Innovative Integration of Residual Networks for Enhanced In-loop Filtering in VVC Using Deep Convolutional Neural Networks

M.K.I. Ibraheem[1], *A.V. Dvorkovich*[1], *A.M.S. Al-Temimi*[2]
[1]*Department of Multimedia Technologies and Telecommunications,*
Phystech-School of Radio Engineering and Computer Technologies (FRKT),
Moscow Institute of Physics and Technology (MIPT), 9, Institutsky Lane, 141701, Dolgoprudny, Russia;
[2] *Iraqi National Data Center, General Secretariat for the Council of Ministers*
Karada Maryam, 10069 , Baghdad, Iraq

Abstract

This paper explores the integration of Residual Networks (ResNets) into the in-loop filtering (ILF) process of the Versatile Video Coding (VVC) standard, aiming to enhance video compression efficiency and video quality through the application of Deep Convolutional Neural Networks (DCNNs). The study introduces a novel architecture, the Residual Deep Convolutional Neural Network (RDCNN), designed to replace conventional VVC in-loop filtering modules, including Deblocking Filter (DBF), Sample Adaptive Offset (SAO), and Adaptive Loop Filter (ALF). By leveraging the Rate Distortion Optimization (RDO) technique, the RDCNN model is applied to every coding unit (CU) to optimize the balance between video quality and bitrate. The proposed methodology involves offline training with specific parameters using the TensorFlow-GPU platform, followed by feature extraction and prediction of optimal filtering decisions for each video frame during the encoding process. The results demonstrate the effectiveness of the proposed RDCNN in significantly reducing the bitrate while maintaining high visual quality, outperforming existing methods in terms of compression efficiency and peak signal-to-noise ratio (PSNR) values across various video files (YUV color space). Specifically, the RDCNN achieved a YUV PSNR of 41.2 dB and a BD-rate reduction of – 2.43 % for the Y component, – 6.96 % for the U component, and – 9.43 % for the V component. These results underscore the potential of deep learning techniques, particularly ResNets, in addressing the complexities of video compression and enhancing the VVC standard. The evaluation across various YUV video files, including Stefan_cif, Soccer, Mobile, Harbour, Crew, and Bus, revealed consistently higher average YUV PSNR values compared to both VTM 22.2 and other related methods. This indicates not only improved compression efficiency but also enhanced visual quality, crucial for diverse video processing tasks.

Key words: Deep Learning, Residual Deep Convolutional Neural Network, Versatile Video Coding, Video Compression, VTM.

Citation: Ibraheem MKI, Dvorkovich AV, Al-Temimi AMS. Innovative Integration of Residual Networks for Enhanced In-loop Filtering in VVC Using Deep Convolutional Neural Networks. Computer Optics 2025; 49 (4): 692-701. DOI: 10.18287/2412-6179-CO-1572.

Introduction

The implementation of lossy video compression often introduces undesired compression artifacts, which degrade visual quality during video decompression. To address this, in-loop filtering has emerged as a critical technique during the encoding phase, enhancing video quality by reducing artifacts such as blocking and ringing. In-loop filters applied at the encoder level can significantly improve motion estimation and motion compensation, ultimately boosting the overall quality of the video output.

Traditional filtering methods, such as the Deblocking Filter (DBF), Sample Adaptive Offset (SAO), and Adaptive Loop Filter (ALF), have proven effective in reducing these artifacts, but they are limited by the non-stationarity of real-world video sequences. This has led to the exploration of deep learning techniques, which have demonstrated exceptional performance in tasks like noise reduction, artifact removal, and image enhancement.

In this paper, we propose a novel approach that leverages a deep convolutional neural network (CNN) architecture to enhance in-loop filtering within the Versatile Video Coding (VVC) standard. By incorporating deep learning, we aim to overcome the limitations of traditional handcrafted filters and improve both compression efficiency and visual quality.

The rest of the paper is organized as follows: Section 2 reviews related work, focusing on traditional and deep learning-based filtering methods. Section 3 presents the proposed methodology, while Section 4 details the neural network architecture. Section 5 discusses the experimental results, and Section 6 provides the conclusions.

1. Related work

In the realm of multimedia applications, the challenge of maintaining optimal video quality is an enduring concern, particularly due to the rapid advancements in various domains such as video gaming, computer vision, and video streaming. The inherent complexities associated with modern video content necessitate the development of robust compression techniques, as lossy video compression frequently introduces artifacts that adversely affect visual quality. Traditional in-loop filtering methods have been instrumental in addressing the challenges posed by lossy compression and the emergence of compression artifacts, thereby enhancing the visual experience during video decompression. Among these methods, the Deblocking Filter (DBF), Sample Adaptive Offset (SAO), and Adaptive Loop Filter (ALF) have been extensively implemented [1, 2, 3, 4]. The DBF effectively mitigates block border artifacts that arise from lossy coding, while the quantization process can induce loss of high-frequency components, resulting in ringing artifacts and boundary distortions. In contrast, the SAO method alleviates these ringing effects by introducing adaptive offsets to sample progressions, whereas the ALF employs a Wiener-based adaptive filtering approach to minimize the mean squared error between the original and decoded samples. Despite the efficacy of these handcrafted filters, which are grounded in established signal processing principles, their effectiveness is often constrained by the non-stationarity inherent in real-world video sequences.

