A Study of Image and Video Compression Using Neural Networks

Ruhiat Sultana
Assistant Professor, Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Science, Hyderabad, India.
Mail: ruhiatsultana@gmail.com

Javeed Mohammad
Associate Professor, Department of Electronics and Communication Engineering, Sree Dattha Institute of Engineering and Science, Hyderabad, India.
Mail: javeed.rahmanee@gmail.com

Nisar Ahmed
Associate Professor, Department of Electronics and Communication Engineering, Sree Dattha Institute of Engineering and Science, Hyderabad, India.
Mail: nisar.ahcet@gmail.com

Abstract

Picture and video coding developments have progressed by leaps and bounds in recent years. However, as image and video acquisition devices become more prevalent, the growth rate of image and video data is outpacing the compression ratio increase. It's been widely acknowledged that seeking more coding performance enhancement within the conventional hybrid coding paradigm is becoming increasingly difficult. Deep convolution neural network (CNN) is a form of neural network that has seen a resurgence in recent years and has seen a lot of success in the fields of artificial intelligence and signal processing. also offers a novel and promising image and video compression solution. We present a systematic, thorough, and up-to-date analysis of neural network-based image and video compression techniques in this paper. For images and video, the evolution and advancement of neural network-based compression methodologies are discussed. More precisely, cutting-edge video coding techniques based on deep learning and the HEVC system are introduced and addressed, with the goal of significantly improving state-of-the-art video coding efficiency. In addition, end-to-end image and video coding frameworks based on neural networks are examined, revealing intriguing explorations on next-generation image and video coding frameworks. The most important research works on image and video coding related topics using neural networks are highlighted, as well as future developments. The joint compression of semantic and visual information, in particular, is tentatively investigated in order to formulate high-efficiency signal representation structures for both human and machine vision. In the age of artificial intelligence, which are the two most popular signal receptors.

Index Terms—*Neural network, deep learning, CNN, image compression, video coding.*

I. INTRODUCTION

IMAGE and video compression is critical for delivering high-quality image/video services when transmission networks and storage capacity are small. Redundancies in photographs and videos, such as spatial redundancy, visual redundancy, and statistical redundancy, are critical for image and video compression. Furthermore, the presence of temporal redundancy in video sequences allows video compression to achieve higher compression ratios than image compression.

Huffman coding [1], Golomb code [2], and arithmetic coding [3] are examples of early image compression methods that rely on direct entropy coding to minimise statistical redundancy within the image. The Fourier transform [4] and Hadamard transform [5] were proposed in the late 1960s for image compression by encoding the spatial frequencies. Ahmed et al. proposed the Discrete Cosine Transform (DCT) for image coding in 1974 [6], which compacts image energy in the low frequency domain, allowing for much more effective compression in the frequency domain.

In addition to entropy coding and transform techniques for reducing statistical redundancy, prediction and quantization techniques for reducing spatial and visual redundancy in images are proposed. JPEG, the most widely used image compression format, is an effective image compression scheme that incorporates previous coding techniques. It divides an image into blocks, which are then transformed into the DCT domain. The differential pulse code modulation (DPCM) [7] is applied to the DC components of each block, compressing the prediction residuals of DC components between neighbouring DCT blocks rather than the DC value directly. Since humans are less susceptible to information loss in high frequency bits, a special quantization table is designed to retain low-frequency information while discarding more high-frequency (noise-like) data to minimise visual redundancy [8]. JPEG 2000 [9], a well-known still image compression standard, uses the 2D wavelet transform rather than the DCT to represent images in a compact form, and EBCOT [10], an effective arithmetic coding process, to reduce the statistical redundancy in wavelet coefficients.

Due to the high similarity between successive frames recorded in a very short time period, temporal continuity, which could be eliminated by inter-frame prediction, becomes the dominant factor in video coding. The block based motion prediction was proposed in the 1970s to obtain inter-prediction efficiently [11]. Motion compensation transform system [12], which is now known as the hybrid prediction/transform coder, was proposed by Netravali and Stuller in 1979. The reader presented an overview of the first generation methods' historical evolution [13].

