International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 11, No. 7, July, 2024, pp. 1989-1995
Journal homepage: http://iieta.org/journals/mmep

Deep Convolutional Neural Network Driven Interpolation Filter for High Efficiency Video Coding

Helen K. Joy[1*], Manjunath R. Kounte[2]

[1] Department of Computer Science, CHRIST Deemed to be University, Bengaluru 560029, India
[2] School of ECE, REVA University, Bengaluru 560064, India

Corresponding Author Email: helenjoy88@gmail.com

https://doi.org/10.18280/mmep.110730

Received: 10 February 2024
Revised: 3 May 2024
Accepted: 15 May 2024
Available online: 31 July 2024

Keywords:
High Efficiency Video Coding, Super-Resolution Convolutional Neural Network, motion vectors, deep Convolutional Neural Network

ABSTRACT

Research in video coding has gained significant importance in recent years, driven by the increasing demand for multimedia transmission. High Efficiency Video Coding (HEVC) has emerged as a prominent standard in this field. Interpolation is a crucial aspect of HEVC, particularly when using fixed half-pel interpolation filters derived from traditional signal processing techniques. In recent times, there has been an exploration of interpolation filters that are based on Convolutional Neural Networks (CNNs). Conventional signal processing techniques are used in traditional HEVC methods to employ fixed half-pel interpolation filters. Recent advancements have delved into the application of Convolutional Neural Networks (CNNs) to enhance interpolation performance. Our proposed method utilises a sophisticated CNN architecture specifically crafted to extract valuable features from low-resolution image patches and accurately predict high-resolution images. The network consists of multiple layers of CNN blocks, which utilise 1×1 and 3×3 convolutional kernels to enable efficient and thorough feature extraction through parallel processing. This architecture improves computational efficiency and greatly enhances prediction accuracy The suggested interpolation filter shows a 2.38% enhancement in bitrate savings, as evaluated by the BD-rate metric, specifically in the low delay P configuration. This highlights the potential of deep learning techniques in improving video coding efficiency.

1. INTRODUCTION

In order to achieve the highest possible level of compression, High Efficiency Video Coding (HEVC) utilises a hybrid method in block-based codecs. This strategy partitions each video frame into blocks. The term "hybrid" [1] refers to the mixture of different tactics that are used to minimise duplicate information within video sequences. This is especially noticeable in video sequences that consist of consecutive frames that have a lot of similarities. In order to reduce the incidence of temporal redundancy, it is necessary to search through past frames in order to identify the image segment that is most similar to each block in the current frame. This segment acts as an estimate for the content of the current block. A residual signal is formed by the difference between the pixels in the current block [2] and this prediction. This residual signal contains a considerable amount of data that is significantly decreased in comparison [3] to the original image, which contributes to effective compression. Furthermore, High-Efficiency Video Coding (HEVC) employs intra-frame or spatial prediction in order to minimise redundancy inside the same frame. Both temporal and spatial redundancies are simultaneously reduced through the utilisation of this hybrid strategy, which results in a significant improvement in the compression efficiency of HEVC.

In addition, High-End Video Coding (HEVC) [4] incorporates in-loop filters such as the de-blocking filter (DB) and the sample adaptive offset (SAO)as mentioned in Figure 1. Image restoration is improved and the ringing effect is reduced thanks to the SAO, while the de-blocking filter helps to reduce the amount of blocking artefacts that occur between image blocks. HEVC was the first to implement SAO, which significantly improves the quality of the video.

With the introduction of the Super-Resolution Convolutional Neural Network (SRCNN) [5], the field of intelligent interpolation experienced its first significant breakthrough. This neural network displayed significant advancements in comparison to traditional methods that were not based on learning, such as bicubic interpolation and sparse coding. The utilisation of this increased super-resolution approach makes it possible to improve the image quality of larger photos, even when the source image has dimensions that are lower. Within the realm of HEVC inter-coding [6-10], researchers have investigated network topologies that are similar to SRCNNs in order to improve frame interpolation. The General HEVC inter prediction coding tree unit can be represented as in Figure 2. Their objective is to enhance the quality of interpolated frames by utilising deep learning models.

The motion-compensated prediction (MCP) approach is highlighted as a crucial step in this investigation, which digs into the dynamic relationship [11] that exists between frames within effective compression techniques. Through the process of identifying the best matching block in previously reconstructed reference frames for the current block, MCP is able to minimise discrepancies, also known as residuals, which are subsequently transferred to the decoder side. Motion vectors (MVs) [12] are able to capture the positional relationship that exists between the present block and its matching reference block. This allows for the acquisition of information regarding block displacements [13-15].

