




Deep Learning Based Video Compression Techniques with Future Research Issues

Helen K. Joy¹ · Manjunath R. Kounte¹ · Arunkumar Chandrasekhar²  · Manoranjan Paul³

Accepted: 8 June 2023 / Published online: 28 June 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

The advancements in the domain of video coding technologies are tremendously fluctuating in recent years. As the public got acquainted with the creation and availability of videos through internet boom and video acquisition devices including mobile phones, camera etc., the necessity of video compression become crucial. The resolution variance (4 K, 2 K etc.), framerate, display is some of the features that glorifies the importance of compression. Improving compression ratio with better efficiency and quality was the focus and it has many stumbling blocks to achieve it. The era of artificial intelligence, neural network, and especially deep learning provided light in the path of video processing area, particularly in compression. The paper mainly focuses on a precise, organized, meticulous review of the impact of deep learning on video compression. The content adaptivity quality of deep learning marks its importance in video compression to traditional signal processing. The development of intelligent and self-trained steps in video compression with deep learning is reviewed in detail. The relevant and noteworthy work that arose in each step of compression is inculcated in this paper. A detailed survey in the development of intra- prediction, inter-prediction, in-loop filtering, quantization, and entropy coding in hand with deep learning techniques are pointed along with envisages ideas in each field. The future scope of enhancement in various stages of compression and relevant research scope to explore with Deep Learning is emphasized.

Keywords CNN · Deep learning · NN · HEVC · Video coding

Abbreviations

VVC	Versatile video coding
HEVC	High efficiency video coding
CNN	Convolutional neural network

✉ Arunkumar Chandrasekhar
arunkumar.c@vit.ac.in

¹ School of Electronics and Communication Engineering, REVA University, Bengaluru, Karnataka 560064, India

² Department of Sensor and Biomedical Technology, School of Electronics Engineering, Vellore Institute of Technology University, Vellore, Tamilnadu 623014, India

³ Charles Sturt University, Bathurst, NSW 2795, Australia

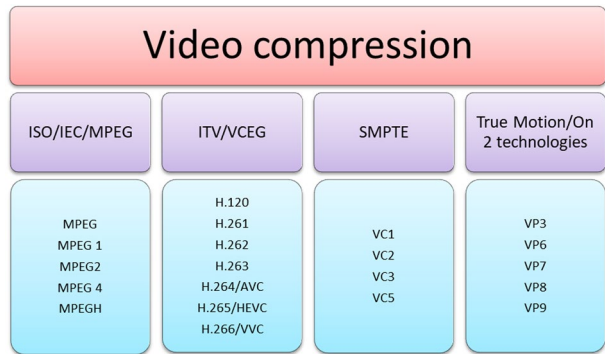
SRCNN	Super-resolution convolutional neural network
GOB	Group of blocks
DCT	Discreet cosine transform
AVC	Advanced video coding
UHD	Ultra-high-definition
DST	Discrete sine transform
DWT	Discrete wavelet transform
HT	Hilbert transform
CABAC	Context-adaptive binary arithmetic coding
NN	Neural network
DL	Deep learning
CTU	Coding tree unit
FRCNN	Faster region based convolutional neural network, fractional-pixel reference generation CNN
SVM	Support vector machine
CNNMCR	Convolutional neural network-based motion compensation refinement
VECNN	Virtual reference frame enhancement CNN
VRF	Virtual reference frame
DVRF	Direct virtual reference frame
FCNN	Fully convolutional neural network
RHCNN	Residual highway convolutional neural network
RDO	Rate distortion optimization
SAO	Sample adaptive offset
MIF	Multi frame in loop filter
DNN	Deep neural network
SSIM	Structural similarity index
VMAF	Video multimethod assessment fusion

1 Introduction

Video coding can be defined as the procedure of compressing and decompressing sequence of raw video frames. Video compression is in its fast-growing stage expect more accuracy, with less bitrate. The era of compression is currently researching on VVC (Versatile Video Coding), advancement of existing HEVC (High Efficiency Video Coding) with much more to focus to develop is also area of research. Various compression methods became so prominent when the transfer of data at a low bit rate became a need, but the quality of the image was also a concern. So, video compression and its variance became an eminent one in the research area. Digging back to the history of video compression [1, 2] the traditional methods started from basic coding like Huffman coding followed by frequency transformation to recent techniques like hybrid video processing techniques, HEVC [3, 4], video compression with deep learning [5, 6]. Based on the need of the user and various other features there are various compression techniques [7, 8], Its divided as in Fig. 1. Incorporating deep learning in various steps of the codec process can result in better efficiency and low bit rate results. The research focuses on the improvisation of video codec by adding the features of the deep learning process into it to get the best out of it.

Video is a set of moving images played in a particular rapid motion. Video can be analyzed or characterized by its various features such as the number of frames per

Fig. 1 Recapitulation of video compression techniques



second/frame rate, the scanning procedure that is interlaced or progressive, aspect ratio, color depth, video quality, and video compression methods, bit rate, video resolution etc. the knowledge about features are a matter of interest for exploring compression. Each of these characteristics is important for the analysis of video and the processing of it. Deep learning can be defined as a subset of machine learning based on neural network that extract more features and train the machine to perform a process. A deep network is designed to extract multiple data in input and convert it into different data representations in output. Deep network learns enormous data, it's convenient for massive unstructured data that adds the advantage to video coding [9–12]. CNN (Convolutional neural network) with deep layers could acquire great success in video coding, that started from 8-layer CNN [9], GAN (Generative adversarial networks) [13] SRCNN (Super-Resolution Convolutional Neural Network) and to many applications such as image classification, object detection, super resolution denoising compression artefact removal etc.

Video compression arose by considering features and making changes in it. Compression helps to avoid high bit rate and spatial and temporal redundancy. This helps to contribute a skeleton to various characteristics of the video. The recent era is focusing on intelligent or smart compression techniques that invites neural network into the working system of video compression. Much research is going on to improvise the existing video compression methods with existing neural network algorithms and some researches are focusing on new algorithms [10] for smart video coding. Considering both set the prominence by using the existing set of algorithms for video compression as the new algorithms totally depend on trial and error and need standardization too as theoretical values cannot be checked as early as the former. Both the kinds of algorithms are discussed in the paper, but mostly focused on the former to align with the research interest and implementation check. Initially the networks were shallow and as the advancement in calculation speed increased, the existence of more layers or deep learning become prominent. It gained it's speed in video coding too as it can extract more features compared to shallow network. Learning about deep learning-based approach not only add advantage in existing video compression standards till HEVC, but also helps in advancement of the recent compression techniques VVC.

The body of the paper is divided into 2 sections to give a detailed view about video compression. In Sect. 2 the various video compression standards, its evolution, and the steps in compression, that can be further used with deep learning. Section 3 focuses on the review of development in video compression techniques in hand with deep learning

in each stage of encoding and its advancement [10, 11]. The subsections explain the various advancement happened in each stage of compression in corporation with neural network, deep learning, its features, and future scope of development. The paper provides the future scope in development of video compression with deep learning techniques that finally culminate this work.

2 Motivation for Survey in Video Compression

2.1 Brief Idea on Video Compression and Deep Learning

Various compression methods became so prominent when the transfer of data at a low bit rate became a need, but the quality of the image was also a concern. Video compression had its way from Huffman coding followed by frequency transformation to recent techniques like hybrid video processing techniques, HEVC, VVC etc. Based on the need of the user and various other features there are various compression its divides as in Fig. 1. Video compression focuses on the reduction of spatial and temporal redundancy. Most of the video compression techniques [12] follows the same working principles, but the way of doing each step changes and become complex as compression technique develops. The universal standard organization has come up with various standards from H.261 to H.266. H.261 that can support only two video frames at a times is the start of video compression and transfer. H.263 the successor, failed to satisfy sending interlaced video. H263+ was developed for low bit rate that used, GOB (Group of Blocks) (176×48), Macroblocks(16×16). The inter and intra MB implies temporal prediction in P and I frame, followed by H.264 mostly called as AVC (Advanced video coding), its block segmentation based, motion compensated with DCT (Discrete Cosine Transform) technique. The aim behind AVC (Advanced Video Coding) was to transfer video in low bit rate with better efficiency for UHD (Ultra-high-definition) videos.

HEVC was the successor of AVC came with a view to have 25–50% of efficiency improvement than AVC maintaining same bit rate. It uses similar techniques in AVC with improvised comparison of patterns, difference in coding from 16×16 to a level of 64 to 64. Versatile video coding is future video compression technique or new generation video coding that expects 30~50% betterment than previous, focusing on 4–16 K and 360-degree video compression in an effective manner for both lossless and lossy compression. But the complexity of the system is expected to be high compared to HEVC.

Video compression mainly have some steps that can be generalized into picture partitioning, Codec set it into micro, macro, and many sub partitions of coding tree. In HEVC it is followed in 64×64 , 16×16 and 4×4 tiles [14]. Prediction based on frames in video compression such as intra prediction and inter-prediction focuses on compression within the frame or between the frames. Intra-prediction will send the motion vectors and the residual information as inter-prediction send prediction direction and the residual. Temporal redundancy is alleviated by interframe prediction and are widely accepted from H262 till VVC as the correlation is high between the neighboring frames.