1.1. Advances in Deep Learning for Video Compression

In recent years, deep learning methodologies have gained prominence, demonstrating remarkable effectiveness in addressing these challenges through advanced techniques such as noise reduction, resolution enhancement, and artifact elimination [5, 6]. For instance, Dong et al. [7] utilized a four-layer convolutional neural network (CNN) to significantly reduce compression artifacts in JPEG-encoded images. Similarly, Dai et al. [8] introduced the Variable-Filter-Size Residue-Learning Convolutional Neural Network (VRCNN) for post-processing within the HEVC standard, reporting an impressive 5% average bitrate reduction in comparison to the HEVC baseline. Lin et al. [9] further advanced artifact elimination through the implementation of a deeper network architecture and the incorporation of partitioning information. Other notable contributions to this field include the work of D. Ma, F. Zhang, and D. Bull [10], who proposed the Multi-Level Feature Review Residual Dense Blocks Network (MFRNet) for in-loop filtering (ILF) and post-processing (PP) tasks, thereby enhancing codec efficiency through the innovative use of residual dense blocks. Moreover, Chen et al. [11] developed a dense residual CNN (DRN) specifically designed for Versatile Video Coding (VVC), focusing on the improvement of coding performance and the minimization of artifacts through the implementation of dense shortcuts, residual learning, and bottleneck layers. A CNN-LSTM (Long Short-Term Memory) network hybrid deep learning model was presented by the authors S. Bouaafia, R. Khemiri, F. E. Sayadi, M. Atri, and N. Liouane [12] for the purpose of predicting HEVC CU inter-mode partitions. At the same time that it successfully improves rate-distortion (RD) performance, this model, which was trained on a huge HEVC inter-mode database, also reduces complexity. On a more particular level, Bouaafia, Khemiri, Sayadi, and Atri [13] proposed two machine learning-based rapid CU partition methods with the intention of reducing the complexity of encoding in inter-mode HEVC. The utilization of the deep CNN and online Support Vector Machine (SVM) methods results in a significant reduction in the complexity and duration of the encoding process. In addition, Bouaafia, Khemiri, Maraoui, and Sayadi [14] have made contributions by introducing a CNN-LSTM learning method to simplify HEVC. These individuals have made additional contributions. Through the utilization of this technique, the complexity of the encoding process is diminished, while the effects on the Bit Error Rate (BER) and Bjøntegaard Delta-Peak Signal-to-Noise Ratio (BD-PSNR) are decreased. In addition, Amna, Imen, Ezahra, and Mohamed [15] published a method for future video coding (FVC) that makes use of a deep CNN model. This method is known as rapid quad-tree (QT) partitioning. Both the intra-mode encoding time and the bitrate escalation performance have seen significant improvements as a result of this strategy. The effectiveness of the Quadtree-Binary tree (QTBT) block partition module in FVC can be improved through the utilization of this technique. Hsu, Lu, Hsieh, and Wang [16] presented a solution for HEVC Intra Frame Coding that is based on a deep convolutional neural network and is extremely effective. When contrasted with Simple Convolutional Neural Network (S-CNN), this method demonstrates that the encoding procedures are completed more quickly. In addition, the in-loop filtering method that was proposed by Pan, Yi, Zhang, Jeon, and Kwong [17] for HEVC by utilizing Enhanced Deep Convolutional Neural Networks (EDCNN) is also very efficient. By utilizing this method, the PSNR as well as the Rejection Detection (RD) are both greatly improved. Additionally, the WSE-DCNN approach was initially proposed by Bouaafia, Messaoud, Khemiri, and Sayadi [18] for the purpose of VVC in-loop filtering. By utilizing this strategy, The Bjontegaard Delta-Rate (BD-rate) decreased while simultaneously the BD-PSNR was enhanced. Zhang, Wang, Huang, Jiang, and Wang [19] developed an effective CU partition determination approach for VVC. This method was developed with the use of an enhanced DAG-SVM model. A new in-loop filter for video coding was presented by Li and Ji [20], who utilized a lightweight multi attention recursive residual CNN at the time of their presentation. The primary objective of this filter is to find solutions to the issues that arise as a result of highly intricate parameters and the necessity for a large number of models to cope with a wide range of quantization parameters (QPs). The solution that has been provided incorporates QPs, Frame Type (FT), and Temporal Layer (TL) into a single cohesive model. This results in considerable reductions in the bit error rate (BER) in all-intra as well as randomaccess configurations. A deep learning approach was created by the authors of the study by Kuanar et al. [21] in order to execute SAO filtering operations in HEVC. The SSIM, BD-BR, and BD-PSNR measurements have all shown that this approach has showed exceptional performance. Numerous studies have demonstrated the potential of deep learning to increase the efficiency of video coding, particularly in reducing bitrates when compared to traditional approaches such as HEVC. Deep Neural Networks (DNNs) have become a powerful tool in video compression by dynamically adjusting to input data and learning optimal filtering strategies for compression tasks. In particular, DNNs enable precise artifact reduction and pattern recognition, offering notable improvements in both compression efficiency and video quality [22]. In the context of this work, DNNs are leveraged to enhance the in-loop filtering process in Versatile Video Coding (VVC), where their ability to optimize parameters through extensive training significantly improves video quality and reduces bitrate.

The key to this approach is the meticulous tuning of model parameters, which ensures optimal learning and enhances performance. This tuning minimizes manual intervention, allowing the network to automatically learn complex filtering tasks and make data-driven decisions. Additionally, transfer learning techniques, where knowledge learned from one domain is applied to another, enhance the model's adaptability and extend its applicability across different video datasets [23], [24]. The findings of these studies indicate that deep learning and machine learning have had a significant influence on video coding, particularly with regard to the enhancement of video quality and the introduction of new levels of complexity. The proposed methods, which make use of CNN structures, had shown good results in terms of lowering the complexity of the encoding process and improving the quality of the video in accordance with the HEVC and VVC standards.

In Tab. 1, we have encapsulated the key insights distilled from prior research.