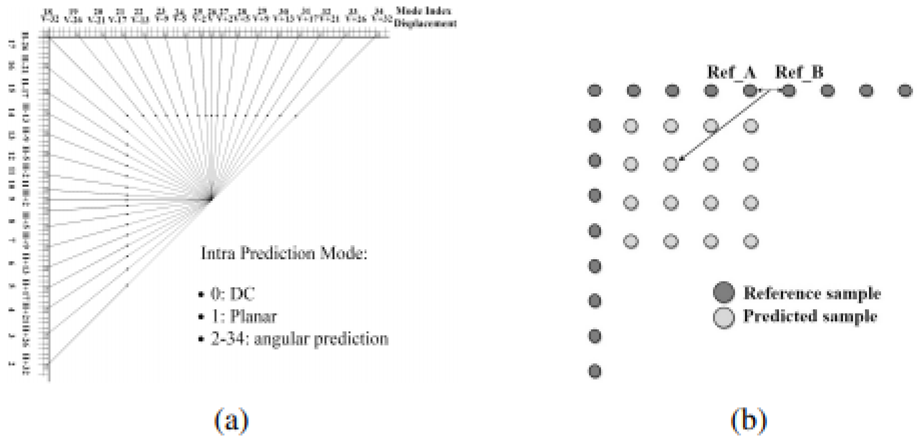

Fig. 1. Illustration of HEVC intra prediction. (a) different intra modes; (b) angular prediction instance.

Hybrid prediction/transform coding methods have had a lot of success after decades of growth. MPEG-1/2/4, H.261/2/3, and H.264/AVC [14], as well as AVS (Audio and Video Coding Standard in China) [15] and HEVC [16], have all been developed and are widely used in various applications. As an example, the most recent video coding standard, HEVC, used neighbouring reconstructed pixels to predict the current coding block, with 33 angular intra prediction modes, including the DC mode and the planar mode, as shown in Fig. 1. In terms of inter-frame coding, HEVC builds on its ancestor, H.264/AVC, by improving it from a variety of angles. For example, increasing the PU division's diversity, using more interpolation filter taps for sub-sample motion compensation [17], and refining the side information coding, including more most probable modes (MPMs) for intra mode coding [18], advanced motion vector prediction (AMVP), and merge mode for motion vector predictor coding [19]. Loop filtering is another modern video coding technique in today's video coding framework, and several loop filters [20]–[25] have been proposed since 2000. Deblock filtering [26], [27], and sample adaptive offset (SAO) [28] have been incorporated into HEVC in this paper. The refinement techniques for conventional hybrid video coding frameworks based on image and video local correlations, on the other hand, are based on image and video local correlations. are becoming increasingly difficult to increase coding quality.

Neural networks, especially convolutional neural networks (CNN), have recently achieved considerable success in a variety of fields, including image/video comprehension, encoding, and compression. One or more convolutional layers are normally present in a CNN. Some activities, in particular, add several completely connected layers after the convolution layers. In an end-to-end approach, the parameters in these layers can be well trained using large image and video samples labelled for specific tasks. With high adaptability, the qualified CNN can be used to solve classification, recognition, and prediction tasks on test data. The efficiency of CNN-generated prediction signals has surpassed that of rule-based predictors. Furthermore, CNNs can be viewed as feature extractors that convert images and videos into feature space with compact representations, which is advantageous for image and video compression. CNN has also been identified as a promising alternative for compression tasks based on these excellent characteristics. This paper offers a thorough analysis of image and video compression using neural networks in order to better understand the current progress of CNN on image and video compression. We divided the main body of the paper into four sections to make the analysis more transparent due to the broad nature of this study. The basic concepts of neural networks and image/video compression are introduced in section II. Section III examines the evolution of neural network-based image compression techniques in depth. In section IV, we go through video compression techniques based on neural networks. In section V, we return to neural network-based image and video compression optimization techniques. The next part of the rationale follows the timeline of network evolution in order to implement neural network-based image compression based on representative network architectures. In section IV, we primarily discuss the CNN-based video coding techniques embedded in the state-of-the-art hybrid video coding system, HEVC, as well as some modern CNN-based video coding frameworks. Section VI, which concludes the paper, discusses the major challenges in deep learning-based image/video compression.

II. INTRODUCTION OF NEURAL NETWORK AND IMAGE/VIDEO COMPRESSION

First, we'll go through some basic concepts and the history of neural networks in this section. The frameworks and basic methodology implementation for block based image coding and hybrid video coding platform are then introduced.

A. Neural Network

The neural network (NN) was invented as a result of interdisciplinary neuroscience and mathematics studies, and it has shown strong abilities in the sense of non-linear transform and classification. The network, on the surface, appears to be made up of multiple layers of simple processing units called neuron (perceptron), which communicate with one another through weighted connections. Weighted associations from previously active neurons unlock the neurons. The activation functions are often extended for all intermediate layers to achieve non-linearity [29].