In transform usually it follows DCT, DST (discrete sine transforms), DWT (discrete wavelet transforms), HT (Hilbert transform) which converts the blocks of pixels into frequencies and create a transform matrix that compress about 67% each prefers different transform, H.26×prefers DCT. In quantization, the lossy part of compression will happen and encoding uses various algorithms aiming to attain reduced bit rate. CABAC

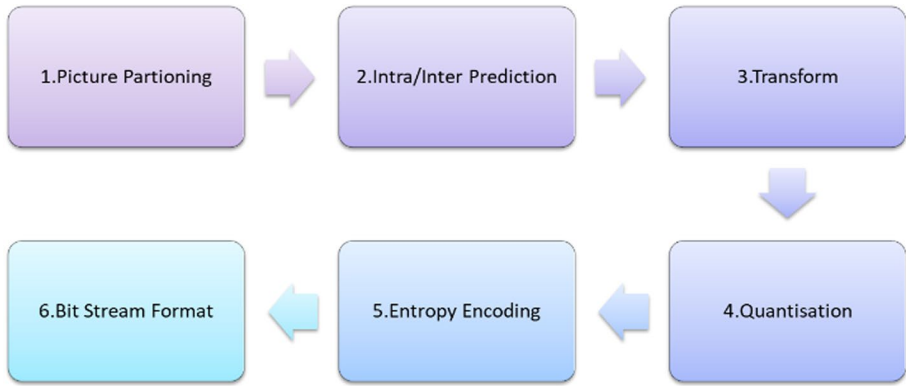


Fig. 2 Basic steps in video compression techniques

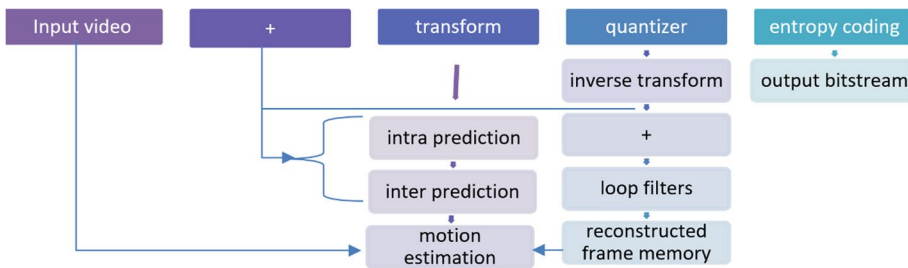


Fig. 3 Simple illustration of Video compression steps. Research gaps and future scopes are evaluated on each of these steps

(Context-adaptive binary arithmetic coding) is one of it. The encoded that is packed and send in various bit stream format. Loop filtering are prominent in AVC [15], HEVC, VVC etc. with deblocking filter and SAO (sample adaptive offset), compression artefact is an issue here to be alleviated.

In each of the compression steps in the Fig. 2, that illustrate compression, the advancement is still as scope, the introduction of neural network in this field started so. Recently deep CNN (convolutional neural network) achieved great success in video compression [11] as the feature extraction from massive video data and its transformation was convenient. It started with shallow network to deep CNN layers, the paper focus to review the recent advancement in NN (neural network) and DL (deep learning) based video compression (Fig. 3).

2.2 Motivation and Contribution

Research in video compression, especially compression with machine learning, neural network, deep learning is important area of research are important area of research. The focus is to know video compressions development from scratch till recent areas with the research gaps and scope of advancement in each area, this enhance the researchers to learn more and work on the needed flaws with proper understanding about overall system flow. The

whole video compression is split and analyzed under four main division for easy understanding, each provide an overview of the development happened in that area its research gap, open challenges and future scope of development. This motivates researches to work more focused as the information's are unified.

3 Recent Video Compression Techniques with Deep Learning

The computational model of deep learning was suitable for video coding because of its dense layer. From 1990's research on neural network-based compression in both video and image was existing [12] but it was not able to prove itself good by providing a better compression efficiency as its network were not too deep. As years went the computational power improved and it could handle and train huge database with deep or more layers. With this note of information an overview of NN/DL [10] based compression techniques and its scope is a worth topic to research. Some of the research scheme is focusing on the implementation of deep learning in traditional coding techniques, i.e., by applying the deep learning techniques in the existing steps of video coding to improve its coding efficiency [16].

A few new schemes were also evolved with deep learning but could not achieve much impact as the compression efficiency was not impressive and calculations were more theoretical and hardware implementations were tough. The recent video coding VVC, i.e., considered as the follower of HEVC, also aims to have deep learning-based video coding. So, the need to discuss on deep learning-based compression is relevant in research [17], that can help in improving existing coding and to develop new compression technique. This paper discusses the effect of deep learning various compression steps like intra prediction, inter prediction, motion estimation, quantization, entropy, loop filtering etc., its scheme proposed, advantages, current status of each steps are analyzed and future research scope.

3.1 Deep Learning Based Intra Prediction

Intra prediction is “the process in which the next part of an image frame is predicted using the previously encoded part of the same frame” [18]. The HEVC aims to make the bitrate by half by keeping the picture quality same compared to AVC and it is achieved mainly through intra /inter prediction. The techniques used in intraprediction varies from SVM (support vector machine), logistic regression, Decision tree, Bayesian classification, Random forest, Convolutional Neural Network and deep learning algorithms. Developmental stages in intra prediction, CTU (Coding tree unit) splitting algorithms and its pros and cons are discussed.

The training of chroma and luma components separately can improvise the efficiency of total system [19]. The computational complexity of the system is seen to be more as the number of computations are more, it's good to train the network to a smaller model to decrease complexity. Deep leaning method can give a better result if incorporated with existing methods [18]. Li et al. [20] is using correlation property of chroma and luma components to predict the successive ones. A neural network extracts the features of the current input and recovers the chroma samples simultaneously by extracting the features of the input (Figs. 4 and 5). The complexity of the network is more as the features are more.

All hybrid video encoding system usually perform an intra-frame prediction that uses the blocks of samples [21, 22] to predict from previously decoded frame of the same video.

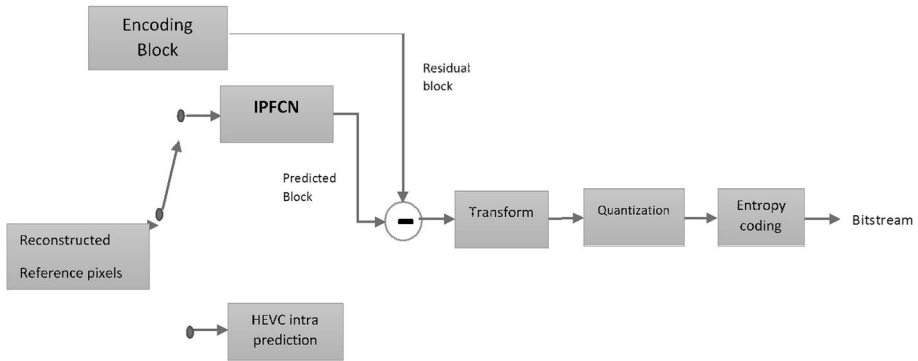


Fig. 4 A representation of embedding IPFCN in HEVC reference software [18]

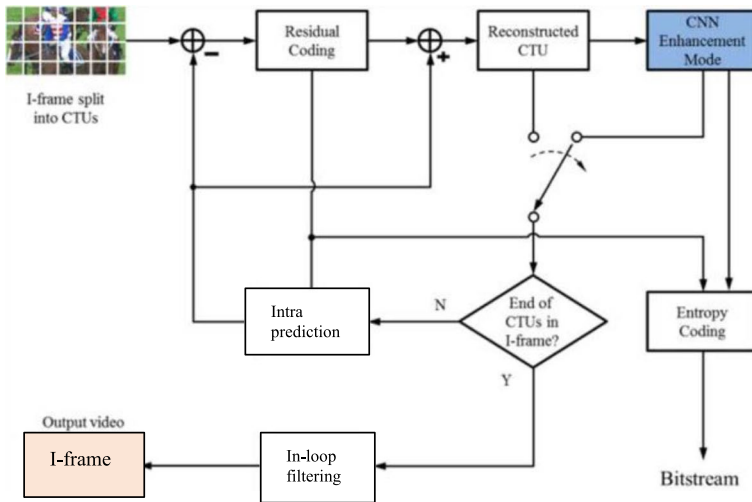


Fig. 5 Deep learning based Intra-frame coding for HEVC [28]

A group of angular prediction patterns is used by HEVC to exploit directional sample correlations [23]. New intra-picture prediction modes construction has the following steps, initially frame pattern as the prediction input. As for each block-shape, several intra prediction modes are proposed, a specific signalling scheme is also proposed, and it provides significant decoded samples. Then these features that are extracted to select a predefined coding gains compared to state-of-the-art video coding technologies. Block based analysis of luma and chroma components are done here, considering its separate and combined approaches the computational complexity was one of the main problems encountered, that need to be rectified. Multiple modes for, predicting blocks are also introduced based on correlation methods, if feature extraction is used here with deep network the prediction become more efficient. Further research possibilities are there to reduce the complexity of the network structure by maintaining same prediction efficiency.

