International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.07, April 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Deep Learning Techniques in HEVC

Besiahgari Dinesh[1], Kavya B[2], Dr. Mohammed Riyaz Ahmed[3]

[1,2]*School of Computing and Information Technology, REVA University-560064, INDIA*

[3]*Associate Professor, School of Electronics and Communications Engineering, REVA University,INDIA*

ABSTRACT

H.265/High Efficiency Video Coding (HEVC), the recent international standard video coding which is under development, has shown a major breakthrough in video compression efficiency. The encoding efficiency of HEVC is higher than that of previous standards. One of the factors for higher efficiency is intra picture prediction which has a large number of prediction directions (35 modes) when compared to the predecessor standards. This high efficiency is made possible with a tradeoff between high complexity of encoders. The main drawback is the inclusive rate distortion (RD). This paper proposes Deep Learning (DL) concept which uses Convolutional Neural Network to predict the best mode which is having the less rate distortion (RD) there by achieving the comprehensive rate distortion optimization (RDO).

Keywords: Deep Learning, HEVC, Rate Distortion (RD), Rate Distortion Optimization (RDO)

I. INTRODUCTION

Video coding (compression & decompression) is used to manipulate video signals to reduce bandwidth and the memory required without compromising on the video quality. The basic video codec is shown in fig:1

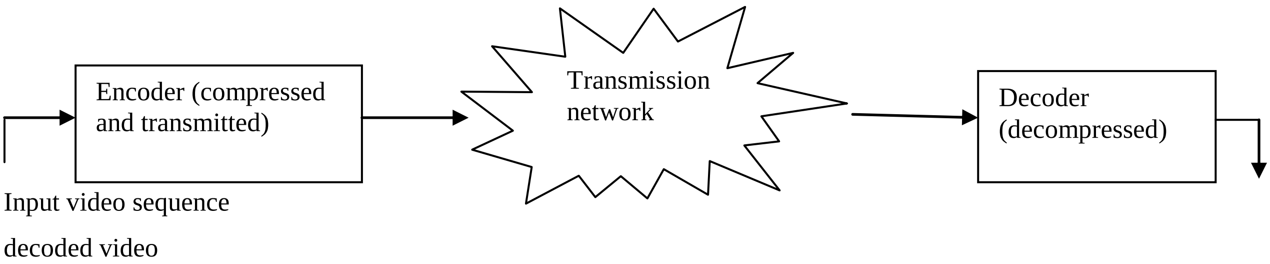

Figure:1 Basic Video Codec

HEVC/H.265, the recent video coding technique, standardized by the Joint Collaborative Team on Video Coding (JCT-VC), ITU-Telecommunication (ITU-T) ,Video Coding Experts Group (VCEG) and ISO/IEC MPEG in January 2013. It has wide range of various applications in the area of video streaming services like internet video uploading and downloading, live streaming channels, video conferencing applications, and terrestrial transmission systems, video, camcorders, mobile networks, security applications.

The video coding standard preceding the HEVC project was H.264/MPEG-4 AVC, standardized in 2003 and showed significant improvement in several crucial applications from 2003–2009. H.264/AVC has been an empowering video coding technology for digital video in all possible domains that the H.262/MPEG-2 could not conquer.

The main aim of HEVC/H.265 is gradually increase in the compression performance compared to the previous standards and reduce the bit rate by 50% for equal perceptual video quality. HEVC has many additional features compared to previous standards. Similar to H.264, H.265 is based on hybrid coding approach which is combination of inter-picture prediction, intra-picture prediction, in-loop filters (Deblocking filter (DF) & Sample Adaptive Offset (SDO)), transformation coding. HEVC has 35 modes (the DC mode:(mode 1), 33 angular modes:(mode2-34) and the planar mode:(mode 0)) when compared with AVC which has 9 modes. Due to this high increase in the modes, rate distortion optimization (RDO) becomes the major intricacy. To overcome this, we intend to circumvent the RD optimization intricacy for intra picture prediction mode decision. The intra prediction modes can be formulated as categorization with different mode classes in which ML suggests themselves as solution. By using the concept of Deep learning in which CNN provides the solution in identifying the best mode with less rate distortion which will lead to increase in the video efficiency by reducing the bit rate and maintaining the same video quality.

