiexiang Tan and Xiaocong Lian, "High Bit Depth Image Acquisition Framework using Embedded Quantization Bias", IEEE Transactions on Computational Imaging, vol. 6, issue 4, 2019.

[8] Lizhi Wang, Zhiwei Xiong, Hua Huang, Guangming Shi, Feng Wu and Wenjun Zeng, "High Speed Hyperspectral Video Acquisition by Combining Nyquist and Compressive Sampling", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, issue 4, 2019.

[9] Alvaro Q Valenzuela and Juan Carlos G Reyes, "Basic Spatial Resolution Metrics for Satellite Imagers", IEEE Transactions on Sensors Journal, vol. 19, issue 13, 2019.

[10] Lingyan Xiu, Bojiao Ma, Konglin Zhu and Lin Zhang, "Implementation and Optimization of Image Acquisition with Smartphones in Computer Vision", Proceedings of 2018 International Conference on Information Networking, 2018.

[11] H. Huang, I. Schiopu and A. Munteanu, "Deep Learning Based Angular Intra-Prediction for Lossless HEVC Video Coding", Proceedings of Data Compression Conference DCC, pp. 579-579, 2019.

[12] Schiopu, H. Huang and A. Munteanu, "CNN Based Intra Prediction for Lossless HEVC", IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 7, pp. 1816-1828, July 2020,

[13] V. Sanchez, M. Hernández-Cabronero and J. Serra-Sagristà, "Hybrid Intra Prediction in Lossless Video Coding using Overfitted Neural Networks", Proceedings of Data Compression Conference (DCC), pp. 369-369, 2021.

[14] Mai Xu, Tianyi Li, Xin Deng, Ren Yang and Zhenyu Guan, " Reducing Complexity of HEVC : A Deep Learning Approach", IEEE Transactions on Image Processing, vol. 27, issue 10, pp. 5044-5059, 2018.

[15] Raz Birman, Yoram Segal, Avishay David and Ofer Hadar, "Intra Prediction with Deep Learning", Proceedings of Applications of Digital Image Processing XLI, 2018.

[16] Thorsten Laude and Jorn Ostermann, "Deep Learning Based Intra Prediction mode Decision for HEVC", Proceedings of 32nd Picture Coding Symposium, Germany 2016.

[17] Damian Ruiz, Gerardo, Jpse L M and Pedro Cuenca, "A Unified Architecture for fast HEVC Intra-Prediction Coding", Journal of Real Time Image Processing, vol. 16, issue 5, pp. 1825-1844, 2019.

[18] Swamy T N and Ramesha K, "Video and Image Acquisition Challenges to define Image Resolutions for HEVC", International Journal of Advanced Science and Technology, vol. 29, issue 5s, 958 – 965, 2020.

[19] Swamy T N, Ramesha K and Diwakar K R, "An Efficient Algorithm for Intra Prediction in HEVC/H.265 Standard for 16 x 16 Pixels", Proceedings of International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2017.

[20] Dai, XU. Visual Saliency Estimation Via HEVC Bitstream Analysis. Diss. University of Sheffield, 2018.

[21] Liang Peng, Ru Lai, Xiaodong Yang, Jie Cui and Guoqianlug Wu, "Design and Implementation of Video Image Acquisition and Processing System based on CPLD", Proceedings of the 29th Chinese Contro Conference, 2010.