In [24], a revolutionary dual-network framework was proposed, that improved the performance of CNN-based super-resolution for compressed HD videos, particularly at

a low bit rate. An enhancement network is delegated prior to the SR network to lessen the effect of compression or to completely decrease the compression artifact effect. These two, i.e. the SR network and enhancement network, gradually optimize for different super-resolution and compression artifact reduction tasks. An enhanced geometric self-ensemble technique is adopted for further enhancement of the SR efficiency. In comparison with HEVC, it could achieve about 31.5% bit-rate savings (Table 1), especially for 4 K video compression, when needed to work with the proposed process, particularly in a super-resolution-based video coding operation, which demonstrates performance. The inter-frame similarities in both super-resolution networks and the enhancement network can be explored by work [25].

In [20], a learning-based approach to super resolution imaging using CNN-CR (convolutional neural network) was introduced. The goal was to mutually minimize the loss of restoration and the loss of regularization. CNN-CR preparation can be performed independently or jointly using CNN for super-resolution image resolution [27]. The network structure and training strategies are intended for this purpose. The objective or subjective assessment of the quality of the CR images is a matter of concern that must be checked in the future. The method can be tried to improvise intra prediction frame SR.

A new down-sampling based video coding scheme was proposed by Minmin Shen et al. in [26] where it encodes the intra-frames at their real resolution and down-samples them in the inter-frames before executing the compression technique. A new example-based SR is used on the decoder side to restore the resolution of down-sampled frames. Philipp Helle et al. [27] trained a neural network to perform the intra-image prediction method for block base video encoding. It requires several modes of prediction that cooperate to mitigate the impact of a loss feature during training. A sigmoid-function and l_1 -norm applied to the prediction residue in the DCT(discreet cosine transform) domain induce the loss function to represent the properties of the coding stages and residual quantization of the pruning of the resulting predictors in the hybrid video coding architecture, which helps to minimize the number of multiplications and make it less computational.

In [30], Cross-component Adaptation states that it is useful to train separately intra-prediction networks for the chroma channels, which can benefit from the luma component knowledge of the respective block and thus performing a trained cross-component prediction. This increases the benefit in luma BD-rate. With a higher prediction accuracy, the network will perform than expected. It is possible to lower the computational complexity. The wise use of reference frame will better reduce the efficiency of prediction, it can help to train the network to choose less artifact frame during training.

In [31], Zhang et al. concentrated on enhancing the compression effectiveness of HEVC intra-frame coding, for which they suggested a deep learning-based architecture. Using a new deep learning model in image super-resolution (SR), the idea brings out to train the CNN (convolutional neural network) model to particularly predict each CTU's residual information (coding tree unit) in a HEVC encoder [32–35]. Due to this, it is possible to achieve better CTU estimation and better reconstruction for the compression of CTUs. By skipping the non-linear mapping layer, computational complexity is reduced here, and the residual learning is incorporated to achieve better residue prediction for CTU encoding. The novel approach achieves 3.2% bitrate reduction with only 37% increase in encoding complexity in the mean of Bjøntegaard delta bit rate i.e., BD-BR. In CNN-based Up sampling for HEVC, block-based up/down sampling for video coding is performed by designing using in a better way when sampling. This will assist in improving performance (Figs. 6, 7, 8, 9).

Table 1 Comparative study of deep learning based intra prediction

Sl. No & Source references	Methods in NN and DL intra prediction	Future scope
1 [18]	Fully connected cross-component based intra-prediction	Train to improve prediction accuracy Training with different input, such as a reference with coding artefact, will increase prediction efficiency at low bitrates Reduction strategies such as pruning quantization and others can help reduce computational complexity
2 [20]	Context based deep learning methods are adopted It provides accurate prediction with average of 4.5%bit rate saving for 4 K sequence	Training separate models for chroma and luma components will probably boost the performance
3 [21]	Fully connected neural network structure with CNN for chroma intra-prediction (hybrid network)	Simple model and training smaller model can reduce complexity
4 [26]	CNN for up or down sampling in block level for video compression	Design a simple structure with less computational complexity with better prediction accuracy Design block based optimum sampling rate analysis algorithm (fast) with reference frame

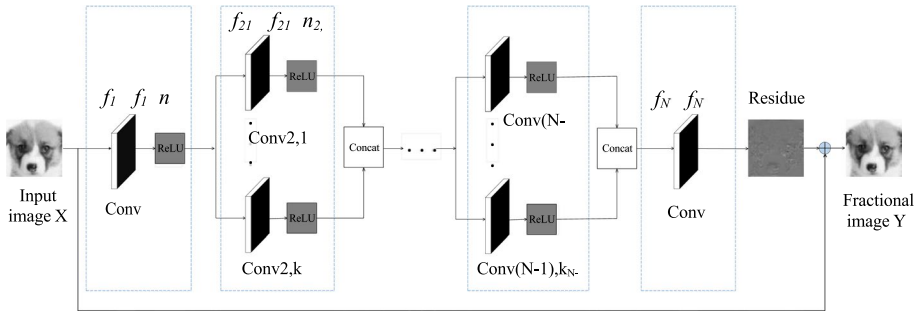


Fig. 6 A schematic representation showing FRCNN including convolutional modules and nonlinear blocks [33]

Fig. 7 CNN based arithmetic coding [29]

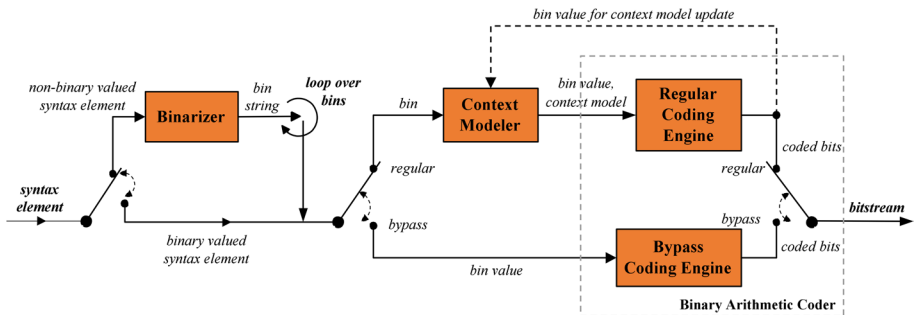
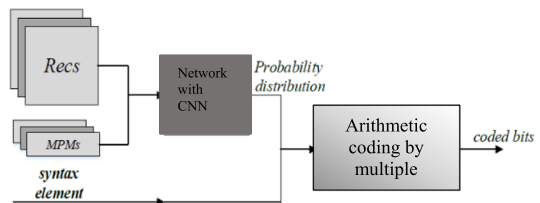


Fig. 8 CABAC method entropy encoding- a representation [43]

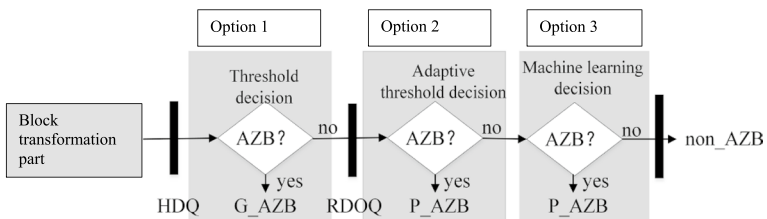


Fig. 9 Multistage three level AZB detection using NN [44]

3.2 Deep Learning Based Inter Prediction and Motion Compensation

Intra prediction defines the prediction of new frame using the previously encoded frame. It uses frame differencing concept to reduce Bit rate reduction with less computational complexity. Various methods used are sorted as CNN based motion compensation reduction, FRCNN (Faster Region based Convolutional Neural Network), FRMCNN, SVM (support-vector networks), ResNet based etc. and the recent techniques focus on better coding gain system. The overview of various methods is discussed here.

A system was proposed on CNN based motion compensation refinement that focus on refining motion compensation prediction by incorporating it with neighbouring reconstructed region, integrating this CNNMCR (Convolutional Neural Network-Based Motion Compensation Refinement) in high efficiency video coding is done and it advantages in its bit reduction. Research can be done to refine Chroma components using same technique in future. It can be applied in multi directional inter prediction under low delay and random-access configuration. In [37], Shuai Huo et al. proposed a motion compensation refinement model using CNN in video coding. Simple CNNMCR and CNNMCR, are the two schemes that is taken in consideration here, in that, CNNMCR uses spatial correlation for the betterment of the prediction accuracy. The two systems considered here are basic CNNMCR and CNNMCR, in that CNNMCR uses spatial correlation. To increase the precision of prediction. The simple CNNMCR method achieves an average 1.8% reduction in the BD rate, followed by HEVC when the results are tested after experimentation with the data provided. On the other hand, an average 2.3% BD-rate reduction is accomplished by the CNNMCR system. A 5.2% BD-rate reduction was achieved by integrating both with the OBMC technique. The CNN-based method for refining bi-directional interaction prediction under random-access configurations and low-delay B will operate on this principle.