In order to effectively portray continuous object motion, fractional-pel precision motion vectors are needed. These vectors allow for a finer depiction of motion, which is essential for maintaining accuracy. In light of this, it is essential to interpolate the reference frame in order to accurately show the complex continuum of motion. Among the first attempts made in this field is the IPCNN project, which was carried out by the Harbin Institute of Technology [16]. This project utilised CNNs for intra prediction for HEVC, and it was successful in reducing the bit rate by 0.70 percent. In a similar vein, Intra Prediction utilising Fully Connected Network (IPFCN), which was created by Peking University and Microsoft Research Asia, was able to achieve a reduction of 1.1% in luma bitrates and a reduction of 1.6% in chroma bitrates, despite large increases in encoding and decoding times [17, 18].

The findings of this study offer a novel convolutional model architecture that was developed for the purpose of training interpolation filter sets. This architecture features numerous layers and branches, which together provide a comprehensive and effective representation of the data. Sharing layers is incorporated into the design, which makes it possible to have a unified training process for the interpolation filter set. This is done with the intention of improving both performance and efficiency [19].

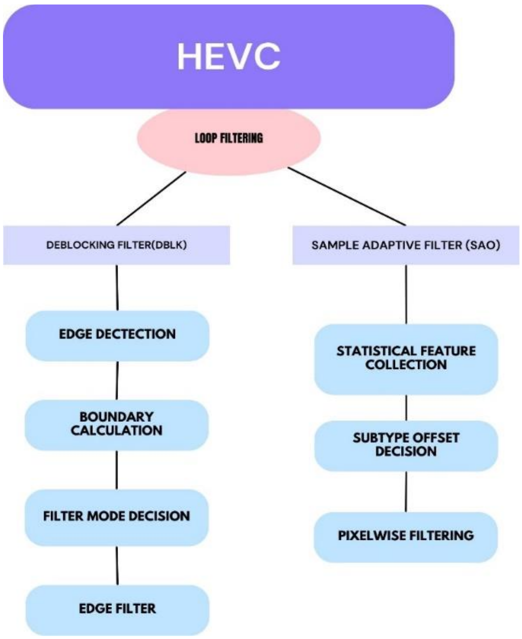
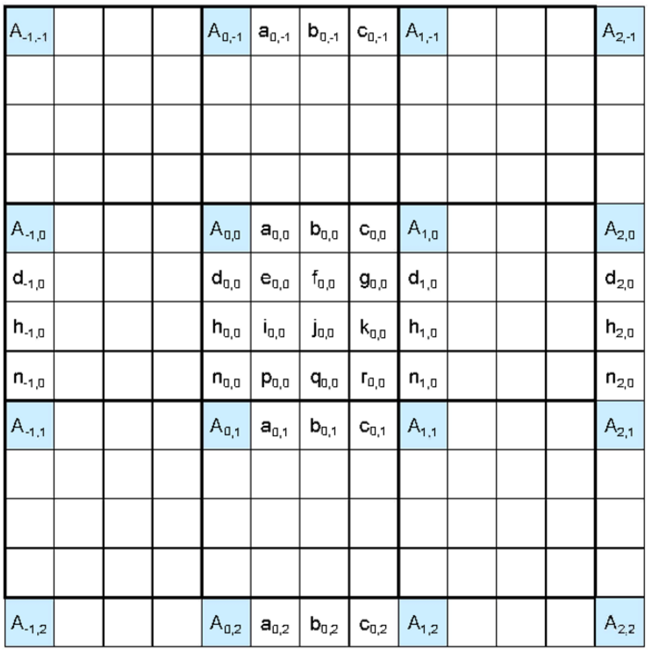


Figure 1. In loop filtering steps in HEVC decoder and encoder

The paper follows this organizational structure: the document is as follows: In the second section, a condensed explanation of the suggested model structure is provided. Section III provides an overview of the approach that was presented, which includes the newly developed network architecture. In the fourth section, the results and findings of the experiment are illustrated. The conclusions that were drawn from these findings are summarised in Section V, which is the final section of the report.



$A_{-1,-1}$				$A_{0,-1}$	$a_{0,-1}$	$b_{0,-1}$	$c_{0,-1}$	$A_{1,-1}$				$A_{2,-1}$
$A_{-1,0}$				$A_{0,0}$	$a_{0,0}$	$b_{0,0}$	$c_{0,0}$	$A_{1,0}$				$A_{2,0}$
$d_{-1,0}$				$d_{0,0}$	$e_{0,0}$	$f_{0,0}$	$g_{0,0}$	$d_{1,0}$				$d_{2,0}$
$h_{-1,0}$				$h_{0,0}$	$i_{0,0}$	$j_{0,0}$	$k_{0,0}$	$h_{1,0}$				$h_{2,0}$
$n_{-1,0}$				$n_{0,0}$	$p_{0,0}$	$q_{0,0}$	$r_{0,0}$	$n_{1,0}$				$n_{2,0}$
$A_{-1,1}$				$A_{0,1}$	$a_{0,1}$	$b_{0,1}$	$c_{0,1}$	$A_{1,1}$				$A_{2,1}$
$A_{-1,2}$				$A_{0,2}$	$a_{0,2}$	$b_{0,2}$	$c_{0,2}$	$A_{1,2}$				$A_{2,2}$