This paper gives us an insight of video encoders deploying DL techniques. This is based on convolutional neural networks classifiers in encoding process. By means of this CNN encoding process, we avoid RDO complexity.

II. BACKGROUND

2.1 Fields of Video coding & decoding

With the fast increment in the network bandwidth and wide spread usage of thin-client devices, Internet and wireless communication, some of the rich applications such as wireless display, video transmission to heterogeneous clients with different screen resolutions, Multicast service, extensive use of mobile device, broadband networks especially 3G/4G networks, have allowed these video coding communications to become important in people everyday life. The Artificial Intelligence configures using the low complexity devices like Personal Digital Assistant, high quality wireless video cameras. Today, computer screen sharing has become very popular in many applications like cloud computing, provides screen sharing technology, power point slides.

2.2 Evolution of Video Coding

The ISO/IEC MPEG & H.26x publishes standards for media coding, transport & storage. These standards consist of different *Parts*. The VCEG, a working group of the ITU-T, published early standards in the H.26x family. The Advanced Video Coding, the video coding standard was co-developed by the JVT, collaborated between VCEG and MPEG.

The below given table describes the main key characteristics of the video coding standards:

Standard	Key Features
MPEG-1	This is standardized in 1993 It supports progressive video Supports YUV 4:2:0 352*288 resolution
MPEG-2	It is standardized in 1995 It supports video on standard definition TVs It divides video signal into two or more coded bits with different resolutions.
MPEG-4	It standardized in 1999 It supports in low bit rate multimedia applications in mobile platform & internet Supports coding of video and audio including animation
H.261	It is standardized in 1988 Developed for conferencing over ISDN Uses block-based hybrid coding with picture motion compensation
H.262	It s standardized as MPEG-2 part-2 in 1995 Handles both YUV 4:2:0 and 4:2:2 formats Supports progressively and interlaced scanned pictures Other features include data partitioning, on-linear quantization
H.263	It was standardized in 1996 and H.263+ was standardized in 1998 Quality is improved compared to H.261 at lower bit rate to enable telephony and video conferencing
H.264(AVC)	Standardized in 2003 Supports video on the hd TVs, mobile devices, internet Improved picture quality, at low bit rates In-loop DF to reduce block discontinuities
H.265/HEVC	It s standardized in 2013 Supports video quality up to 8k resolution Rich flexibility in prediction Reduction of 50% more bit rate compared to AVC

Table:1 Evolution of video coding family

III. Literature Survey

3.1 Introduction to research motivation

With the advancements in the video coding standards, due to the current network bandwidth the video cannot be transmitted efficiently with less band width. This is the motivation to the researches to reduce the maximum redundancy possible through intra picture prediction and inter picture prediction . In HEVC, the modes have been increased when compared to AVC and so the complexity is increased. Many techniques and algorithms came into existence to reduce the bit rate by 50% so that the encoding and decoding computational time is reduced without reducing the video quality.