FRCNN, i.e., a CNN-based approach for fractional-pixel MC to aggravate the efficiency of video coding i.e., in inter prediction step [33]. Instead of interpolation the novel approach introduces the fractional-pixel MC as a regression problem between frames. A CNN was proposed to solve the problem, regression. The proposed method introduces a CNN models to generate a relationship between integer-pixel values and fractional references, each model in that accounts for a proper and fractional MV. For bi-directional MC, the CNN has been trained properly. There are several problems to be further studied, such as how to build a fair network structure to achieve better compression efficiency and minimize complexity, and the mutual optimization of CNN for compression artifact reduction in FRCNN, which also reduces compression noise. Furthermore, the principle can be generalized to the simple idea of formulating fractional-pixel MC as an inter-picture regression problem in the case of integer pixel MC or even general MC without involving MV (Table 2).

The FRMCNN predicts the pixel value of the current frame to be coded from the integer pixel value of reference pixel. Fractional pixel reference regeneration convolutional neural network can be implemented for bi-directional and unidirectional motion compensation in video coding. The virtual reference frame that is created with the deep neural network, does not send the motion data directly but explicitly signalled for inter prediction method.

Apart from the available methods, such as the generation of interactive reference frames, Zhao et al. [35] uses the reconstructed bi-directional frames to provide a more efficient prediction of the to-be-coded frame. A CNN called VECNN (Virtual Reference

Table 2 Comparative study of deep learning based inter prediction and motion compensation

Sl. No & Source references	Methods in NN and DL based inter prediction	Future scope
1 [35]	VRCNN to enhance the training and quality of network Post processing of HEVC with CNN	Ease the network yet preserving its effectiveness in coding Processing P and B frames by extending VRCNN
2 [36]	CNN-based approach for fractional-pixel MC to increase the performance of inter-prediction video coding	Achieve improved compression efficiency and lower complexity Joint CNN optimization for the elimination of compression artifacts
3 [37]	A new CTU level inter-prediction approach with advanced deep learning based on the FRUC method is introduced The innovation lies in the fact that the virtual reference frame is tailored to the advanced deep learning FRUC process, so that better predictive output can be achieved without the cost of signaling any motion information	Methods can be introduced to improve coding gain
4 [40]	Non block based inter prediction with switchable motion prediction option improve the frame prediction accuracy	To incorporate methods to switch between different motion model
5 [41]	Fast coding decision for CU prediction in HEVC by 3 stage networks to reduce computational complexity	Artifact removal can be enhanced Deep learning techniques can be included in deblocking, bitrate saving etc

Frame Enhancement CNN) is designed to improve the consistency of VRF (Virtual Reference Frame), concentrate on alleviating the compression artefacts and improve the accuracy of VRF predictions. The development of the CTU level coding mode DVRF (Direct Virtual Reference Frame), which eliminates the extra ME procedure on VRF, would help to achieve a better trade-off between difficulty and coding efficiency. A fast CU algorithm based on Online SVM [36] to minimize HEVC complexity by training it online with a deep CNN, was designed to predict HEVC CU partitioning. The goal of this approach was to reduce the complexity of the HEVC inter-mode and to reduce the complexity of the estimation and the time of execution as per results. SVM- achieve time saving of 52.28% with 1.928% of bitrate reduction. Encoding time is reduced by 53.99%. It can be improved without affecting BR. To note as a point to develop is the degradation in bitrate and that can be reduced in future.

A CNNVPN (Convolutional Neural Network video prediction network) to generate [37] VRF (virtual reference frame) which is synthesized using previously coded frame to improve coding efficiency. Main VPN uses two sub-VPN architecture in cascade that predict the current video frame in same time instance, it is incorporated in HEVC and that helps to predict PV through rate distortion optimization (RDO) only. Modification is done in the network to predict advanced motion vector prediction and merge mode adaptively. It shows improved random access and low delay value when compared to HEVC reference software HM 1.6.6 version [38]. The research option here can be pointed as, extensively experimenting on the developed system can be done to improve system efficiency.

A three stage DL algorithm that is applied for prediction of CU size. The developed fast learning method reduce computational complexity [40]. Deep learning framework of tensor flow was built and embedded into the HM simulation to form a runtime strategy for performance improvement. In future artifact reduction of HEVC encoders, deep learning technology can be incorporated in deblocking and sample adaptive offset, to improve the objective quality and to achieve bit rate saving.

In [41], it states that a framework is created to identify I frames and other types of P frames with two types of information maps, CU size map (CSM) and CU preliminary map (CPM). A ResNet with two streams is attention based is proposed to capture CSM and CPM. Then LSTM will capture the relocated I frames temporal dynamics. Feature to be noted Bi prediction to identify I frames in a video using B frame. More fine info like TU transforms unit), PU (prediction unit) can be considered particularly for research.

Non block based inter prediction with switchable motion prediction option helped to increase the frame prediction but the challenge is the difficulty in switching between the different motion [40] models. Interpolation of a new DL based reference frame to reduce the prediction error was another advancement in this field. This could improve the coding performance and bit reduction of 4.5% from normal HEVC [39]. The next evolution was to send only differential data of motion vector [35] and recovering it by special residual coding network. This shows an advancement in coding performance. The new era of development is open with DL, CNN, RNN etc.

3.3 Deep Learning-Based Quantization and Entropy Coding for Video Coding

Quantization is rounding up into whole number and this is the lossy part of compression and its coded to save the bitrate. Context-adaptive binary arithmetic coding (CABAC) is used by HEVC, research is done on this by using FCNN (fully convolutional neural

network), CNN based CABAC, multilevel ABZ algorithm and some state of art methods are the recent development in this area, as discussed follows.

A strategy for HEVC i.e fast convolutional-neural-network based quantization. It explains FCNN based quantization for HEVC encoding. In this local artifact removal will be done. It is done by using a network trained on data i.e., extracted from an improved contrast gain control model. Local artifact visibility was predicted through convolutional neural network trained on data obtained from an improved contrast gain control model. The model is trained with natural images. Computational time analysis in real-time for predicting distortion visibility is the future work focused on this.

HEVC uses context-adaptive binary arithmetic coding (CABAC) for entropy coding [46] method. CABAC depends on manually designed binarization process and handcrafted context models, that can affect or restrict its compression efficiency [42]. A CNN based arithmetic code [29] is introduced to work on HEVC. Instead of binarization a probability distribution of 35 intra prediction model is videos estimated with adoption of multilevel arithmetic codec (Table 3). The future the work can be extended to other syntax of HEVC like quantization coefficient, motion vector etc. The convolutional neural network is used to execute the process of probability distribution than using handcrafted context models. The Fig. 7 shows CNN based arithmetic learning scheme. Simulation results prove the proposed arithmetic coding leads to as high as 9.9% bits saving compared with CABAC. Simon Wiedemann et al. [43] developed Deep CABAC for compressing neural network, the aim was to (a) Quantize each weight parameter by minimizing [45] a weighted rate distortion function. (b) It an increase compression rate by 7.4% on average. (c) Boost compression gain to presparsified deep NN. In future, to benchmark deep CABAC on non-scarified network and to apply distributed learning scenario where memory complexity is critical.

A new multi-stage ABZ detection algorithm was proposed by Haibing Yin et [44]. for RDOQ blocks with a strong trade-off between complexity and precision. If there are several stages, the procedure is as follows.

Stage 1- G-ABZ is split using the adaptive classification threshold based on HDQ.

Stage 2- The adaptive threshold model is built for PABZ detection based on QP.

Final stage-ML is used from TU level, coefficient level to context syntax element level, for the classification model utilizing higher level functionality.

A CNN based nonlinear intra-coding [48] was introduced with 2 main focus (1) to reduce the directionality in the residue of intra-prediction signal, (2) to improve efficiency by introducing a novel loss function. The method worked well for natural videos. A recent study [50] shows that intra-prediction with entropy coding shows better results than convolutional autoencoders and this method uses the border information of entropy coding in the tensor that showed better results than autoencoders. The convolutional autoencoders showed poor results in small image blocks because of not utilizing spatial redundancy in better way.

3.4 Deep Learning-Based Loop Filtering

Loop filters focus to alleviate the error between original video and decoded with the help of adaptive filtering techniques. Various novel methods like NN based loop filtering, RHCNN (Residual Highway Convolutional Neural Network) [47], content based loop

Table 3 Comparative study of deep learning BASED quantization and entropy coding

Sl. No & Source references	Methods in NN and DL based quantization and entropy coding for video coding	Future scope
1 [42]	CNN predicts the optimal transformation index probability distribution from the quantified coefficient blocks, and then uses the probability of binarizing the transform index using variable length instead of fixed length coding to increase entropy coding efficiency	CNN based work on the dominant syntax, to improve coding efficiency using quantized transform coefficients
2 [43]	Deep CABAC for compression that boosts compression rate, compression gain etc	To apply Deep CABAC on a non-specified network, where memory complexity is critical
3 [44]	Novel multi(three) stage ABZ detection algorithm for RDOQ blocks possessing better tradeoff between complexity and accuracy	Various classes of algorithm can be checked and accessing
4 [46]	HEVC's fast convolutional-neural-network-based quantization technique. Visibility of local objects was projected by a network trained on data extracted from an enhanced model of contrast gain control	Computational time analysis for clarity estimation of distortion in real-time videos
5 [45]	Arithmetic coding technique based on CNN for intra-prediction modes in HEVC	The other syntax components in HEVC, such as quantified coefficients and motion vectors, can be coded by CNN-based arithmetic coding

filtering, Dense net based multi frame in loop filtering, CResNet (compressed representation-ResNet), MACNN were in research.