Figure 2. HEVC inter prediction CTU structure for video compression techniques

2. MOTIVATION FOR PROPOSED MODEL

HEVC, a commonly utilised standard for video compression, utilises a hybrid block-based coding technique to divide each video frame into blocks in order to enhance compression efficiency. This approach effectively minimises both the repetition of information within each frame and the duplication of information across frames through the use of spatial and temporal predictions. In traditional HEVC, an interpolation filter is used to estimate intermediate pixel values. This involves using an 8-tap symmetric DCT-based filter for interpolating half-samples and bilinear interpolation for quarter-pel pixels.

Nevertheless, the predetermined parameters of the DCT-based filter may not consistently yield the best results for the various attributes of distinct pixel blocks, resulting in less than optimal interpolation results. The need for improvement in interpolation accuracy and flexibility has sparked interest in exploring alternative methods.

A significant advancement in intelligent interpolation was achieved through the use of Convolutional Neural Networks (CNNs). The Super-Resolution Convolutional Neural Network (SRCNN) [20] played a pivotal role in enhancing image quality compared to conventional techniques. Building upon the SRCNN framework, we present a customised shared model specifically designed for HEVC interpolation. Through the utilisation of advanced deep learning techniques, the proposed model seeks to improve the interpolation process in HEVC. This will result in more precise predictions and improved handling of various pixel block characteristics.

In the field of HEVC inter-coding, researchers have delved into SRCNN-like network architectures to enhance frame interpolation. Their goal is to use deep learning models to improve the quality of interpolated frames. Research writing often involves a crucial step called motion-compensated prediction (MCP). This step focuses on finding the most suitable matching block in previously reconstructed reference frames for the current block. The goal is to minimise any discrepancies, known as residuals, that are sent to the decoder side. The motion vectors with fractional-pel precision [21] reveal the positional connection between the current block and its reference block, offering valuable information about block displacements [22-25]. The use of fractional-pel precision motion vectors is crucial in accurately capturing and depicting the smooth and continuous motion of objects. To achieve this, it is necessary to interpolate the reference frame in order to faithfully portray the intricate details of the motion.

The Harbin Institute of Technology conducted a study on intra prediction for HEVC, using CNNs in their IPCNN model. Their research resulted in a significant reduction in bit rate. Similarly, the Intra Prediction using Fully Connected Network (IPFCN), developed by Peking University and Microsoft Research Asia, showcased impressive reductions in bitrate for both luma and chroma components. However, it is worth noting that this came at the expense of longer encoding and decoding times [26-29].

Our study presents an innovative convolutional model architecture specifically developed for training interpolation filter sets [30-34]. This architecture includes multiple layers and branches to effectively represent data in a comprehensive and efficient manner. The architecture incorporates shared layers, allowing for a cohesive training process for the interpolation filter set, with the goal of improving both performance and efficiency [35]. This method combines bilinear interpolation techniques with adjacent half-pel pixels to enhance precision and versatility.

Our model showcases remarkable improvements in bitrate savings and peak signal-to-noise ratio (PSNR) through the use of a streamlined approach with CNN blocks and shared layers [36]. This highlights the immense potential of deep learning techniques in pushing the boundaries of video coding standards. The effectiveness of this innovative approach is thoroughly assessed through extensive experiments, demonstrating encouraging enhancements compared to conventional methods. This study emphasises the potential of using deep learning-driven interpolation filters to overcome the limitations of traditional DCT-based filters and improve the overall performance of HEVC.

3. INTRODUCTION OF THE DEEP CNN BASED INTERPOLATION FILTER

3.1 Details of the dataset

The dataset utilised in this study comprises of static images derived from a dynamic video, consolidated into a single file for convenient analysis. The dataset offers the convenience of segmenting images into various sizes, such as 64×64 pixels, 32×32 pixels, and even smaller dimensions if required. The labels for these segmented images are neatly organised in a directory called "pkl." Every label is associated with a particular Coding Tree Unit (CTU), which is an image file measuring 64×64 pixels. Each 64×64 CTU is associated with a Python list that consists of 16 items, which correspond to the labels for the CTU. In this setup, every 64×64 CTU is split into 16 smaller blocks, each measuring 16×16 pixels. Each block is assigned its own label.