Tab. 1. Summary of Related Work

Reference	Methodology/Model	Application	Performance Metrics
[10]	MFRNet: Multi-level feature review residual dense blocks (MFRBs)	In-loop filtering and post-processing in video compression	Compression efficiency, Visual quality
[11]	Dense residual convolutional neural network (DRN)	In-loop filtering in VVC	Coding performance, Artifact reduction
[12]	CNN-LSTM hybrid model for predicting HEVC CU inter-mode partitions	Predictive coding in HEVC	Rate-distortion performance, Complexity reduction
[13]	Machine Learning-based CU partition methods: Deep CNN and online SVM methods for reducing complexity in inter-mode HEVC	CU partitioning in HEVC	Complexity reduction, Encoding duration reduction
[14]	CNN-LSTM learning method for simplifying HEVC	Video coding	Complexity reduction, Bjøntegaard Delta-Rate, BD-PSNR improvement
[15]	Deep CNN model for FVC using Rapid quantum tunneling QT partitioning	Video coding	Intra-mode encoding time, Bitrate performance improvement
[16]	Deep CNN-based solution for HEVC intra frame coding	Video coding	Encoding speed improvement
[17]	Enhanced Deep CNN (EDCNN) for in-loop filtering in HEVC	In-loop filtering in HEVC	PSNR improvement, Rejection Detection enhancement
[18]	WSE-DCNN Deep CNN model for VVC in-loop filtering	In-loop filtering in VVC	Bjontegaard Delta-Rate reduction, BD-PSNR enhancement
[19]	Enhanced DAG-SVM model for CU partition determination in H.266/VVC	CU partitioning in VVC	Efficiency improvement in video coding
[20]	Lightweight multi attention recursive residual CNN for in-loop filtering in video coding	In-loop filtering in video coding	Bit error rate reduction
[21]	Deep learning approach for SAO filtering operations in HEVC	SAO filtering in HEVC	SSIM, BD-BR, BD-PSNR improvement

2.Proposed methodology

This section presents the proposed methodology for enhancing video compression through the integration of a Residual Deep Convolutional Neural Network (RDCNN) as a replacement for the conventional VVC loop filtering module. The methodology is organized into three main phases:

2.1. Framework Configuration and Training Parameters

The deep learning framework is initially configured for offline training, utilizing the TensorFlow-GPU platform to optimize computational efficiency for processing high-dimensional video data. In this configuration, a batch size of 128 is employed to balance computational efficiency and the capacity to capture complex data patterns. The training process consists of 50 epochs, ensuring adequate learning while mitigating the risk of overfitting. A learning rate of 0.001 is maintained to facilitate stable convergence, allowing the model to learn effectively from the provided data. Additionally, a weight decay of 0.1 is applied after 10 epochs as a regularization strategy. This approach enables the model to initially learn without regularization for potentially better performance, followed by stabilization through regularization techniques. The Adam optimization algorithm [27] is utilized for its adaptive learning rate capabilities, allowing for individual adjustment of the learning rate for each parameter, thereby facilitating faster convergence and stabilization during training.

2.2. Feature Extraction and Post-Training Prediction

In this phase, the CNN is specifically designed to extract significant features from video frames, focusing on attributes such as luminance, chrominance, and motion vectors. These features are critical for video content inference, which informs the development of optimal filtering strategies. As shown in Figure 1, we introduce the Residual Deep Convolutional Neural Network (RDCNN) model as a replacement for the conventional VVC loop filtering module. This module encompasses functionalities for deblocking filtering (DBF), sample adaptive offset (SAO), and adaptive loop filtering (ALF). The primary objective of the RDCNN is to enhance visual appearance while preserving the advantages of the coding process. The feature extraction process begins with the input of raw video frames into the CNN. During training, the model learns to identify and represent these attributes through multiple convolutional layers, where the features are progressively refined. The architecture includes several convolutional layers, followed by activation functions such as ReLU (Rectified Linear Unit) to introduce non-linearity, and pooling layers to reduce spatial dimensions while preserving essential features. Upon completion of training, the CNN is capable of accurately predicting optimal filtering decisions for each video frame during the encoding process. This prediction involves analyzing the extracted features to dynamically adjust filtering intensity and orientation, thereby guiding the encoder in determining appropriate compression levels and preservation strategies. By leveraging the RDCNN's capabilities, we achieve a refined balance between video quality and bitrate, ultimately resulting in higher-quality videos that are compressed at lower bitrates. This synergy of deep learning techniques with traditional video coding methods significantly enhances compression efficiency and effectiveness.

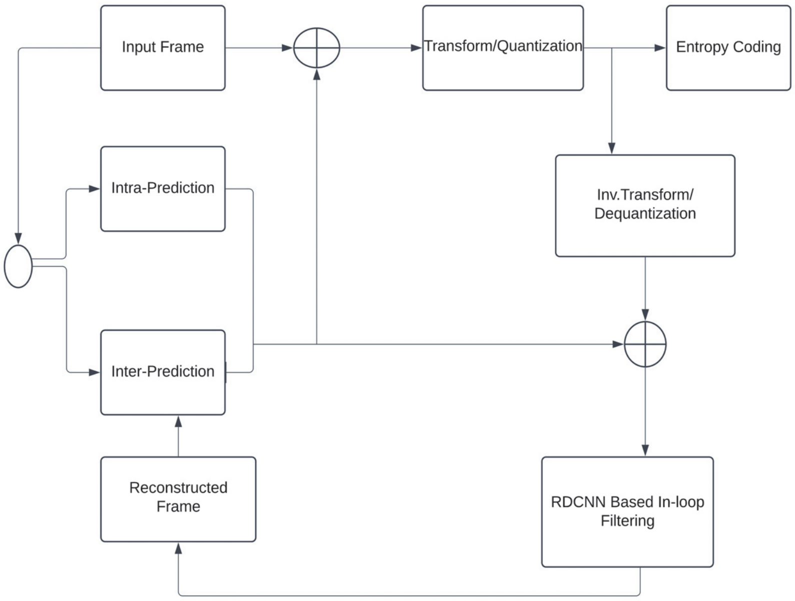

Fig. 1. The proposed model

2.3. Rate Distortion Optimization (RDO) Strategy

The Rate Distortion Optimization (RDO) technique is a critical component of our video coding methodology, guiding the effective application of the RDCNN in-loop filters across the entire video frame. The RDO criterion is mathematically represented as

$$J = D + \lambda R \quad (1)$$

where J is the overall cost function, D represents the distortion between the original and reconstructed frames, and R denotes the number of bits used for coding. The Lagrange multiplier (λ) adjusts the balance between these two components, enabling a controlled trade-off between video quality and bitrate. During the encoding process, the RDO mechanism evaluates potential filtering strategies generated by the CNN in conjunction with the overall distortion and bitrate. By optimizing this trade-off, the RDO technique ensures that the chosen filtering methods enhance visual quality while adhering to bitrate constraints. The application of RDO is particularly vital in scenarios where maintaining quality is essential, as it helps to prevent significant quality degradation that can arise from excessive compression.