RHCNN was introduced [55], that constitutes of two units, residual highway unit and convolutional layer [52, 53]. Residual highway is designed with some path that allow some free flow of data through some layers. There exists an identity skip connection from beginning till end by using CNN, this will not affect the deblocking filter and the SAO (sample adaptive filter). RHCNN is a high dimensional filter following DF and SAO to enhance resolution frames. The entire QP is divided into various bands or levels to have better performance. to lower the QP value early tracing and their weights are used in progressive manner. “Compression Artifacts Reduction by a Deep Convolutional Network” states that lossy compression introduces ringing and blocking artifacts and blurring formulates network to activate the different compression artefact. It uses a shallow network with feature training with “easy to hard” ideas followed.

In [49] CNN and image/video processing blend together and give tremendously good results and acquire more success in image/video processing. The purpose is content aware CNN based loop filtering for HEVC. Here in the proposed model the quantitative analysis in multiple dimensions will be done to make more efficient and worth for content aware CNN based loop filtering.

The CTU is considered as a separate region here, the Fig. 10 explains the working. Here the CNN model analyses the resolution of different region with different CNN model using a discriminative network. This is the working process of content aware multimodal filtering mechanism. The deep learning model helps to adaptively select the content-based region based on discriminative neural network. RDO (rate distortion optimisation) i.e., rate distortion optimisation is the key for analysing CTU level control. CNN based loop filter is implemented after the SAO(sample adaptive offset) in HEVC (Table 4). The performance of the total system increases by the content aware system.

Multi frame in loop filter (MIF) is proposed by Tianyi Li et al. [57] that enhanced visual quality of each frame by leveraging multiple adjacent frame [58]. MIF network uses the large database (HIF database) with encoding frames and its corresponding transform REF, reference frame selector is designed to identify frame the pool, MIF is developed or

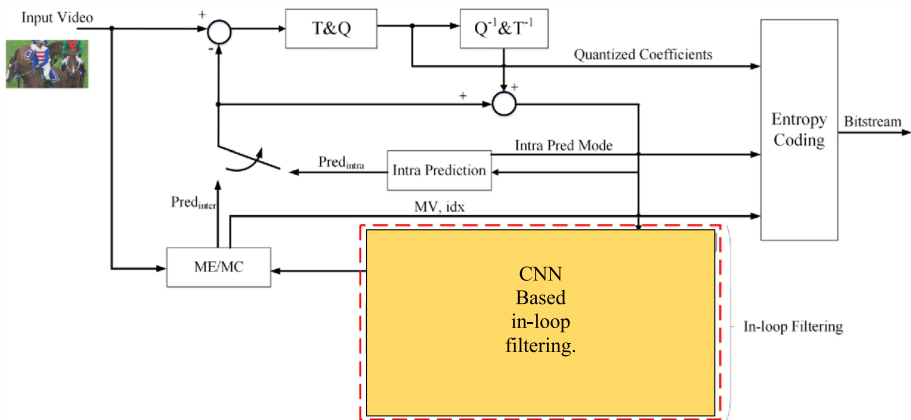


Fig. 10 A CNN based in-loop filtering within the hybrid coding framework of HEVC. Neural Network-Based Arithmetic Coding of Intraprediction Modes in HEVC [51]

Table 4 Comparative study of deep learning based loop filtering

Sl. No & Source references	Methods in NN and DL based loop filtering	Future scope
1 [55]	In-loop filtering in HEVC by very deep residual highway convolutional neural networks (RHCNN) consisting of multiple highway units of cascading, convolutional layers and shortcuts	During the preparation of the proposed RHCNN, which improves the output of the loop filter, it is worth using texture sophistication, motion forms of moving objects and cameras, motion numbers, etc
2 [52]	In-loop filtering based on Content-aware CNN for High Efficiency Video Coding (HEVC). Study of several aspects of the configuration of the theoretical CNN algorithm to make the model interpretable and suitable for CNN-based loop filtering	Coding performance and bit-rate reduction can be improvised with DL algorithms
3 [53]	Compact and effective network for smooth attenuation of multiple artifacts of compression. This shows superior results on both benchmark datasets and real-world usage cases than the state-of-the-art	SRCNN, notice that greater filter sizes often tend to increase efficiency
4 [57]	Multi frame in loop filter (MIF) with improved visual consistency of all frames by exploiting multiple adjacent frames with Dense Net	Compression artifacts such as residual frame, motion vector, prediction unit partitioning skip modes can be used to make the performance of loop filter better. To speed up MIF by accelerating DNN with some existing technology

DenseNet that uses the spatial info in encoded frame and the temporal info in the neighboring high-quality frame. High computational efficiency is high here. Block adaptive convolution layer is added to MIF Net for managing artifact removal because of CTU. Future scope of research noted are as follows, various details related to compression artifacts like residual frame, motion vector, prediction unit partitioning skip modes can be utilized to improve the performance of loop filter. Next is to speed up MIF by accelerating DNN (deep neural network) with some existing technology.

EDNN is proposed by Zhaoqing Pan et al. [59] to reduce the artifact in HEVC encoded video by loop filtering. It has a weighted normalization method, an efficient feature information fusion block and a precise loss function. This reduces loss function and provides a High-quality video. To reduce the limitation of current loop filter Dhanalakshmi et al. proposed a SHVC ie scalable extension of HEVC and CResNet (combined residual network) in loop filtering, CResNet [60] central layer info from spatial temporal domain to avoid visual artifact like blocking, ringing etc., the blocks are correctly collocated using deep CNN [61].

MACNN is introduced by Peng-Ren Lai et al. in [62], to replace in loop-filtering, it has 3 stage design fashion and different artifacts are removed in different stages and all building blocks can focus on artifact. Result shows its better than HEVC in loop filtering by 6.5% BD rate reduction. In future deeper the network can be created and can leverage the self-attention mechanism by simultaneously keeping a best tradeoff between performance gain and computational cost. Scope to research is noted as follows. Each network is trained separating according to different configuration with various QP and coding approach (Figs. 11, 12, 13).

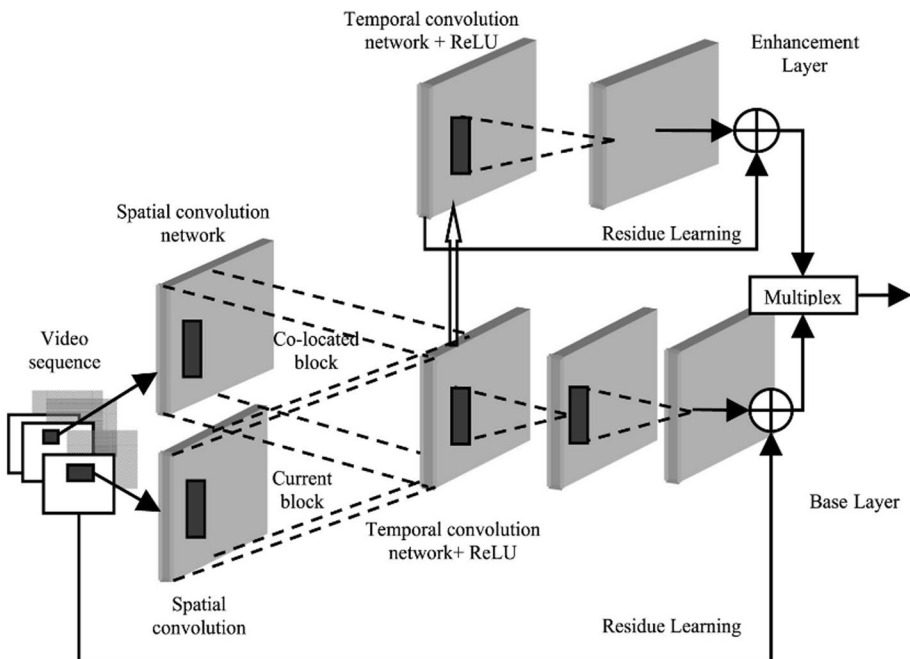


Fig. 11 The CResNet architecture [60]