When the images are divided into 64×64 CTUs, the dataset becomes quite large in practice. The training dataset alone consists of around 110,000 images, which offers a substantial sample size for model training. In addition, the validation dataset consists of approximately 40,000 images, which allows for a thorough evaluation of the model's performance and ability to generalise. This comprehensive dataset enables a meticulous training and validation process, essential for the development and improvement of deep learning models for tasks like image interpolation in video coding. With the help of this extensive and organised dataset, the model can grasp complex patterns and greatly enhance its predictive accuracy.

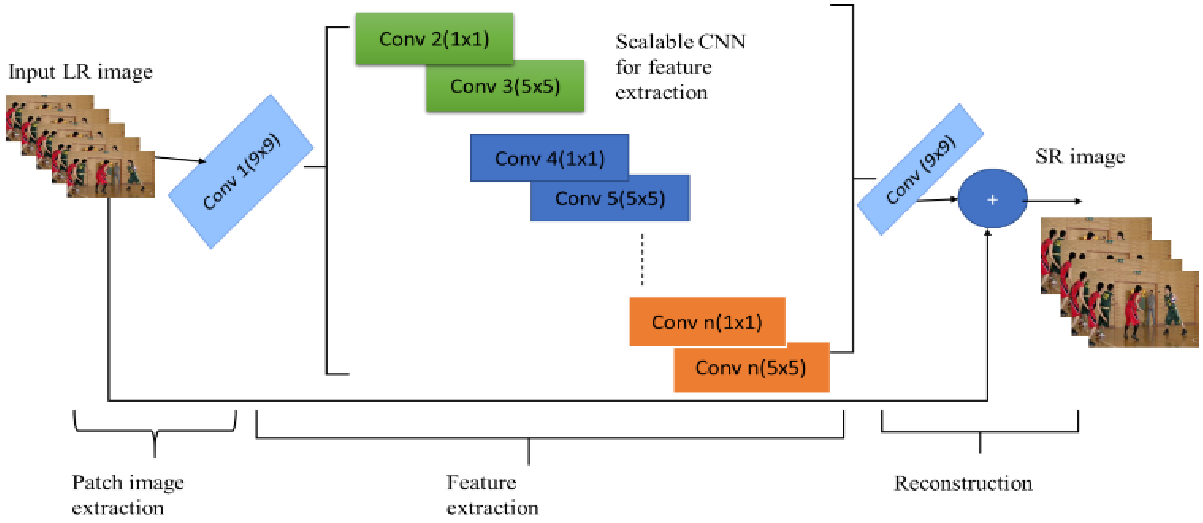


Figure 3. Deep CNN based inter prediction model with n- layers of CNN network to extract low and high frequency features to produce high resolution output

3.2 Modelling of network

The proposed method incorporates three convolution layers in its architecture as in Figure 3. In the initial layer, the focus is on patch extraction and representation, where the features are extracted from the low-resolution image. In this study, we consider the dimensions of W1 to be 9×9×64, indicating a tensor of size 9 in the first dimension, 9 in the second dimension, and 64 in the third dimension. The first phase can be represented as:

$$(I * K)(x, y) = \sum_{i=-4}^{n} \sum_{j=-4}^{n} I(x-i, y-j) \cdot K(i, j) \tag{1}$$

where, $(I*K)$ (x, y) represents the output value at position (x, y) after convolution (n=4); $I(x\text{-}i, y\text{-}j)$ is the value of the input signal at position $(x{-}i, y{-}j)$; $K(i, j)$is the value of the 9×9 kernel at position (i, j); the summation is performed over all positions (i, j) within the 9×9 kernel.

The second stage of convolution can be evaluated by:

$$(I * K)(x, y) = I(x, y) \cdot K \tag{2}$$

where, $(I*K)$ (x, y) represents the output value at position (x, y) after convolution; $I(x, y)$ is the value of the input signal at position (x, y); K is the value of the 1×1 kernel.

The third stage can be represented by:

$$(I * K)(x, y) = \sum_{i=-2}^{n} \sum_{j=-2}^{n} I(x-i, y-j) \cdot K(i, j) \tag{3}$$

where, $(I*K)(x, y)$ represents the output value at position (x, y) after convolution; $I(x{-}i, y{-}j)$ is the value of the input signal at position $(x{-}i, y{-}j)$; $K(i,j)$ is the value of the 5×5 kernel at position (i,j); The summation is performed over all positions (i,j) within the 5×5 kernel.