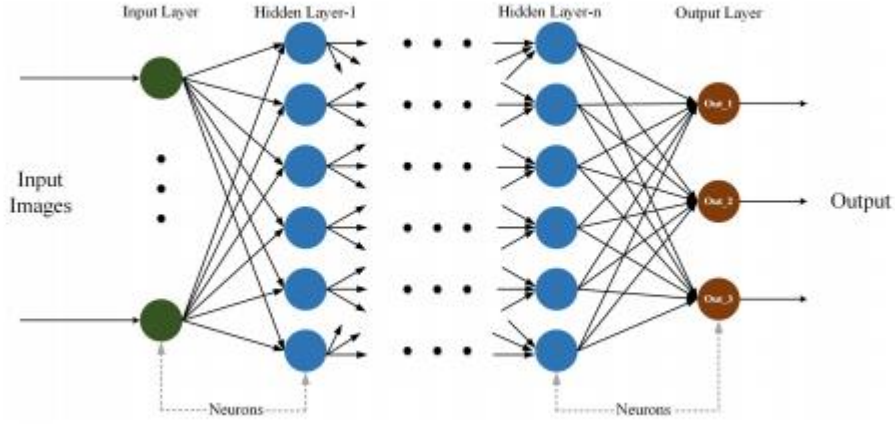

Fig. 2. Illustration of the neural network architecture.

Figure 2 depicts a basic neural network architecture with one input layer, one output layer, and several hidden layers, each with a different number of neurons. In the 1960s, the basic perceptron learning protocol was proposed and studied [30]. During the decade of the 1970s, Backpropagation procedures [31], [32], inspired by the chain rule for derivatives of the training goals, were proposed to solve the multilayer perceptron training problem in the 1970s and 1980s (MLP). Then, stochastic gradient descent with backpropagation is used to train multi-layer architectures, despite the fact that it is computationally intensive and suffers from bad local minima. The dense connections between adjacent layers in neural networks, on the other hand, cause the number of model parameters to quadratically increase, preventing the creation of neural networks in terms of computational efficiency. The implementation of parameter-sharing for MLP 1990 [33] was a game-changer. Convolutional neural networks are a lighter-weighted variant of neural networks that have been proposed and used in document recognition, allowing for large-scale neural network training.

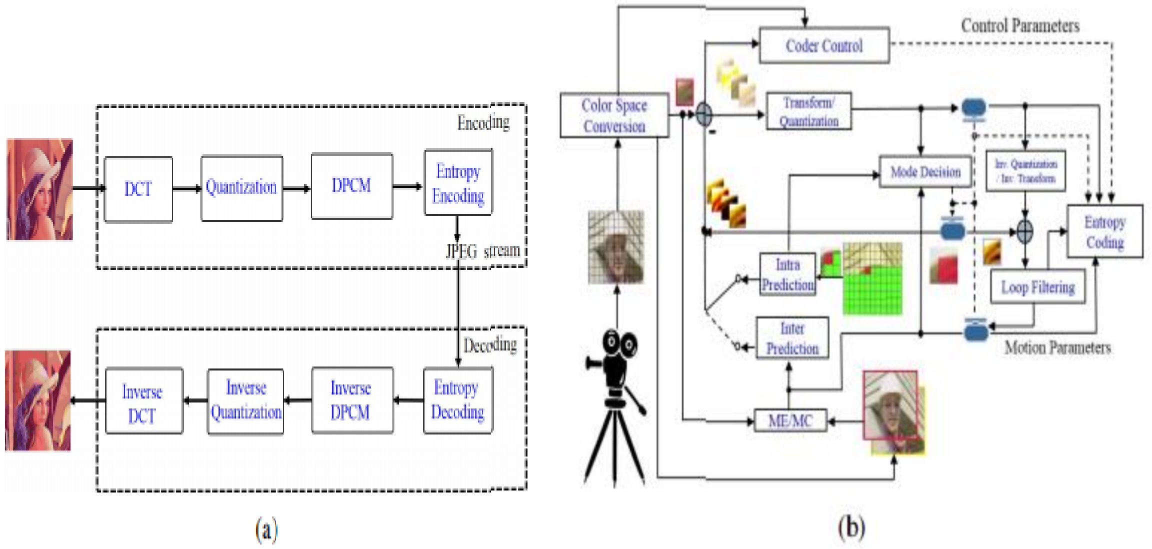

Fig. 3. Image and video compression framework (a) JPEG compression, (b) hybrid video compression