Algorithm	Characteristics
Early termination of RD mode decision process	Intra prediction searches from top down and total no of RD calculations is reduced by checking the correlation between neighboring depths 1% BD rate loss and 67% faster encoding
Adaptive Quantization technique	On an average of 0.35% & 0.46% bit rate reduction BD PSNR improvement of 0.065db & 0.038db
Encoding a base layer using H.265- compliant encoder and encode an enhancement layer using BCEC with exploitation of base layer info	8% reduction in bit rate on an average 18.9% bit rate reduction for screen content video as compared with SHEVC
Using Efficient inter prediction mode (EIPM) for Motion Estimation(ME)	Computes the priority of all inter prediction modes and performs ME only on a selected inter prediction mode
Introduction of two chroma intra prediction modes LML & LMA	0.2%,5.9%,6.7% BD rate reduction for Y, Cb, Cr respectively Complexity added at decoder side is negligible
CU Depth decision based on latest sum of absolute decision	Implemented by comparing a derived threshold with SAD between the upper & sub SAD costs Increase in bit rate of average 1.61% and 2% Reduction in computation time by 52% and 58.4%
Combining most probable mode flag signaling and intra mode signaling	Takes neighboring intra modes into account to obtain a prioritization of different modes 0.33% reduction in bit rate Minimal increase in complexity in encoder and decoder
Tree depth inter-level	Inter-level correlation observed between predominant edge orientation of current PU and already coded PU in lower levels. Edge orientation is used to choose the subset of modes that will be used in intra coding. 5 subsets of angular modes are defined; each 1 is composed of 9 angular modes from 33 angular modes. Each subset is associated with 1 edge orientation (horizontal, vertical, 45,135°, non-directional. Increase in bit rate of average 1.2% & decrease up to 40% of processing time
Using rectangular partitioning for intra prediction	Early skips are used for reducing complexity and by achieving better compression efficiency 1.8 % bit rate loss with 2-point prediction and transform, 1.4% bit rate loss without 2-point prediction and transform Coding gain is achieved
Using RD-cost	48% computation complexity reduction comparing to original SHVC & 12% time saving
Adaptive template selection	Extend the neighboring pixels and remove unavailable pixels before the calculation of parameters 0.3%,2%,3% for high QP (0.3%,1.2%,1.8% for normal QP) for Y, Cb, Cr on average

Coding unit depth decision algorithm	Negligible impact on rate distortion
Transform unit depth decision	Uses rough mode cost(RMC) Increases bit rate by 0.21% and encoding time is reduced by 3.8%

Table:2 Algorithms used in Intra-prediction and Inter-prediction

3.2 Decoding the working of HEVC

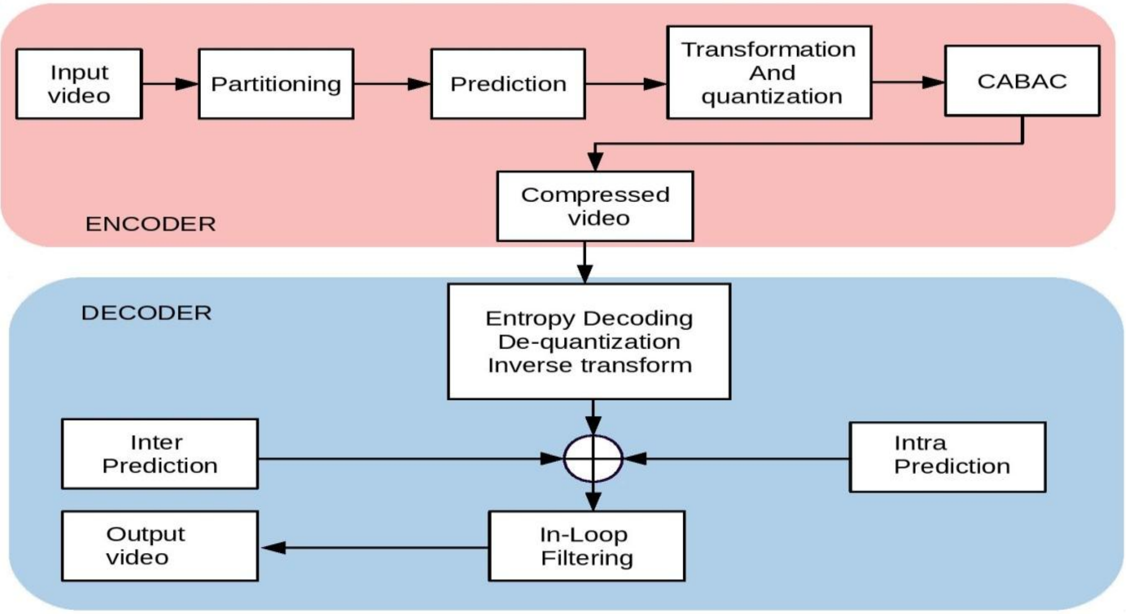

Figure: 2 Block diagram of H.265

HEVC is a hybrid coding approach having intra picture prediction, inter picture prediction and 2D transform in its video coding layer.