This repeated process extracts the features from the video frame that is needed for reconstruction. The summation of the output of this and the previous input gives the reconstructed frame Additionally, B1 is represented as a vector with 64 dimensions. The second layer can be conceptualised as a non-linear mapping mechanism that transforms the low-resolution image features into high-resolution image features. In this context, W2 refers to a tensor with dimensions 1×1×32, indicating that it has a single element along the first and second dimensions, and 32 elements along the third dimension. B2, on the other hand, is a vector with a dimensionality of 32, meaning it consists of 32 scalar values. In the proposed research, a third layer is incorporated into the model architecture. This layer, denoted as W3, has dimensions of 5×5×1. This W2and W3 gets cascaded for n layers that extracts the features. Its purpose is to facilitate the reconstruction of the high-resolution image using the high-resolution features obtained from previous layers. The model is utilised in this study, albeit with slight modifications in the preparation of input and output labels.

3.3 Test, train, validation and loss function

In this phase of the project, all three CNN models are trained using the same methodology. Stochastic gradient descent, coupled with backpropagation, is employed to optimize the loss function. Evaluating the model is a critical aspect of system development. To maximize effectiveness, the test, training, and validation sets are designed to be entirely independent of each other. For models aimed at prediction, the mean squared error (MSE) is an appropriate metric for assessing model quality. Training a CNN involves adjusting the parameters defined for the SRCNN to minimize the loss function over the training set, thereby improving the CNN's accuracy. Let Y represent the output of the proposed approach and I the input labels. The mean squared error is used as the loss function, where N denotes the total number of training data points.

$$L = \frac{1}{N} \sum_{i=1}^{N} [E|(I - Y)^2|] \tag{4}$$

4. EXPERIMENTAL RESULTS

For this study, the model undergoes evaluation through a training process spanning 10 epochs. Subsequently, the trained model is employed for validation purposes. The training process utilises the ReLU (Rectified Linear Unit) activation function to incorporate non-linearity and capture intricate patterns within the data. During training, the ADAM optimizer is utilised for its adaptive learning rate and efficient convergence. Following the training phase, the model demonstrates an impressive accuracy of 90.2% on the validation set, showcasing its ability to accurately classify or predict the desired outcomes. After conducting the necessary model training and validation, the evaluation process is centred around assessing the BD-bitrate by analysing different resolution video frames. BD-bitrate is a metric that quantifies the difference in bitrate between various coding methods or configurations. It is used to evaluate the coding efficiency.

The evaluation is carried out in two different contexts: within the HEVC (High-Efficiency Video Coding) baseline and by utilising the existing intraprediction method. HEVC is a popular video coding standard known for its impressive compression efficiency. Utilising spatial redundancies within a frame, intraprediction enhances coding efficiency. Through an analysis of the BD-bitrate, we can gain insights into the trained model's efficiency and effectiveness when compared to the HEVC baseline and other intraprediction methods. This evaluation provides a holistic view of the model's performance in terms of reducing bitrate and compressing videos as in Figure 4.

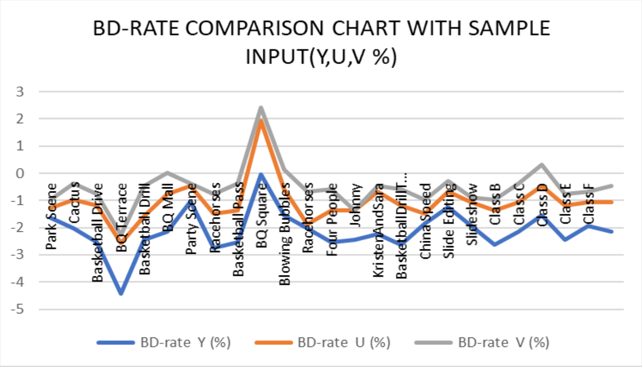


Figure 4. BD-bit rate comparison chart with the test input

The ReLU activation function and the ADAM optimizer are being utilised in this scenario in order to evaluate a model that has been trained on a dataset for a total of ten epochs. The model is able to obtain an accuracy of 90.2% on validation data, which demonstrates that it is effective in predicting the outputs that are wanted. Following the training phase, we assess the effectiveness of the model in terms of video compression by contrasting the BD-bitrates inside the HEVC baseline with the intraprediction approaches that are currently in use.

The first thing that we do is examine the BD-rate for each of the five classes over a period of ten epochs by making use of YUV data and four quantization parameters (QPs): 22, 27, 32, and 34. The fractional variation in bitrate that exists between various encoding techniques is denoted by the abbreviation BD-rate. A BD-rate that is lower suggests that the coding efficiency is higher. The model demonstrates a reduction in luminance bitrate of 2.14%, which indicates a more efficient representation without affecting the quality of the image quality. There is also a drop of 1.08% and 0.478% in the chromaticity components, which indicates that the encoding efficiency has been increased.