The effectiveness of the RDO strategy is further enhanced by iterative optimization, allowing the encoder to refine its filtering decisions based on the outcomes of previous iterations. This iterative process contributes to improved visual quality in the final compressed output, ensuring that the advantages of the RDCNN's deep learning capabilities are fully realized in the video coding framework.

3. RDCNN architecture

In the **Residual CNN (ResCNN)** architecture, adapted from the widely recognized **ResNet architecture** introduced by He et al. [29], each layer plays a crucial role in feature extraction and the filtering process, tailored specifically for enhancing video frames in **Versatile Video Coding (VVC)**. While the original ResNet architecture was designed for image classification tasks, we have modified it to better capture both **spatial and temporal dependencies** in video data, which are essential for making accurate predictions regarding optimal filtering decisions. Below, we detail the contribution of each layer to the enhancement of video frames in VVC, with particular emphasis on residual layers.

Input Layer: This layer receives preprocessed video frames that have been standardized and normalized, denoted as X. It initiates the network, where raw pixel data is passed into subsequent layers for processing.

Convolutional Layers: These layers, represented as $\mathrm{Conv}(X, W)$ are responsible for extracting spatial features from the input frames. The architecture includes a total of five convolutional layers, each with 64 filters, set with padding='same' and using ReLU activation. Here, W represents the filters applied to the data, with each filter designed to detect specific patterns within the frames. The convolution operation is defined as:

$$Y = \sum_{i=0}^{k-1} X_i \cdot W_i, \tag{2}$$

where X_i represents the input values, W_i denotes the filter weights, and Y is the resulting feature map. Deconvolution layers are not required in this architecture because the task does not involve frame upscaling or reconstruction. Instead, the architecture focuses on making pixel-wise filtering decisions, maintaining spatial alignment through the convolutional and residual layers without needing deconvolution.

Residual Layers: Inspired by the ResNet architecture [29], these layers learn the residual function F(X, {Wi}), capturing the difference between the input and the desired output. The residual connections help the network bypass vanishing gradients and enable deeper network training by learning the identity function. The output of a residual block is expressed as:

$$Y = F(X, \{W_i\}) + X. \tag{3}$$

Activation Functions: ReLU (Rectified Linear Unit), defined as $g(X) = \max(0, X)$, introduces non-linearity into the network, allowing it to learn complex patterns. ReLU is widely used due to its simplicity and efficiency in mitigating the vanishing gradient problem. Fan Zhang's study [28] further highlights ReLU's efficacy in enhancing video coding efficiency through CNN-based post-processing techniques.

Pooling Layers: These layers, denoted as Pool(X), reduce the spatial dimensions of the feature maps, thus lowering computational complexity and preventing overfitting. Max pooling is used to select the maximum value within a designated region of the feature map:

$$Y = \mathrm{Max}(X_i). \tag{4}$$

Fully Connected Layers: After feature extraction, fully connected layers FC(X, W, b)) are used to make final predictions for pixel-wise filtering decisions, determining the intensity and direction of filtering. These layers are not intended for spatial reconstruction, but rather for making adjustments to each pixel during the compression process. The output is computed as:

$$Y = WX + b, \tag{5}$$

where W denotes the weight matrix, X signifies the input vector, and b represents the bias vector.

Output Layer: The output layer generates the final filtering predictions for each pixel in the video frame. These decisions influence filtering strength and direction during the encoding process, enhancing visual quality while maintaining bitrate efficiency.

Our Residual Deep Convolutional Neural Network (RDCNN) architecture synergizes convolutional and residual layers to effectively extract features from video frames and predict optimal filtering decisions. By harnessing the principles of residual learning, the architecture significantly improves the compression efficiency of Versatile Video Coding (VVC), enabling the delivery of higher-quality video at lower bitrates. The model employs the Adam optimizer, which is renowned for its adaptive learning capabilities, facilitating stable convergence during training. To ensure pixel-level accuracy in filtering predictions, we utilize the Mean Squared Error (MSE) as the loss function. This approach aligns the model's outputs with the ground truth values, thereby enhancing the model's precision in video compression tasks and reinforcing the integrity of the visual content.

4. Data collection and preprocessing

The proposed method is trained using the publicly available BVI-DVC dataset [26], which comprises 800 video sequences spanning resolutions from 270p to 2160p. This dataset is specifically designed for training CNN-based video compression systems, emphasizing machine learning tools that enhance conventional coding architectures, including spatial resolution and bit depth up-sampling, post-processing, and in-loop filtering. 80% of these sequences are designated for training, while the remaining 20 % are held for validation. Employing a random-access scenario, the VVC VTM-22.2 test model [25] compresses these sequences using various QP values (29, 35, 38, 40, and 41), resulting in reconstruction video images with both luma and chroma components. These images are then divided into 64×64 patches in a randomized manner.

5. Results

During the evaluation, particular emphasis is placed on the loss curve (Fig. 2), which shows the model's gradual improvement in performance as training progresses. The decrease in the loss values reflects the model's ability to iteratively optimize its parameters and improve its performance through each training epoch. This optimization occurs within the context of training the Convolutional Neural Network (CNN) as an in-loop filter for Versatile Video Coding (VVC).