The encoding algorithm proceeds as follows:

Each picture is split into block shaped regions called Coding Tree Units (CTUs). The CTU is the basic unit of coding which are similar to the macro blocks used in the earlier video coding standards (AVC/H.264). It is further split into regions called Coding Units (CU).

HEVC supports two kinds of predictions

- Intra: Where every unit is estimated from adjacent image data within the current picture.
- Inter: Employs image data from other reference pictures.

The first frame in video sequence is employed with intra picture prediction as there are no other frames to compare (reference frames), the other pictures of a sequence or between any random points are predicted using inter picture predictive coding modes.

The encoding process for inter picture prediction consists of choosing motion data comprising the selected reference picture and motion vector(MV) to be applied for predicting the samples of each block. The encoder is given a sequence of frames as an input in which each frame is segmented as Quad-tree structure and each segment is fed to the encoder to one by one to decide whether the block is compressed as P-frame or I-frame. By using mode decision and motion vector identical inter picture prediction is generated by encoder and decoder applying MC. In H.265, the original block image frame is subtracted from prediction block image frame or reference frame to obtain the difference between the frames. This resulted block is called error image block or residual image block. This residue block of intra/inter is transformed by spatial transform. These residual transform coefficients, passes through scaling, quantization, CABAC and are dispatched together with the prediction information.

The encoder replicates the decoder processing loop such that both will generate identical predictions for upcoming data. Later, the quantized transform coefficients are reconstructed by inverse scaling and transform to duplicate the decoded approximation of decoded signal. Both the error image block and the predicted block is added and this is fed to in-loop filters to remove the artifacts which are obtained due to quantization. The decoded picture buffer will store the final picture which helps for inter picture prediction to compare the predicted block with reference pictures.

3.3 Finding Loop holes in INTRA prediction

Intra prediction is a highly complex tool. The increase in the number of modes increases signaling overhead in the bit stream. Selecting the best mode with less RD using manual techniques is a difficult task and time consuming and this can be overcome by using machine learning techniques.

3.4 Machine Learning approaches as solutions

HEVC encoder should select the best intra picture prediction modes using UDI prediction supporting 35 modes, including two non directional modes and 33 angular modes (directional). The eminent mode is selected by choosing the minimum mode resulting from the RDO process. In order to minimize the computational time of the encoder we go for ML based accelerated intra picture prediction algorithm.

IV. Related Work

4.1 ML approaches

4.1.1 K-medoid clustering algorithm

This uses an early winding up criterion based on Hadamard cost statistics. We cluster the intra picture prediction modes into K-clusters using this machine learning algorithm K-medoid. The center of the clusters is taken as candidate for rate distortion optimization (RDO) process. This K-medoid algorithm uses the depth range prediction method to find all the correlations between the neighboring coding units (CUs). To find whether the coding un it is split we make use of HSAD & the RD-costs of the encoded coding units. At the final step, this K-medoid algorithm selects the intra prediction mode (IPM) which is having lower precision in RMD and based on correlation between neighboring coding units (CUs), RDO is reduced for the IPMs. A gradient algorithm in HEVC is used to reduce the computational complexity by maintaining the same video quality.

4.1.2 Random Forest algorithm

As we know that the HEVC has 35 intra prediction modes and therefore increase in the encoding computational time. This encoding intricacy can be minimized if the parts of the decoder can automatically predicting from the encoder of original HD video. We see the prediction of CTUs into CUs by training the recent 10 frames of newly viewed by using this algorithm. This random forest algorithm creates classifier from N decision tress. Each tree has subsets for each input features. To identify the output label of new sample given as input, the probability of splitting of the samples is returned from the each tree in forest. The average of probabilities is considered and if the average is greater than 50% CU is split

4.1.3 K-NN algorithm

This algorithm uses search strategy that shows a strong impact on the quality of the predictor. This is used as template search representation as sequential combination of K-NN templates (K-templates) taken from SW. K-nearest neighbor is found by computing distance between candidate block in reference frames and template of current block. Different approaches rely on K-NN, since it s explored on inter frame and intra frame such as nonnegative matrix factorization or sparse representation. This is mainly used for prediction units especially large size unit so that it provides better RD optimization. This method is also highly parallelizable and can reduce the computational time.