The second thing that we do is evaluate BD-rate in comparison to other methods such as VDSR and SRCNN. This table provides a comprehensive comparison of the bitrate reduction obtained by the proposed method in comparison to the techniques that are being discussed. Graphs illustrate the reduction in bitrate for the luminance component (Y), demonstrating that the suggested method is superior to VDSR and SRCNN while still preserving quality.

In order to improve the performance of video compression, these results reveal that the proposed method is effective in lowering the bitrate without compromising the quality of the video as in Figure 5. As a result of surpassing previously established methods, it demonstrates potential for optimising video coding in order to enhance the efficiency of transmission and storage noted in Table 1.

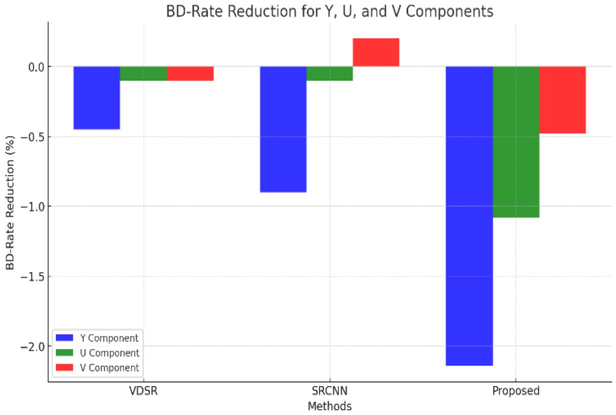


Figure 5. BD-bit rate reduction for Y, U, V components

Table 1. Comparison of BD -bit rate of various methods over proposed method

Classes	Sequences	BD-Rate [%] VDSR			BD-Rate [%] SRCNN			BD-Rate [%] Proposed		
		Y (%)	U (%)	V (%)	Y (%)	U (%)	V (%)	Y (%)	U (%)	V (%)
Class C	Basketball Drill	-0.9	-0.7	-0.3	-1.2	-0.6	0.2	-2.14	-0.78	0.022
	BQMall	-0.3	-0.3	0.2	-0.9	0.2	0.7	-1.04	-0.48	-0.378
	PartyScene	-0.1	-0.2	-0.3	0.2	0.5	0.3	-2.74	-1.48	-0.778
	RaceHorses	-0.4	0.4	0.4	-1.5	-0.5	-0.1	-2.54	-1.38	-0.378
Class D	BasketballPass	-1.2	-0.2	-1.5	-1.3	-0.4	0.3	-0.04	1.92	2.422
	BQSquare	0.3	1.1	0.5	1.2	2.9	3.1	-1.54	-0.58	0.122
	BlowingBubbles	-0.2	-0.1	-0.7	-0.3	0.4	0.8	-2.04	-1.88	-0.678
	RaceHorses	-0.7	-0.9	0.3	-0.8	-0.9	0	-2.54	-1.38	-0.578
Class E 720p	FourPeople	-0.6	0.1	-0.2	-1.3	-0.4	0.1	-2.44	-1.38	-1.378
	Johnny	-0.7	-0.5	0.1	-1.2	-0.4	-0.7	-2.24	-0.68	-0.478
	KmsenAndSara	-0.4	0.3	0.3	-1	0.3	0.2	-2.64	-1.18	-0.578
Class F	BasketballDrillText	-0.4	-0.1	0.1	-1.4	-0.2	0.1	-1.84	-1.48	-0.978
	ChinaSpeed	-0.6	-0.4	-0.2	-0.6	-0.5	-0.3	-1.24	-0.68	-0.278
	Slideñditing	0	0	0.1	0	0.3	0.4	-1.94	-1.08	-0.878
	ShdeShow	-0.2	-0.1	0.3	-0.7	-0.1	-0.2	-2.64	-1.38	-0.978
Average	Class C	-0.4	-0.2	0	-1.4	-0.4	-0.3	-2.14	-1.08	-0.378
	Class D	-0.5	0	-0.4	-0.9	-0.1	0.3	-1.54	-0.48	0.322
	Class E	-0.6	-0.1	0.1	-0.3	0.5	1	-2.44	-1.18	-0.778
	Class F	-0.3	-0.1	0.1	-1.2	-0.2	-0.1	-1.94	-1.08	-0.678
Overall	All	-0.45	-0.1	-0.1	-0.9	-0.1	0.2	-2.14	-1.08	-0.478