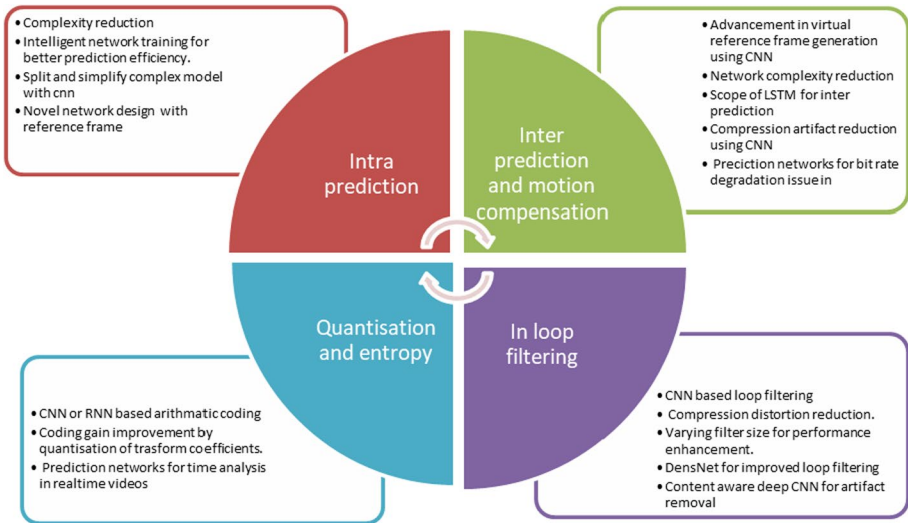


Fig. 12 Representation of research scope in compression in each area of video compression

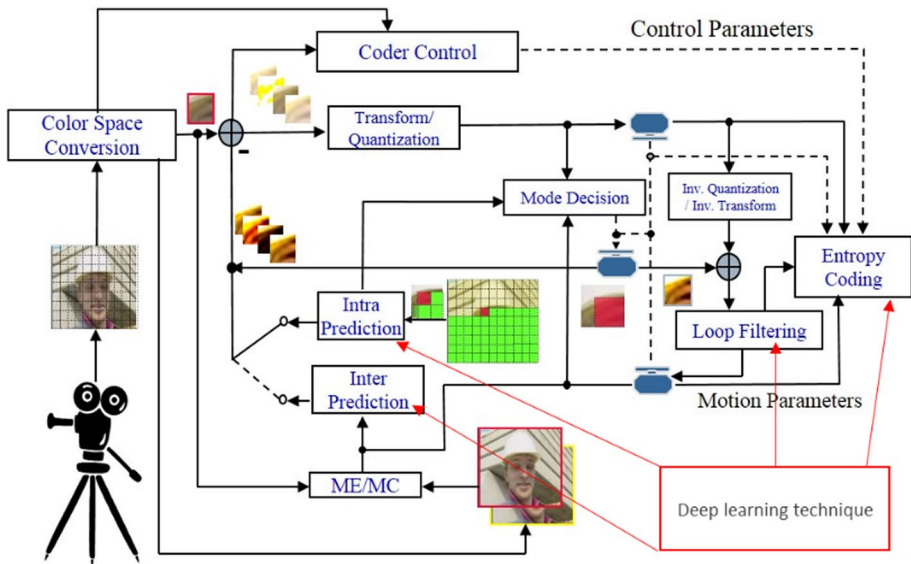


Fig. 13 The proposed architecture model of deep learning-based video compression

A 3 stage YUV post loop filtering was introduced for compression of YUV frames [54] with improved luminance and chrominance channel. This showed better results in recent video codecs. A CNN based loop filtering was introduced with two streams i.e., appearance-based filtering and coding distortion-based filtering [56]. This both are fused and pipelined with existing network that helped in its better performance.

4 Open Research Scope and Future Trends

Considering the evolution of deep learning based intra prediction, the advancements evolved can be noted as content based intra prediction, fully connected network CNN based chroma intra prediction, unique chroma and luma training for channel block level prediction etc. The scope of development in intra-prediction lies in the following ideas, training separate models for chroma and luma to boost the performance of intra-prediction, developing smaller models with DL to reduce complexity and to improved accuracy, prediction accuracy enhancement by training in DL based IP etc. The usage of reference frame with coding artifact can improve the quality of prediction. Design of fast algorithm for sampling rate decision and usage of reference frame for up sample are areas having future scope to work on. The area open for research for,

- Complexity reduction of intra prediction.
- Training network intelligently to attain better prediction efficiency
- Splitting and modeling simple network to get simplified network.
- A novel network design for intra prediction with optimal sampling rate analysis with a reference frame.

Intra prediction and motion compensation process have developed its way long with NN techniques from normal correlation related motion compensation to three stage DL based algorithms with ResNet. To improve the performance and accelerate training, CNN based fractional pixel MC (improve efficiency), CTU prediction with virtual reference frame (for high efficiency coding) deep FRCU based VRF etc. were introduced. Multiple areas still need advancement in inter-prediction and motion compensation are, network complexity, reference frame generation in both direction, artifact removal, bitrate degradation, coding gain and compression efficiency values. Improvisation in interpretation is proposed for,

- Simplification of network by maintaining coding efficiency
- An extended version of VRCNN for HEVC and for VCC.
- CNN based joint optimization for compressing artifact reduction in CNN based FRMC and to achieve better compression efficiency
- To improve coding gain in FRCU and to alleviate complexity.
- The degradation of bitrate in fast CU based prediction approach is open topic to research.
- An LSTM based network for intra prediction.

As a future scope, different class algorithm can be checked and accessed in quantization and entropy coding. There is development in the step of quantization and entropy coding with CNN, as fast CNN based quantization in HEVC predicted local visibility by network trained on data that are from improved vision of contrast gain control model, followed by CNN based arithmetic coding [45–50] and variable length coding that is CNN based to predict the probability distribution from quantized efficient block. Quantization and entropy need refinement by,

- Computation time analysis for predicting distribution in real time video
- CNN based arithmetic coding for other parameter like quantization coefficient and motion vector.

- To improve coding gain by quantizing the transform coefficient in video coding
- Deep learning-based classification algorithms can be tried to generate smart ABZ detection with n.

The area of exploration in DL based loop filtering focus on a very deep RHCNN loop filtering in HEVC with cascaded highway and CNN layers, followed by content aware CNN based loop filtering [46] for HEVC in multi dimension and CNN for different compression artifacts [63] etc. The scope of improvisation in loop filtering can be noted as.

- Focus on improving the training to get better to avoid compression distortion.
- Coding performance and bitrate reduction can be improved in content aware CNN based loop filtering
- Larger filter size in SRCNN can improve performance.
- Extra features related to compression artifact can be used to train the DenseNet to improve loop filter performance.
- For better performance, each network can be trained separately with different configurations with QP.
- Artifact removal is a main issue in loop filtering stage that can be done with content aware deep CNNs.

Here a model is proposed for smart video compression. The steps in video compression is upgraded with the advancement in deep learning. The predictions line intra prediction and inter prediction, encoding, quantization and loop filtering are enhanced by replacing the traditional model by various deep learning techniques based on the performance of each steps. Various research methods are proposed in this paper for each of these steps separately. The best out of it can be chosen and can design this smart video compression model. It has been observed that CNN/RNN perform better for inter prediction whereas LSTM and deep learning techniques which holds the previous memory and compare previous features can perform good in inter prediction as it depends on other frames for its prediction process. For encoding and loop filtering CNN, ResNet etc. can be used to enhance the performance of traditional techniques. The total model will be designed with deep learning techniques and combined to perform as a system. The system will perform smart than the traditional techniques.

4.1 Quality Enhancement

The video quality analysis can be categorized as subjective and objective analysis. Though objective analysis focus on quality measurement, in lossy operations like quantization (eg: RD cost optimization in HEVC), subjective analysis looks into the visual quality of the video. BD PSNR, SSIM (structural similarity index), VMAF (Video Multimethod Assessment Fusion), Bitrate, encoding time etc. are some of the objective analysis. The quality analysis with combined subjective and objective method should be a focus in the future deep learning-based video coding techniques that helps to give a better perspective of quality analysis.

4.2 Computational Complexity

HEVC has its advantage of saving bitrate of 50% compared to its predecessor, as it has replaced the macroblock with CTU of different sizes, but the computational complexity is high. Various algorithms were introduced to reduce the complexity in calculation, some are heuristic and some are ML/DL based. Scope to improve the algorithm are still open for research to reduce the time for encoding by reducing the number of calculations by novel methods.

4.3 Scalability

Video coding is a compression technique that is upgrading its features as per various needs. The update incorporating scalability add advantage in many areas for cost reduction. The device scalability and scalability of updated application to other video coding techniques are areas to improve. Scalability can be focused as an area of research as it provides an advancement without hardware replacement.

4.4 Application Agnostic

Video codec, when used in various applications, the parameters used and needs changes based on application, platform used etc. Some application demands quality rather than time while other demands in reduction of bitrate or lowering streaming time etc. The demand of each application will be different. The models developed can focus on flexibility to adapt to the application by syncing with the needs of that platform. This challenges in that can be considered as a scope to research to avoid environmental dependencies in various platforms.