4.2 Introduction to Deep Learning

Deep learning, a novel technique under machine learning helps in reduction of complexity in H.265. The CU partitioning features can be automatically predicted or extracted by deep learning rather than using manual techniques. The advantage of using the deep learning approach in HEVC for reducing the complexity is that it can take large scale data as advantage to automatically mine extensive CU related partitioning. Considering, mode selection in the HEVC as a classification problem, this approach can give an effective solution in deciding the best mode having less RD. For superior results in classification problems deep convolutional neural network is employed.

4.3 Introduction to Convolutional Neural Network

This neural network consists of neurons connected together which can result in deep structural network. The advantage of CNN is that during Back-propagation, the network has to adjust a number of parameters equal to a single instance of the kernel which reduces connections from the typical neural networks. In order to address the problem of blocking artifacts resulted from block wise inter or intra picture prediction, transform scaling and quantization, deep convolutional neural network is employed which approximates the reverse function of video encoding. CNN is independently computable from any predecessor encoder decision and reconstructed sample values. By removing the unwanted common coupling between the encoder decisions from actual encoding process all decisions are carried out concurrently for all the blocks and so there is no additional latency required. Convolutional Neural Network (CNN) is applicable for SCC systems.

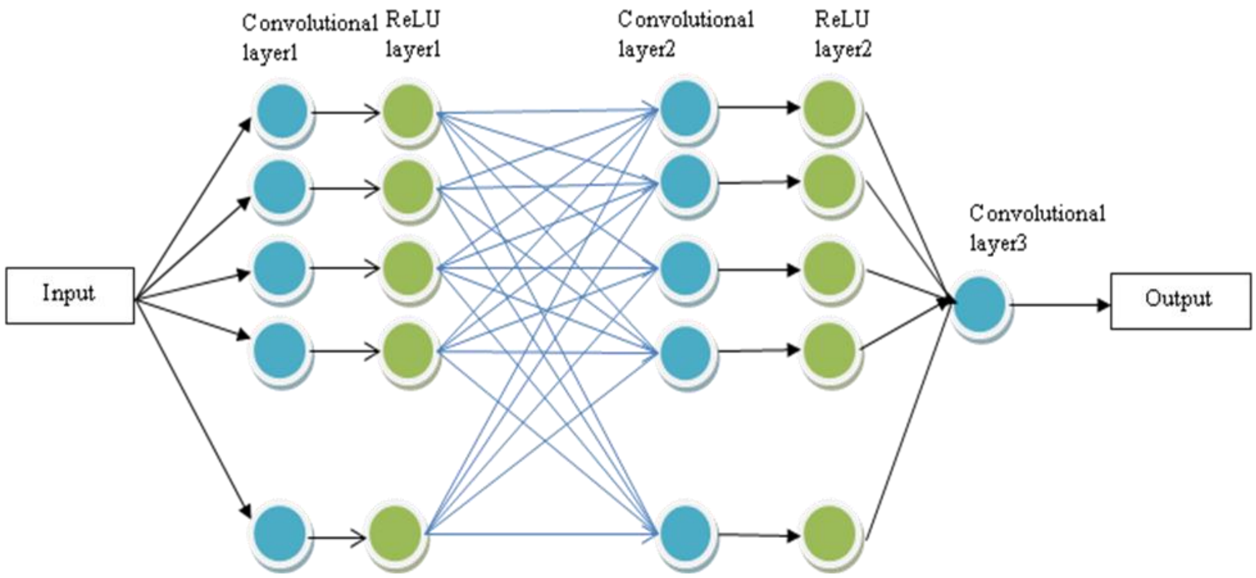

Figure:3 CNN architecture

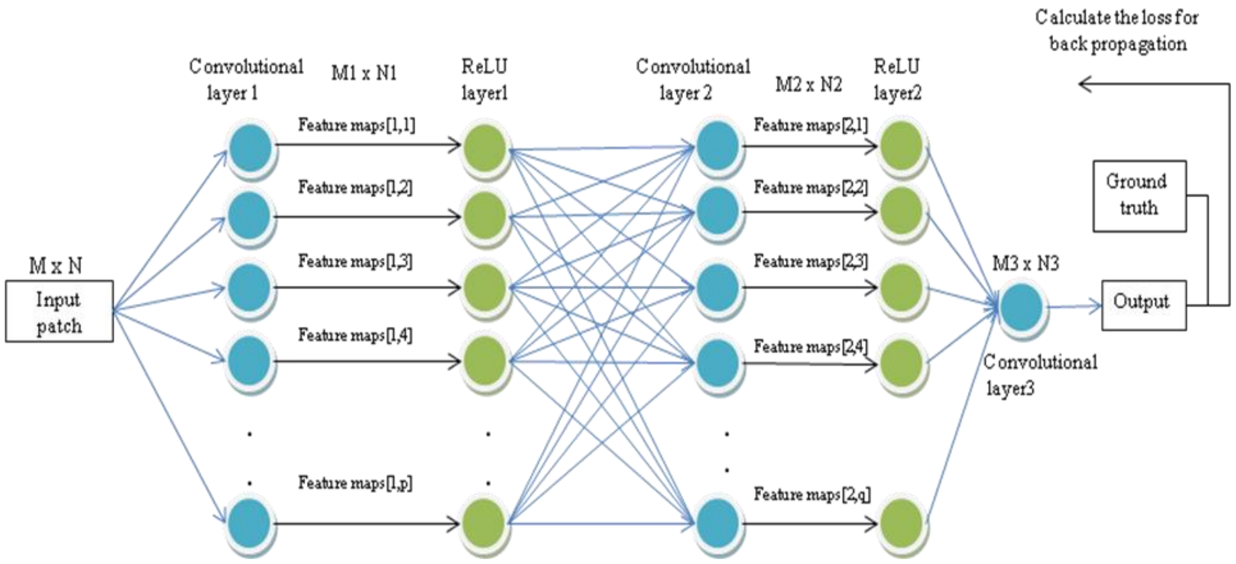

Figure:4 Network training