5. CONCLUSIONS

The purpose of this research was to present a unique interpolation filter that is based on deep learning. The filter was designed to improve inter prediction within the framework of High Efficiency Video Coding (HEVC). Compared to the conventional fixed half-pel interpolation filters, the method that we have proposed presents a significant improvement thanks to the utilisation of Convolutional Neural Networks (CNNs). In order to extract patches and features from low-resolution photos, the network that was created makes use of these images. This makes it easier to make accurate predictions about high-resolution images. By employing this method, the network is educated to produce high-resolution outputs from the patches that are provided, thereby effectively reconstructing an entire frame within the HEVC format. In order to effectively manage the complexity of the computations involved, our system make use of a simplified architecture that incorporates many CNN layers for the purpose of feature extraction. Additionally, in order to successfully capture a wide variety of information, each layer combines two distinct kernels, namely 1×1 and 3×3. The performance of our network was subjected to a comprehensive evaluation that utilised a variety of inputs, and the findings suggested a significant increment in performance. To be more specific, our approach was successful in achieving a significant reduction in bitrate of 2.38 percent, as determined by the BD-rate metric, particularly in the condition of low delay P configuration. It is clear that deep learning-driven interpolation filters have the potential to advance video coding approaches, as evidenced by the demonstrated gain in bitrate reductions. By incorporating this forward-thinking strategy into the High-Efficiency Video Coding (HEVC) framework, our approach paves the way for more effective transmission of multimedia material, which is in response to the growing need for high-quality video content in the digital era. Work that will be done in the future will concentrate on further optimising the network architecture and investigating whether or not it is applicable to various settings and coding standards.

ACKNOWLEDGMENT

The authors would like to express their gratitude to CHRIST University and REVA University for extending research facilities to carry out this research.

REFERENCES

[1] Ma, S., Zhang, X., Jia, C., Zhao, Z., Wang, S., Wang, S. (2019). Image and video compression with neural networks: A review. IEEE Transactions on Circuits and Systems for Video Technology, 30(6): 1683-1698. https://doi.org/10.1109/TCSVT.2019.2910119

[2] Joy, H.K., Kounte, M.R., Chandrasekhar, A., Paul, M. (2023). Deep learning based video compression techniques with future research issues. Wireless Personal Communications, 131(4): 2599-2625. https://doi.org/10.1007/s11277-023-10558-2

[3] Huffman, D.A. (1952). A method for the construction of minimum-redundancy codes. Proceedings of the IRE, 40(9): 1098-1101. https://doi.org/10.1109/JRPROC.1952.273898

[4] Pandey, S.S., Singh, M.P., Pandey, V. (2015). Image transformation and compression using Fourier transformation. International Journal of Current Engineering and Technology, 5(2): 1178-1182.

[5] Joy, H.K., Kounte, M.R. (2022). Deep CNN based video compression with lung ultrasound sample. Journal of Applied Science and Engineering, 26(3): 313-321. https://doi.org/10.6180/jase.202303_26(3).0002

[6] Joy, H.K., Das, S.L. (2013). A novel approach for biomedical web image super resolution. In 2013 International Conference on Circuits, Power and Computing Technologies (ICCPCT), Nagercoil, India, pp. 876-879. https://doi.org/10.1109/ICCPCT.2013.6528858

[7] Qu, Z., Liu, W.J., Cui, L.Z., Yang, X.H. (2024). Video frame interpolation via spatial multi-scale modelling. IET Computer Vision, 18(4): 458-472. https://doi.org/10.1049/cvi2.12281

[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. https://doi.org/10.1109/TCSVT.2003.815165

[9] Sullivan, G.J., Ohm, J.R., Han, W.J., Wiegand, T. (2012). Overview of the high efficiency video coding (HEVC) standard. IEEE Transactions on Circuits and Systems for Video Technology, 22(12): 1649-1668. https://doi.org/10.1109/TCSVT.2012.2221191

[10] Joy, H.K., Kounte, M.R., Sujatha, B.K. (2022). Design and implementation of deep depth decision algorithm for complexity reduction in high efficiency video coding (HEVC). International Journal of Advanced Computer Science and Applications, 13(1).

[11] Ma, C., Liu, D., Peng, X., Li, L., Wu, F. (2019). Convolutional neural network-based arithmetic coding for HEVC intra-predicted residues. IEEE Transactions on Circuits and Systems for Video Technology, 30(7): 1901-1916. https://doi.org/10.1109/TCSVT.2019.2927027

[12] 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 Transactions on Image Processing, 25(11): 5088-5103. https://doi.org/10.1109/TIP.2016.2601264

[13] Song, N., Liu, Z., Ji, X., Wang, D. (2017). CNN oriented fast PU mode decision for HEVC hardwired intra encoder. In 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, Canada, pp. 239-243. https://doi.org/10.1109/GlobalSIP.2017.8308640