5 Conclusion

In this paper, the advancement in video compression with artificial intelligence, especially with deep learning, neural network etc. are emphasized for each stage of video compression, along with the scope of enhancement. The impact of deep learning network and CNN in video compression is a detailed thing to discuss in current era because the advancement and efficiency improvement acquired and going to acquire from that is expected as tremendous. So, its significant study is needed in current situation as research are booming in the same. Some recomences of involving the technique of neural network for video processing points on content adaptivity i.e., present neural network compared to traditional methods. CNN and deep learning model can use both near and far pixel details instead traditional signal processing can utilize only neighboring pixel. Content analysis is an advantage in CNN. The video compression stage in each step has a scope to develop and its noted and scrutinized in order. Intra prediction mainly needs focus on chroma, luma component, block-based approaches that needs refinement in reducing computational complexity, bit rate saving, enhancing various deep learning algorithms for better block prediction etc. whereas in inter prediction and motion compensation the scope can be briefed in developing a simple network structure with reduced artifacts with VRF in bi-direction. Quantization and entropy coding the research scope is to create a DL based quantizer which handle multiple levels and a smart coder. A content aware DL based in loop filter with less complexity is the main area research. Deep learning can combine multiple features and texture for analysis, that is not available in traditional models. While combined with video

processing it shoes an excellent improvement by high efficiency prediction, compact representation, and improved coding gain. In future an end-to-end deep learning-based system is expected for video compression that can make a change in compression world.

Acknowledgements The authors would like to thank the anonymous referees for providing valuable suggestions which helped clarify the exposition of the material. The authors are greatly indebted to the anonymous reviewers whose thought-provoking and encouraging comments have motivated them to modify significantly and update the paper. They also like to express their gratitude to REVA University for extending research facilities to carry out this research.

Author Contributions HKJ Analysis of existing video coding techniques and it evolution. Extensive research on multiple intra and inter prediction techniques with deep learning methodology. MRK Content on Deep learning-based Quantization and entropy, Future trends and open research scope identification. AC Overall review of section wise contents in the paper. Content contribution on open research scope identification. MP Overall review of section wise contents in the paper. Content contribution on recent techniques and reviews on loop filtering and its research white spaces.

Funding The authors have not disclosed any funding.

Data Availability Enquiries about data availability should be directed to the authors.

Declarations

Competing interest The authors declare that they have no competing interests.

References

1. Ma, S., Zhang, X., Jia, C., Zhao, Z., Wang, S., & Wanga, S. (2019). Image and video compression with neural networks: A review. *IEEE Transaction on Circuits and System for Video Technology*, 8215(SEP-TEMBER 2018), 1–1.
2. Reader, C. (2002). History of video compression (Draft), document JVT-D068, Joint video team (JVT) of ISO/IEC MPEG & ITEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6).
3. Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9), 1098–1101.
4. Andrews, H., & Pratt, W. (1968). Fourier transform coding of image in *Proc. Hawaii Int. Conf. System Sciences*, pp. 677–679.
5. Pratt, W. K., Kane, J., & Andrews, H. C. (1969). Hadamard transform ima coding. *Proceedings of the IEEE*, 57(1), 58–68.
6. Ahmed, N., Natarajan, T., & Rao, K. R. (1974). Discrete cosine transform. *IEEE Transaction on Computers*, 100(1), 90–93.
7. Joy, H.K., & Kounte, M.R. (2019). An overview of traditional and recent trends in video processing, in *Proceedings of the 2nd International Conference on Smart Systems and Inventive Technology, ICSSIT 2019*. pp. 848–851.
8. Wiegand, T., Sullivan, G. J., Bjontegaard, G., & Luthra, A. (2003). Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7), 560–576.
9. Sullivan, G. J., Ohm, J., Han, W.-J., & Wiegand, T. (2012). Overview of the high efficiency video coding (HEVC) standard. *IEEE Transaction on Circuits and Systems for Video Technology*, 22(12), 1649–1668.
10. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
11. Dong, L., Yue, L., Jianping, L., Houqiang, L., & Feng, W. (2020). Deep learning-based video coding: A review and a case study. *ACM Computer Survey*, 53(1), 1–34.
12. Kumar, B. S., & Shree, V. U. (2020). An end-to-end video compression using deep neural netowrk. *JAC: A Journal of Composition Theory*, XIII(XI), 209–215.
13. Agustsson, E., Tschannen, M., Mentzer, F., Timofte, R., & Van Gool, L. (2018). Extreme learned image compression with GANs, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2587–2590.

14. Zhang, X., Ma, S., Wang, S., Zhang, X., Sun, H., & Gao, W. (2017). A joint compression scheme of video feature descriptors and visual content. *IEEE Transaction on Image Processing*, 26(2), 633–647.
15. Li, Y., Jia, C., Zhang, X., Wang, S., Ma, S., & Gao, W. (2018). Joint rate-distortion optimization for simultaneous texture and deep feature compression of facial images, in *IEEE International Conference on Multimedia Big Data (BigMM)*, pp. 334–341.
16. Li, X., & Gong, N. (2020). Run-time deep learning enhanced fast coding unit decision for high efficiency video coding. *Journal of Circuits, Systems and Computers*, 29(3), 1–19.
17. Srivastava, N., Mansimov, E., & Salakhudinov, R. (2015). Unsupervised learning of video representations using LSTMs,” in *International conference on machine learning*, pp. 843–852.
18. Li, J., Li, B., Xu, J., Xiong, R., & Gao, W. (2018). Fully connected network- based intra prediction for image coding, *IEEE Transaction on Image Processing*.
19. Joy, H.K., Kounte, M.R., & Joy, A.K. (2020). Deep learning approach in intra -prediction of high efficiency video coding, in *2020 International conference on smart technologies in computing, electrical and electronics (ICSTCEE), Bengaluru*, pp. 134–138, doi: <https://doi.org/10.1109/ICSTCEE49637.2020.9277189>
20. Li, Y., Li, L., Li, Z., Yang, J., Xu, N., Liu, D., & Li, H. (2018). A hybrid neural network for chroma intra prediction, in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 1797–1801.
21. Pfaff, J., Helle, P., Maniry, D., Kaltenstadler, S., Stallenberger, B., Merkle, P., Siekmann, M., Schwarz, H., Marpe, D., & Wiegand, T. (2018). Intra prediction modes based on neural networks, in *JVET-J0037. ISO/IEC JTC/SC 29/WG 11*, April, pp. 1–14.
22. Li, Y., Liu, D., Li, H., Li, L., Wu, F., Zhang, H., & Yang, H. (2017). Convolutional neural network-based block up-sampling for intra frame coding. *IEEE Transaction on Circuits and Systems for Video Technology*. <https://doi.org/10.1109/TCSVT.2017.2727682>
23. Hu, Y., Yang, W., Xia, S., Cheng, W.H., & Liu, J. (2018). Enhanced intra prediction with recurrent neural network in video coding, in *IEEE Data Compression Conference (DCC)*, pp. 413–413.
24. Feng, L., Zhang, X., Zhang, X., Wang, S., Wang, R., & Ma, S. (2018) A dual-network based super-resolution for compressed high-definition video, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
25. Huang, H., Schioppa, I., & Munteanu, A. (2020). Frame-wise CNN-based filtering for intra-frame quality enhancement of HEVC videos. *IEEE Transaction on Circuits and System Video Technology*, 8215(c), 1–1.
26. Shen, M., Xue, P., & Wang, C. (2011). Down-sampling based video coding using super-resolution technique. *IEEE Transaction on Circuits and Systems for Video Technology*, 21(6), 755–765.
27. Pfaff, J., Helle, P., Maniry, D., Kaltenstadler, S., Samek, W., Schwarz, H., Marpe, D., & Wiegand, T. (2018). Neural network based intra prediction for video coding, in *Applications of Digital Image Processing XLI*, vol. 10752. International Society for Optics and Photonics, 2018, p. 1075213.
28. Zhang, Z.T., Yeh, C.H., Kang, L.W., & Lin, M.H. (2017). Efficient CTU- based intra frame coding for HEVC based on deep learning, in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, pp. 661–664
29. Ma, C., Liu, D., Peng, X., Li, L., & Wu, F. (2020). Convolutional neural network-based arithmetic coding for HEVC intra-predicted residues. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(7), 1901–1916.
30. Meyer, M., Wiesner, J., Schneider, J., & Rohlfing, C. (2019). Convolutional neural networks for video intra prediction using cross-component adaptation, in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom*, 2019, pp. 1607–1611, doi: <https://doi.org/10.1109/ICASSP.2019.8682846>.
31. Liu, Z., Yu, X., Gao, Y., Chen, S., Ji, X., & Wang, D. (2016). CU partition mode decision for HEVC hardwired intra encoder using convolution neural network. *IEEE Transaction on Image Processing*, 25(11), 5088–5103.
32. Song, N., Liu, Z., Ji, X., & Wang, D. (2017) CNN oriented fast PU mode decision for HEVC hardwired intra encoder, in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 239–243.
33. Yan, N., Liu, D., Li, H., Li, B., Li, L., & Wu, F. (2018). Convolutional neural network-based fractional-pixel motion compensation. *IEEE Transaction on Circuits and Systems for Video Technology*. <https://doi.org/10.1109/TCSVT.2018.2816932>
34. Zhao, L., Wang, S., Zhang, X., Wang, S., Ma, S., & Gao, W. (2018). Enhanced CTU-level inter prediction with deep frame rate up-conversion for high efficiency video coding,” in *25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 206–210.
35. Alexandre, D., Hang, H.-M., Peng, W.-H., & Domański, M. (2021). Deep video compression for inter-frame coding. *IEEE International Conference on Image Processing (ICIP), 2021*, 2124–2128. <https://doi.org/10.1109/ICIP42928.2021.9506275>