At first, the classifier's CNN is fed with original input samples, thus enabling it to be independent from the encoders decision or the reconstructed samples of the encoder. This also avoids delay in processing. Out of one luma and two chroma components of each CTU, the luma components contain major information compared to the chroma components, only the luma components are used for processing. As mentioned earlier the main motive of introducing CNN here is to solve the classification problem considering each of the 35 different intra-prediction modes as different classes. Each input sample will be made to pass through three layers including one max pooling and two convolutional layers which help in filtering the data. The Rectified Linear Unit being hidden is used to activate neurons .At last, the data passes through the two fully-connected layers and the output is retrieved.

V. Discussion

5.1 Summary

HEVC /H.265 being the next generation video coding achieves superior coding influence compared to previous video coding standards at the expense of complex encoders, where rate distortion optimization(RDO) is a major contributor. In order to overcome the mentioned short coming, we aim to avoid RD and embrace machine learning (ML) algorithms for intra-prediction mode decision. Deep Learning being an attractive technique for most of the research scholars provides striking results for classification problem by employing deep Convolutional neural networks (CNN's) exploiting CNN's for intra-prediction mode decision.

5.2 Issues

- Coding gains obtained by the hybrid motion- compensated block transform architecture using scalar quantization and entropy are expected to reach saturation
- Intermediate storage buffers increase the implementation cost
- Freedom of encoders having multiple ways for motion–compensated picture partitioning increases complexity as well as the encoding time.
- As HEVC supports large block sizes it demands for higher cache memory.