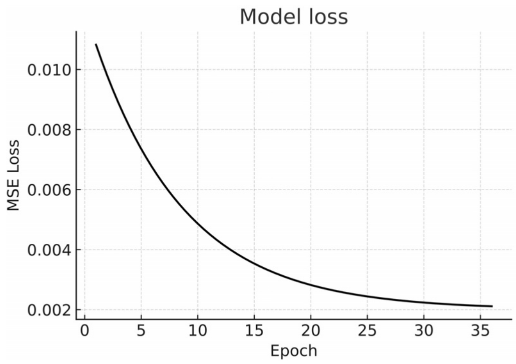

Fig. 2. Loss function

The effectiveness of the proposed filter was evaluated on a diverse set of YUV video files, including **BUS_352×288**, **CREW_352×288**, **HARBOUR_352×288_30**, **MOBILE_352×288_30**, **SOCCER_352×288**, and **stefan_cif**. These sequences, commonly used as benchmarks in video processing and encoding, have a resolution of 352×288 pixels and a frame rate of 30 frames per second. The evaluation followed the configurations outlined in the **encoder_randomaccess_vtm.cfg** file, to ensure uniform testing conditions, we conducted experiments across various QP values (29, 35, 38, 40, and 41), allowing bitrate to vary accordingly with each QP. This approach provided a balanced comparison of the impact of QP on video quality and compression efficiency under consistent encoding settings. This approach systematically explores the impact of different QP settings on video encoding quality and efficiency. Higher QP values, like 41, result in higher compression rates but may introduce more visible artifacts. In contrast, lower QP values, such as 29, preserve more detail but at the cost of larger file sizes. Maintaining a fixed bit rate ensures that the observed differences are solely due to QP variations, allowing for a clear comparison of the encoding performance under different conditions. The analysis, conducted over 10 frames for each sequence, revealed a compression ratio of approximately 15.36, which corresponds to the lower quantization parameter (QP) values used in the experiments. This indicates a significant reduction in video size while maintaining acceptable quality levels. The **YUV Peak Signal-to-Noise Ratio (PSNR)** was employed as the primary metric for evaluating video quality, with compression ratios varying depending on the PSNR achieved for each video. As shown in Fig. 3 through 9, the relationship between PSNR and compression ratio is illustrated for different test sequences, where higher PSNR values generally correspond to lower compression ratios.

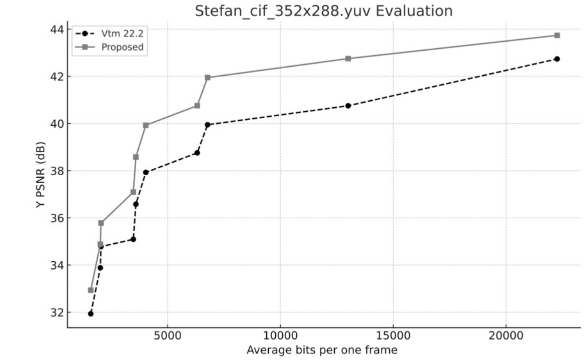

Fig. 3. Proposed vs VTM22.2 on Stefan.yuv

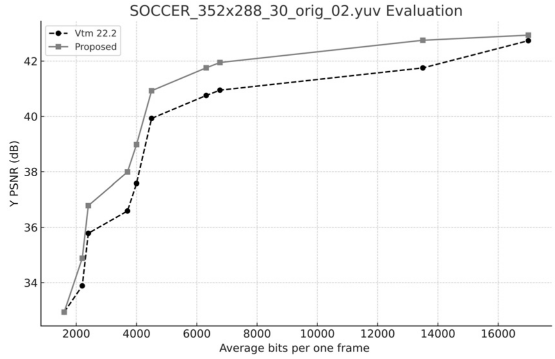

Fig. 4. Proposed vs VTM22.2 on Soccer.yuv

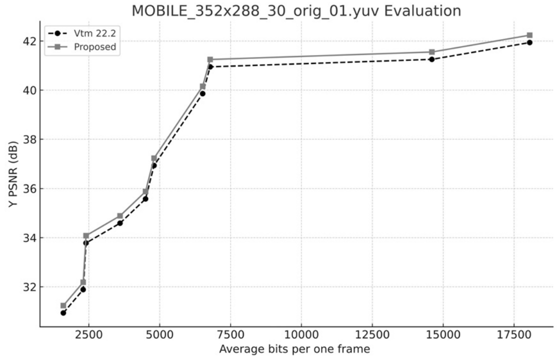

Fig. 5. Proposed vs VTM22.2 on Mobile.yuv

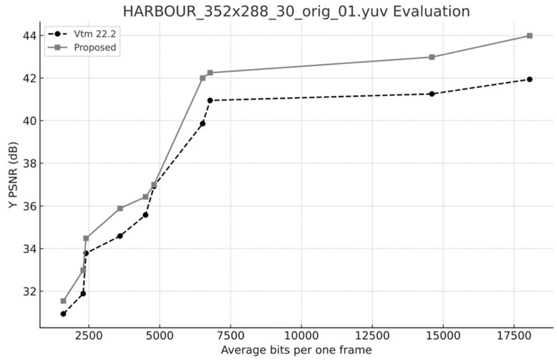

Fig. 6. Proposed vs VTM22.2 on Harbour.yuv

The results presented in Tab. 2 compare the performance of the proposed method against **VTM22.2**. The proposed filter consistently achieved higher average YUV PSNR values, indicating improved video quality. For example, the proposed method achieved a PSNR of 41.2 on the **Stefan_cif** sequence, compared to 38.4 for VTM22.2.