[14] Yan, N., Liu, D., Li, H., Li, B., Li, L., Wu, F. (2018). Convolutional neural network-based fractional-pixel motion compensation. IEEE Transactions on Circuits and Systems for Video Technology, 29(3): 840-853. https://doi.org/10.1109/TCSVT.2018.2816932

[15] Joy, H.K., Kounte, M.R. (2022). Deep CNN based video compression with lung ultrasound sample. Journal of Applied Science and Engineering, 26(3): 313-321. https://doi.org/10.6180/jase.202303_26(3).0002

[16] 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 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, pp. 206-210. https://doi.org/10.1109/ICIP.2018.8451465

[17] Joy, H.K., Kounte, M.R. (2020). A comprehensive review of traditional video processing. Advances in Science, Technology and Engineering Systems Journal, 5(6): 272-279. https://doi.org/10.25046/aj050633

[18] Hu, Y., Jung, C., Qin, Q., Han, J., Liu, Y., Li, M. (2024). HDVC: Deep video compression with hyperprior-based entropy coding. IEEE Access, 12: 17541-17551. https://doi.org/10.1109/ACCESS.2024.3350643

[19] 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: 185-196. https://doi.org/10.1007/s11554-019-00936-0

[20] 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. https://doi.org/10.1109/ACCESS.2020.2993566

[21] Zhao, L., Wang, S., Zhang, X., Wang, S., Ma, S., Gao, W. (2019). Enhanced motion-compensated video coding with deep virtual reference frame generation. IEEE Transactions on Image Processing, 28(10): 4832-4844. https://doi.org/10.1109/TIP.2019.2913545

[22] Wang, S.W., Yang, X.H., Feng, Z.Q., Sun, J.D., Liu, J. (2024). EMCFN: Edge-based multi-scale cross fusion network for video frame interpolation. Journal of Visual Communication and Image Representation, 103: 104226. https://doi.org/10.1016/j.jvcir.2024.104226

[23] Li, K., Bare, B., Yan, B. (2017). An efficient deep convolutional neural networks model for compressed image deblocking. In 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, pp. 1320-1325. https://doi.org/10.1109/ICME.2017.8019416

[24] He, P., Li, H., Wang, H., Wang, S., Jiang, X., Zhang, R. (2020). Frame-wise detection of double HEVC compression by learning deep spatio-temporal representations in compression domain. IEEE

Transactions on Multimedia, 23: 3179-3192. https://doi.org/10.1109/TMM.2020.3021234

[25] Ioffe, S., Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning, Lille France, pp. 448-456.

[26] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, pp. 105-114. https://doi.org/10.1109/CVPR.2017.19

[27] Salimans, T., Kingma, D.P. (2016). Weight normalization: A simple reparameterization to accelerate training of deep neural networks. Advances in Neural Information Processing Systems, pp. 901-909.

[28] Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K. (2017). Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492-1500.

[29] 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. https://doi.org/10.1109/TIP.2020.2982534

[30] Ding, D., Kong, L., Wang, W., Zhu, F. (2021). A progressive CNN in-loop filtering approach for inter frame coding. Signal Processing: Image Communication, 94: 116201. https://doi.org/10.1016/j.image.2021.116201

[31] Zhu, L., Zhang, Y., Wang, S., Yuan, H., Kwong, S., Ip, H.H.S. (2018). Convolutional neural network-based synthesized view quality enhancement for 3D video coding. IEEE Transactions on Image Processing, 27(11): 5365-5377. https://doi.org/10.1109/TIP.2018.2858022

[32] Li, K., Bare, B., Yan, B. (2017). An efficient deep convolutional neural networks model for compressed image deblocking. In 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, pp. 1320-1325. https://doi.org/10.1109/ICME.2017.8019416

[33] Brand, F., Seiler, J., Kaup, A. (2021). Switchable motion models for non-block-based inter prediction in learning-based video coding. In 2021 Picture Coding Symposium (PCS), Bristol, United Kingdom, pp. 1-5. https://doi.org/10.1109/PCS50896.2021.9477475

[34] 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. https://doi.org/10.1109/TIP.2019.2921877

[35] Yang, K., Liu, D., Wu, F. (2020). Deep learning-based nonlinear transform for HEVC intra coding. In 2020 IEEE International Conference on Visual Communications and Image Processing (VCIP), Macau, China, pp. 387-390. https://doi.org/10.1109/VCIP49819.2020.9301790

[36] 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. In 2021 Picture Coding Symposium (PCS), Bristol, United Kingdom, pp. 1-5. https://doi.org/10.1109/PCS50896.2021.9477503

NOMENCLATURE

$I(x, y)$	value of the input signal at position (x, y)
$K(i, j)$	value of the kernel at position (i, j)
K	value of the 1×1 kernel
Y	output of the proposed approach
I	input labels
N	total number of training data points