5.3 Challenges

- Reducing motion estimation complexity which occupies 77-81% of HEVC encoder implementation.
- It is unavoidable to completely eradicate complexity since it remains intact with RD performance and arises difficulty in compensating complexity at the cost of RD performance.
- Complex inter prediction brings heavy computing cost and leads to restriction on the promotion of HEVC

5.4 Tradeoff

- The HEVC encoder that serves to be more complex however provides a substantial benefit in improving rate distortion performance.
- Disabling time consuming features and tools in HEVC simultaneously brings reduction in coding efficiency gain compared to H.264.
- There exists a tradeoff between compression efficiency and encoding complexity when early skips are used for the reduction of encoder search complexity.

5.5 Expectations

- A real time encoder which is efficient in compression is highly expected in the near future so that the computational time to encode and decode is reduced.
- Researchers should also focus on the improvement in the coding speed of Intra prediction for Quality SHVC.

VI. CONCLUSION

Through this paper, we formally introduce CNN based classification for predicting the best directional mode for H.265. The original video samples are fed into the neural network as input. The decision to elect the best intra picture prediction mode is formulated as a categorization problem. Bjϕntegaad-Delta rate losses become insignificant with a reduction upto 0.521% over H.265. Thus, CNN provides the suitable video encodes. This can be mainly used in application scenarios like streaming video involving simultaneous production and retrieval of digital content over Internet. Hence, a single categorization for the finest intra picture prediction mode can be employed in any encoder.

VII. ACKNOWLEDGEMENTS

The authors would like to express our gratitude to Dr. P. Shyama Raju, Chancellor, REVA University for providing us with the best infrastructure in successful completion of this paper work.

REFERENCES

1. Sullivan, Gary J., et al. "Overview of the high efficiency video coding (HEVC) standard." *IEEE Transactions on circuits and systems for video technology* 22.12 (2012): 1649-1668
2. Al-Barwani, Maryam, and E. A. Edirisinghe. "A machine learning based framework for parameter based multi-objective optimisation of a H. 265 video CODEC." *Future Technologies Conference (FTC)*. IEEE, 2016.
3. Wang, Tingting, Mingjin Chen, and Hongyang Chao. "A Novel Deep Learning-Based Method of Improving Coding Efficiency from the Decoder-End for HEVC." *Data Compression Conference (DCC), 2017*. IEEE, 2017.
4. Lin, Rongqun, et al. "Deep Convolutional Neural Network for Decompressed Video Enhancement." *Data Compression Conference (DCC), 2016*. IEEE, 2016
5. Laude, Thorsten, and Jörn Ostermann. "Deep learning-based intra prediction mode decision for HEVC." *Picture Coding Symposium (PCS), 2016*. IEEE, 2016.
6. Zhang, Yun, et al. "Machine learning-based coding unit depth decisions for flexible complexity allocation in high efficiency video coding." *IEEE Transactions on Image Processing* 24.7 (2015): 2225-2238.
7. Jens-Rainer Ohm and Gary J. Sullivan. "*High Efficiency Video Coding: The Next Frontier in Video Compression" IEEE* Signal Processing Magazine, January 2013
8. Martin Alain, Christine Guillemot, Dominique Thoreau, Philippe Guillotel."*Inter Prediction Methods based on linear Embedded for Video Compression",* Signal Processing: Image Communication , April 24, 2105
9. Javed Moazzam, Asha Mishra." *A Study on Fast Intra Prediction Mode Decision Algorithm in HEVC (2320-9801)",* IJIRCCE.2016, 0405126.
10. Jiawen Gu , Minhao Tang, Jiangtao Wen and Yuxing Han." *Adaptive Intra Candidate Selection With Early Depth Decision for Fast Intra Prediction in HEVC"* IEEE SIGNAL PROCESSING LETTERS, *VOL 25, NO.2,* February 2018.
11. Xingyu ZHANG, Oscar C. AU, Jingjingh DAI, Chao PANG, Feng ZOU." *NEW CHROMA INTRA PREDICTION MODES BASED ON LINEAR MODEL FOR HEVC*" IEEE ICIP 2012.
12. Mai Xu, Tianyi Li, Zulin Wang, Xen Deng, Zhenyu Guan." *Reducing Complexity of HEVC: A Deep Learning Approach"* arXiv: 1710.01218v1 19 September 2017 .