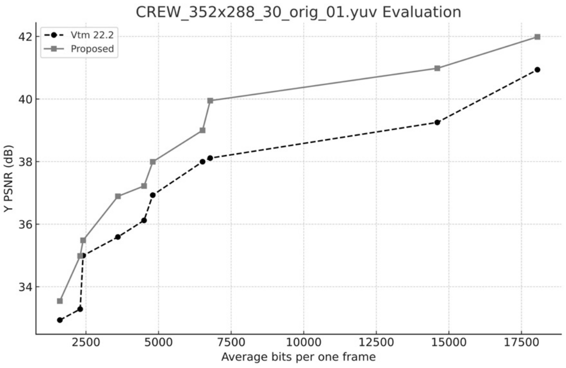

Fig. 7. Proposed vs VTM22.2 on Crew.yuv

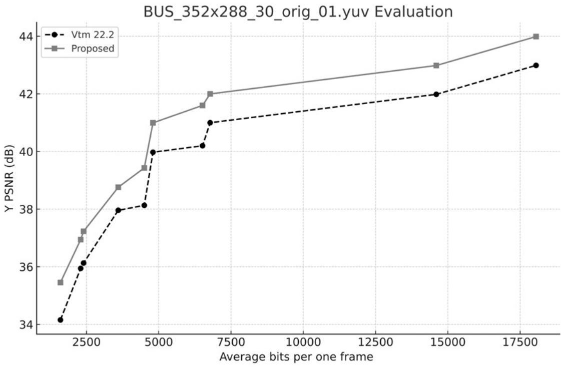

Fig. 8. Proposed vs VTM22.2 on Bus.yuv

Tab. 2. Results for all samples (Proposed vs VTM22.2)

YUV sample	Average YUV PNSR	
	Proposed	VTM 22.2
Stefan_cif	41.2	38.4
Soccer	39.9	38.1
Mobile	38.2	37.3
Harbour	38.9	36.3
Crew	38.2	36.9
Bus	39.9	38.1

The **Residual Deep Convolutional Neural Network (RDCNN)** system, when integrated into the VVC framework, outperformed existing methods in terms of both luma and chroma component compression. This improvement in compression efficiency translated into enhanced visual quality, as demonstrated in Tab. 3, where the proposed method surpassed the baseline VTM22.2 and other related approaches.

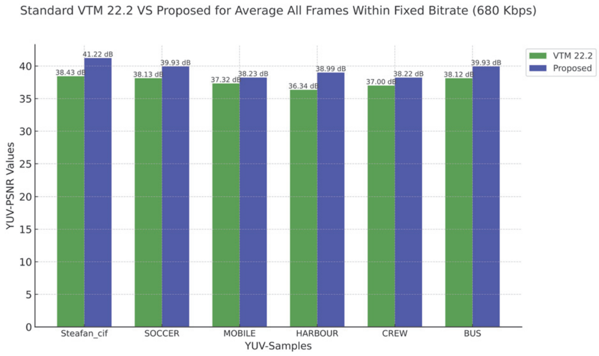

Fig. 9. Bar-Plot for all samples (Proposed vs VTM22.2)

Tab. 3. Comparison with other works

Algorithm	YUV PSNR (db)
[18]	40.0
VTM 22.2	38.4
Proposed	41.2

Tab. 4 shows another comparison of the proposed approach with a number of CNN-based filtering strategies. Tab. 4 shows how encoding performance is compared to other methods using the Versatile Video Coding (VVC) architecture in terms of minimizing RD [30, 31]. Chen et al. [30] presented a method that uses a dense residual convolutional neural network's (DRN) in-loop filter to enhance the quality of reconstructed films. Using the DIV2K dataset [32] for training, this network is placed after the deblocking filter (DBF) and before SAO and ALF in the VVC VTM reference software. Furthermore, an additional in-loop filtering technique based on CNN is suggested [31], designed for both inter and intra frames, which operates in the VVC VTM prior to the ALFs, with the DBF and SAO disabled.

Table. 4. Comparison of the proposed approach with a number of CNN-based filtering strategies

	Approach [30]			Approach [31]			Proposed Approach		
Class	(Y)	(U)	(V)	(Y)	(U)	(V)	(Y)	(U)	(V)
A1	-1.27	-3.38	-5.10	0.87	0.12	0.22	-1.33	-8.66	-9.05
A2	-2.21	-5.74	-2.88	-1.12	-0.52	-2.11	-1.10	-11.02	-8.08
B	-1.13	-4.73	-4.55	-0.83	-0.47	-1.20	-2.82	-7.78	-14.64
C	-1.39	-3.63	-4.36	-1.76	-3.64	-6.80	-2.14	-4.42	-7.57
D	-1.39	-1.96	-3.08	-2.95	-3.27	-7.35	-2.53	-5.53	-8.50
Over all	-1.47	-3.88	-3.99	-1.16	-1.56	-3.44	-2.43	-6.96	-9.43