36. Bouaafia, S., Khemiri, R., Sayadi, F. E., & Atri, M. (2020). Fast CU partition-based machine learning approach for reducing HEVC complexity. *Journal of Real-Time Image Processing*, 17(1), 185–196.
37. Lee, J. K., Kim, N., Cho, S., & Kang, J. W. (2020). Deep video prediction network based inter-frame coding in HEVC. *IEEE Access*, 8, 95906–95917.
38. Lee, J.K., Kim, N., Cho, S., & Kang, J.W. (2018). Enhanced motion-compensated video coding with deep virtual reference frame generation, *submitted to IEEE Transaction on Image Processing*.
39. Guo, Y., Liu, Z., Chen, Z., & Liu, S. (2020). Deep inter coding with interpolated reference frame for hierarchical coding structure. *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020, 302–305. <https://doi.org/10.1109/VCIP49819.2020.9301769>
40. Li, K., Bare, B., & Yan, B. (2017). An efficient deep convolutional neural networks model for compressed image deblocking, in *International Conference on Multimedia and Expo (ICME)*, 2017, pp. 1320–1325.
41. He, P., Li, H., Wang, H., Wang, S., Jiang, X., & Zhang, R. (2020). Frame-wise detection of double HEVC compression by learning deep spatiotemporal representations in compression domain. *IEEE Transaction on Multimediations*, 9210(65), 1–14.
42. Brand, F., Seiler, J., & Kaup, A. (2021). Switchable motion models for non-block-based inter prediction in learning-based video coding. *Picture Coding Symposium (PCS)*, 2021, 1–5. <https://doi.org/10.1109/PCS50896.2021.9477475>
43. Wiedemann, S., et al. (2019). DeepCABAC: Context-adaptive binary arithmetic coding for deep neural network compression,” *arXiv*, pp. 2–5.
44. Yin, H., Yang, H., Huang, X., Wang, H., & Yan, C. (2019). Multi-stage all-zero block detection for HEVC coding using machine learning. *Journal of Visual Communication and Image Representative*, 73(September), 102945.
45. Wang, M., Fang, X., Tan, S., Zhang, X., & Zhang, L. (2020). Low complexity quantization in high efficiency video coding. *IEEE Access*, 8, 145159–145170.
46. Puri, S., Lasserre, S., & Le Callet, P. (2017). CNN-based transform index prediction in multiple transforms framework to assist entropy coding, in *Signal Processing Conference (EUSIPCO), European*, pp. 798–802.
47. Y. Zhang, T. Shen, X. Ji, Y. Zhang, R. Xiong, and Q. Dai, “Residual Highway Convolutional Neural Networks for in-loop Filtering in HEVC,” *IEEE Trans. on Image Processing*, 2018.
48. Yuan, Z., Liu, H., Mukherjee, D., Adsumilli, B., & Wang, Y. (2021). Block-based learned image coding with convolutional autoencoder and intra-prediction aided entropy coding. *Picture Coding Symposium (PCS)*, 2021, 1–5. <https://doi.org/10.1109/PCS50896.2021.9477503>
49. Dong, C., Deng, Y., Change Loy, C., & Tang, X. (2015). Compression artifacts reduction by a deep convolutional network, in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 576–584.
50. Yang, K., Liu, D., & Wu, F. (2020). Deep learning-based nonlinear transform for HEVC intra coding. *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020, 387–390. <https://doi.org/10.1109/VCIP49819.2020.9301790>
51. Jia, C., Wang, S., Zhang, X., Liu, J., Pu, S., Wang, S., & Ma, S. (2019). Content-aware convolutional neural network for in-loop filtering in high efficiency video coding. *IEEE Trans. on Image Processing*. <https://doi.org/10.1109/TIP.2019.2896489>
52. Song, X., Yao, J., Zhou, L., Wang, L., Wu, X., Xie, D., & Pu, S. (2018). A practical convolutional neural network as loop filter for intra frame, *arXiv preprint arXiv:1805.06121*.
53. Park, W.-S., & Kim, M. (2016). CNN-based in-loop filtering for coding efficiency improvement, in *Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, pp. 1–5.
54. Cui, K., Koyuncu, A. B., Boev, A., Alshina, E., & Steinbach, E. (2021). Convolutional neural network-based post-filtering for compressed YUV420 images and video. *Picture Coding Symposium (PCS)*, 2021, 1–5. <https://doi.org/10.1109/PCS50896.2021.9477486>
55. Zhu, L., Zhang, Y., Wang, S., Yuan, H., Kwong, S., & Ip, H.H.-S. (2018). Con- volutional neural network-based synthesized view quality enhancement for 3d video coding. *IEEE Transactions on Image Processing*, 27(11), 5365–5377.
56. Yue, J., Gao, Y., Li, S., & Jia, M. (2020). A mixed appearance-based and coding distortion-based CNN fusion approach for in-loop filtering in video coding. *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020, 487–490. <https://doi.org/10.1109/VCIP49819.2020.9301895>
57. Li, T., Xu, M., Zhu, C., Yang, R., Wang, Z., & Guan, Z. (2019). A deep learning approach for multi-frame in-loop filter of HEVC. *IEEE Transactions on Image Processing*, 28(11), 5663–5678.
58. Joy, H. K., & Kounte, M. R. (2022). Decision algorithm for intra prediction in high-efficiency video coding (HEVC). *Journal of Southwest Jiaotong University*, 57(5), 180–193. <https://doi.org/10.35741/issn.0258-2724.57.5.15>

59. Pan, Z., Yi, X., Zhang, Y., Jeon, B., & Kwong, S. (2020). Efficient in-loop filtering based on enhanced deep convolutional neural networks for HEVC. *IEEE Transactions on Image Processing*, 29, 5352–5366.
60. Dhanalakshmi, A., & Nagarajan, G. (2020). Combined spatial temporal based In-loop filter for scalable extension of HEVC. *ICT Express*, 6(4), 306–311.
61. Lai, P.R., & Wang, J.S. (2020). Multi-stage attention convolutional neural networks for HEVC in-loop filtering,” in *Proceedings - 2020 IEEE International Conference on Artificial Intelligents Circuits System AICAS 2020*, pp. 173–177.
62. Cavigelli, L., Hager, P. & Benini, L. (2017). CAS-CNN: A deep convolutional neural network for image compression artifact suppression, in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 752–759.
63. Joy, H. K., & Kounte, M. R. (2020). A comprehensive review of traditional video processing. *Advances in Science, Technology and Engineering System Journal*, 5(6), 274–279.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Helen K. Joy received the Bachelor of Engineering in Electronics and Communication Engineering from National college of engineering, Anna University, Chennai, India, and Master of engineering from Satyabhama University, Chennai, India, in 2013, and currently working toward the Ph.D. degree at REVA University. The research interests include video compression, video processing, deep learning.



Manjunath R. Kounte received his Bachelors in Electronics and Communication Engineering from PDA College of Engineering, Kalburgi, India and master's in computer network engineering from AMC College of Engineering, Bengaluru and PhD in Machine Vision from JAIN University, Bengaluru in 2017. He is currently serving as Associate Professor and Head, Department of Electronics and Computer Engineering in REVA University, Bengaluru. He has over 12+ years of Teaching, Research and Administrative experience. He has published around 35 peer reviewed publications including 12 Journals. He is also heading Machine Learning and Blockchain Lab, REVA University, India. His research interests include Machine learning, Video processing, IoT Edge Computing, Blockchain technologies etc.



Arunkumar Chandrasekhar is currently an Associate Professor in Department of Sensors and Biomedical Technology, Vellore Institute of Technology, Vellore, India and he is a recipient of VIT SEED Grant research project. He worked as a postdoctoral researcher at Nanomaterials and Systems Lab, South Korea. He received his Ph.D. in Mechatronics Engineering from Jeju National University, Jeju, South Korea where he was a scholarship recipient from the Korean Government Scholarship Program. He is interested in wearable triboelectric nanogenerator, battery free electronic devices, energy storage devices, MEMS and self-powered devices.



Manoranjan Paul received the Ph.D. degree from Monash University, Australia, in 2005. He was a Postdoctoral Research Fellow with the University of New South Wales, Monash University, and Nanyang Technological University. He is currently a Full Professor, the Director of the Computer Vision Laboratory, and the Leader of the Machine Vision and Digital Health (MaViDH) Research Group, Charles Sturt University, Australia. He has published around 200 peer reviewed publications, including 72 journals. He has supervised 15 Ph.D. students to completion. He was an Invited Keynote Speaker in IEEE DICTA 2017 and 2013, CWCN 2017, WoWMoM 2014, and ICCIT 2010. His major research interests include video coding, image processing, digital health, wine technology, machine learning, EEG signal processing, eye tracking, and computer vision. He was awarded the ICT Researcher of the Year 2017 by Australian Computer Society. He obtained more than \$3.6 million competitive external grant, including the Australian Research Council (ARC) Discovery grants and the Australia-China grant. He was the General Chair of PSIVT 2019 and the Program

Chair of PSIVT 2017 and DICTA 2018. He is currently an Associate Editor of three top ranked journals such as the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and the EURASIP Journal on Advances in Signal Processing