6. Discussion

In light of the proposed methodology, the results obtained underscore the effectiveness of integrating deep learning techniques, specifically the RDCNN architecture, into the VVC framework for video compression enhancement. Through meticulous training and optimization on the TensorFlow-GPU platform, our model demonstrates remarkable performance improvements over conventional methods. The chosen parameters, including batch size, training epochs, and learning rate adjustment strategy, play pivotal roles in guiding the model towards learning optimal filtering strategies for video frames. By leveraging the inherent capabilities of the Convolutional Neural Network (CNN) to extract significant features from video frames, the proposed RDCNN architecture autonomously learns to discern and predict optimal filtering decisions. These decisions, accurately predicted during the encoding process, dynamically adjust filtering intensity and orientation, ultimately enhancing visual quality while maintaining bitrate efficiency. The observed convergence of loss metrics across training epochs further validate the model's incremental progression towards optimal performance, aligning with established neural network training principles. Moreover, the detailed assessment conducted across various YUV files showcases consistently higher average YUV PSNR values compared to both VTM 22.2 and other related methods. This indicates not only improved compression efficiency but also enhanced visual quality, which is crucial for diverse video processing tasks. The meticulous design of the RDCNN architecture, with its assortment of layers including residual layers, effectively captures spatial and temporal dependencies in video data, contributing to precise predictions concerning optimal filtering decisions. The comparison between the proposed method and VTM 22.2 across various YUV samples reveals significant improvements in average YUV PSNR values, indicative of enhanced visual quality achieved by our approach. For instance, in the Stefan_cif sample, our proposed method achieves an average YUV PSNR of 41.2, surpassing VTM 22.2's score of 38.4. Similarly, in the Soccer sample, our method demonstrates superior performance with an average YUV PSNR of 39.9 compared to VTM 22.2's 38.1. This trend persists across other samples, including Mobile, Harbour, Crew, and Bus, where our proposed method consistently outperforms VTM 22.2. Additionally, we calculated the BD-rate, which further corroborates our findings. The overall BD-rate results indicate that the proposed approach enhances compression efficiency more effectively than the existing methods presented in references [30] and [31], achieving BD-rate improvements of -2.43, -6.96, and -9.43 for Y, U, and V, respectively. These results reinforce the efficacy of our method, highlighting its capacity to enhance compression efficiency while preserving visual quality. The superior performance observed in comparison to the existing algorithm further validates the efficacy of our approach. By achieving higher average YUV PSNR values and improved BD-rates across multiple YUV samples, our method demonstrates its potential to deliver superior-quality videos at reduced bitrates, addressing key challenges in video compression and encoding.

7. Conclusion

In conclusion, this work presents a novel approach to enhancing video compression through the integration of deep learning techniques, specifically leveraging a Residual Deep Convolutional Neural Network (RDCNN) model within the Versatile Video Coding (VVC) framework. The methodology involves configuring a deep learning framework for offline training with carefully selected parameters to optimize computational efficiency and memory usage. The use of a Convolutional Neural Network (CNN) for feature extraction from video frames, combined with the RDCNN model for in-loop filtering, aims to improve video quality while maintaining efficient compression. The proposed method demonstrates significant advancements in compression efficiency and visual quality, as evidenced by the achieved compression ratio and the improvement in YUV Peak Signal-to-Noise Ratio (PSNR) across various video sequences. The integration of deep learning methodologies, such as the RDCNN model, into traditional video coding processes represents a significant step forward in the field of video compression. By dynamically adjusting to input data and autonomously solving problems without constant human oversight, deep learning models like the RDCNN offer a promising solution for enhancing video compression techniques. The empirical evidence presented in this work underscores the potential of deep learning in transforming video compression, paving the way for more efficient and high-quality video encoding solutions.

References

[1] Norkin A, Bjontegaard G, Fuldseth A, Narroschke M, Ikeda M, Andersson K, Zhou M, Van der Auwera G. HEVC deblocking filter. IEEE Trans Circuits Syst Video Technol 2012; 22(12): 1746-1754. DOI: 10.1109/TCSVT.2012.2223053.

[2] Fu C-M, Alshina E, Alshin A, Huang Y-W, Chen C-Y, Tsai C-Y, Hsu C-W, Lei S-M, Park J-H, Han W-J. Sample adaptive offset in the HEVC standard. IEEE Trans Circuits Syst Video Technol 2012; 22(12): 1755-1764. DOI: 10.1109/TCSVT.2012.2221529.

[3] Tsai C-Y, Chen C-Y, Yamakage T, Chong IS, Huang Y-W, Fu C-M, Itoh T, Watanabe T, Chujoh T, Karczewicz M, Lei S-M. Adaptive loop filtering for video coding. IEEE J Sel Top Signal Process 2013; 7(6): 934-945. DOI: 10.1109/JSTSP.2013.2271974.

[4] Bross B, Chen J, Liu S, Wang YK. Versatile video coding (draft 10). ITU-T and ISO/IEC JVET-S2001 2020.

[5] Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans Image Process 2017; 26(7): 3142-3155. DOI: 10.1109/TIP.2017.2662206.

[6] Wang Z, Chen J, Hoi SCH. Deep learning for image super-resolution: A survey. IEEE Trans Pattern Anal Machine Intell 2020; 43(10): 3365-3387. DOI: 10.1109/TPAMI.2020.2982166.

[7] Dong C, Deng Y, Loy CC, Tang X. Compression artifacts reduction by a deep convolutional network. 2015 IEEE International Conference on Computer Vision (ICCV) 2015: 576-584. DOI: 10.1109/ICCV.2015.73.

[8] Dai Y, Liu D, Wu F. A convolutional neural network approach for post-processing in HEVC intra coding. In Book: Amsaleg L, Guðmundsson GÞ, Gurrin C, Jónsson BÞ, Satoh S, eds. MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4-6, 2017, Proceedings, Part I. Cham, Switzerland: Springer In-

ternational Publishing AG; 2017: 28-39. DOI: 10.1007/978-3-319-51811-4_3.

[9] Lin W, He X, Han X, Liu D, See J, Zou J, Xiong H, Wu F. Partition-aware adaptive switching neural networks for post-processing in HEVC. IEEE Trans Multimed 2019; 22(11): 2749-2763. DOI: 10.1109/TMM.2019.2962310.

[10] Ma D, Zhang F, Bull DR. MFRNet: a new CNN architecture for post-processing and in-loop filtering. IEEE J Sel Top Signal Process 2020; 15(2): 378-387. DOI: 10.1109/JSTSP.2020.3043064.

[11] Chen S, Chen Z, Wang Y, Liu S. In-loop filter with dense residual convolutional neural network for VVC. 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) 2020: 149-152. DOI: 10.1109/MIPR49039.2020.00038.

[12] Bouaafia S, Khemiri R, Sayadi FE, Atri M, Liouane N. A deep CNN-lstm framework for fast video coding. In Book: Moataz AE, Mammass D, Mansouri A, Nouboud F, eds. Image and Signal Processing. 9th International Conference, ICISP 2020, Marrakesh, Morocco, June 4–6, 2020, Proceedings. Cham, Switzerland: Springer International Publishing AG; 2020: 205-212. DOI: 10.1007/978-3-030-51935-3_22.

[13] Bouaafia S, Khemiri R, Sayadi FE, Atri M. SVM-based inter prediction mode decision for HEVC. 2020 17th Int Multi-Conf on Systems, Signals & Devices (SSD) 2020: 12-16. DOI: 10.1109/SSD49366.2020.9364153.

[14] Bouaafia S, Khemiri R, Sayadi FE, Atri M. Fast CU partition-based machine learning approach for reducing HEVC complexity. J Real-Time Image Process 2020; 17(1): 185-196. DOI: 10.1007/s11554-019-00936-0.

[15] Amna M, Imen W, Ezahra SF, Mohamed A. Fast intra-coding unit partition decision in H.266/FVC based on deep learning. J Real-Time Image Process 2020; 17(6): 1971-1981. DOI: 10.1007/s11554-020-00998-5.

[16] Hsu T-Y, Lu Y-J, Hsieh T-H, Wang C-C. An efficient HEVC intra frame coding based on deep convolutional neural network. 2021 IEEE/ACIS 22nd Int Conf on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) 2021: 138-141. DOI: 10.1109/SNPD51163.2021.9704928.

[17] Pan Z, Yi X, Zhang Y, Jeon B, Kwong S. Efficient in-loop filtering based on enhanced deep convolutional neural networks for HEVC. IEEE Trans Image Process 2020; 29: 5352-5366. DOI: 10.1109/TIP.2020.2982534.

[18] Bouaafia S, Messaoud S, Khemiri R, Sayadi FE. VVC in-loop filtering based on deep convolutional neural network. Comput Intell Neurosci 2021; 2021(1): 9912839. DOI: 10.1155/2021/9912839.

[19] Zhang Q, Wang Y, Huang L, Jiang B, Wang X. Fast CU partition decision for H.266/VVC based on the improved DAG-SVM classifier model. Multimedia Systems 2021; 27(1): 1-14. DOI: 10.1007/s00530-020-00688-z.

[20] Li M, Ji W. Lightweight multiattention recursive residual CNN-based in-loop filter driven by neuron diversity. IEEE Trans Circuits Syst Video Technol 2023; 33(11): 6996-7008. DOI: 10.1109/TCSVT.2023.3270729.

[21] Kuanar S, Rao KR, Conly C, Gorey N. Deep learning based HEVC in-loop filter and noise reduction. Signal Process: Image Commun 2021; 99: 116409. DOI: 10.1016/j.image.2021.116409.

[22] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521(7553): 436-444. DOI: 10.1038/nature14539.

[23] Ma S, Zhang X, Jia C, Zhao Z, Wang S, Wang S. Image and video compression with neural networks: A review. IEEE Trans Circuits Syst Video Technol 2019; 30(6): 1683-1698. DOI: 10.1109/TCSVT.2019.2910119.

[24] Bidwe RV, Mishra S, Patil S, Shaw K, Vora DR, Kotecha K, Zope B. Deep learning approaches for video compression: A bibliometric analysis. Big Data Cogn Comput 2022; 6(2): 44. DOI: 10.3390/bdcc6020044.

[25] jvet/VVCSoftware_VTM/Tags/VTM-22.2. 2025. Source: <https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tags/VTM-22.2>.

[26] Ma D, Zhang F, Bull DR. BVI-DVC: A training database for deep video compression. IEEE Trans Multimed 2021; 24: 3847-3858. DOI: 10.1109/TMM.2021.3108943.

[27] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv Preprint. 2014. Source: <https://arxiv.org/abs/1412.6980>. DOI: 10.48550/arXiv.1412.6980.

[28] Zhang F, Feng C, Bull DR. Enhancing VVC through CNN-based post-processing. 2020 IEEE Int Conf on Multimedia and Expo (ICME) 2020: 1-6. DOI: 10.1109/ICME46284.2020.9102912.

[29] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2016: 770-778. DOI: 10.1109/CVPR.2016.90.

[30] Chen S, Chen Z, Wang Y, Liu S. In-loop filter with dense residual convolutional neural network for VVC. 2020 IEEE Conf on Multimedia Information Processing and Retrieval (MIPR) 2020: 149-152. DOI: 10.1109/MIPR49039.2020.00038.

[31] Kawamura K, Kidani Y, Naito S. CE13-2.6/CE13-2.7: Evaluation results of CNN based in-loop filtering. Document JVET-N0710, 14th JVET meeting, Geneva, Switzerland 2019: 19-27.

Authors' information

Murooj Khalid Ibraheem Ibraheem, A postgraduate student at Moscow Institute of Physics and Technology (MIPT), Phystech School of Radio Engineering and Computer Technologies (FRKT), Department of multimedia technologies and telecommunications. Assistant teacher at Mustansiriyah University, Collage of Engineering, Department of Computer Engineering, Iraq. Research interests: Computer Network, Computer Vision, Video Processing, Machine Learning, Deep Learning, Software Engineering. E-mail: *ibragim.m@phystech.edu*

Alexander Viktorovich Dvorkovich, Head of the Department of multimedia technologies and telecommunications, Phystech School of Radio Engineering and Computer Technologies (FRKT), Moscow Institute of Physics and Technol-

ogy (MIPT), Doctor of Technical Sciences, corresponding member of the Russian Academy of Sciences. Research interests: Video compression, Audio compression, Wireless media transmission, Telecommunication systems, Assessment of the quality of multimedia processing and transmission, Satellite communications.
Email: dvork.alex@gmail.com

Ammar Mudheher Sadeq Al-Temimi, Candidate of Technical Sciences, Director of Iraqi National Data Center, General Secretariat for the Council of Ministers. Research interests: Digital Image Processing and Computer Vision, Ciphering, Data Compression, Pattern Recognition, Database Systems. E-mail: ammar.m.altemimi@gmail.com

Received: June 03, 2024. The final version – January 04, 2025.