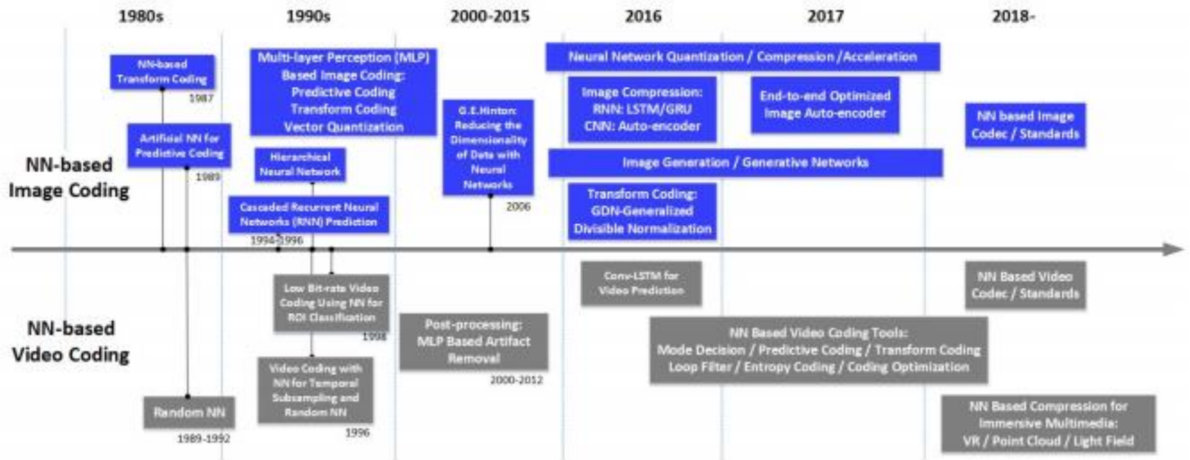
1980s
1990s
2000-2015
2016
2017
2018-
NN-based Image Coding
NN-based Video Coding
NN-based Transform Coding
1987
Artificial NN for Predictive Coding
1989
Multi-layer Perception (MLP) Based Image Coding: Predictive Coding Transform Coding Vector Quantization
Hierarchical Neural Network
Cascaded Recurrent Neural Networks (RNN) Prediction
1994-1996
G.E.Hinton: Reducing the Dimensionality of Data with Neural Networks
2006
Neural Network Quantization / Compression /Acceleration
Image Compression: RNN: LSTM/GRU CNN: Auto-encoder
End-to-end Optimized Image Auto-encoder
NN based Image Codec / Standards
Image Generation / Generative Networks
Transform Coding: GDN-Generalized Divisible Normalization
Low Bit-rate Video Coding Using NN for ROI Classification
1998
Post-processing: MLP Based Artifact Removal
2000-2012
Conv-LSTM for Video Prediction
NN Based Video Codec / Standards
NN Based Video Coding Tools: Mode Decision / Predictive Coding / Transform Coding Loop Filter / Entropy Coding / Coding Optimization
Random NN
1989-1992
Video Coding with NN for Temporal Subsampling and Random NN
1996
NN Based Compression for Immersive Multimedia: VR / Point Cloud / Light Field

Fig. 4. The technical roadmap of neural network based compression algorithms

Backpropagation procedures [31], [32], inspired by the chain rule for derivatives of the training goals, were proposed to solve the multilayer perceptron training problem in the 1970s and 1980s (MLP). Then, stochastic gradient descent with backpropagation is used to train multi-layer architectures, despite the fact that it is computationally intensive and suffers from bad local minima. The dense connections between adjacent layers in neural networks, on the other hand, cause the number of model parameters to quadratically increase, preventing the creation of neural networks in terms of computational efficiency. With the advent of parameter-sharing for MLP 1990 [33], a lighter-weighted version of neural network known as convolutional neural network was proposed and used in document recognition, allowing for large-scale neural network training.

B. Image and Video Compression

The key techniques in image and video compression are transform and prediction, which can be used in a variety of coding frameworks. The most widely used image compression format is JPEG [34], which is made up of the simple transform/prediction modules shown in Fig. 3. (a). The input image is divided into 88 non-overlapping blocks in JPEG, each of which is translated into the frequency domain using block-DCT. The DCT coefficients are then compressed into a binary stream for each transformed block using quantization and entropy coding. Most common video coding standards, such as MPEG-2, H.264/AVC, and HEVC, use a transform-prediction based hybrid video coding system, as shown in Fig. 3(b). Unlike JPEG, HEVC makes use of more intra prediction modes from neighbouring reconstructions of blocks. instead of DC estimation in the spatial domain, as shown in Fig. 1. Aside from intra prediction, high-efficiency inter prediction, which uses motion estimation to find the most similar blocks as a prediction for the to-be-coded block, provides more video compression coding benefits. Furthermore, HEVC uses two loop filters to eliminate compression artefacts sequentially: the deblocking filter and the SAO.

The compression in the above block-based image and video coding standards is typically block-dependent and must be implemented block by block sequentially, limiting the parallelism of compression using parallel computing platforms, such as GPUs. Furthermore, as compared to end-to-end compression, the separate optimization technique for each individual coding method limits the compression efficiency improvement. In essence, as seen in Fig. 4, there is another technical development trajectory focused on neural network techniques for image and video compression. The marriage of conventional image/video compression and CNN accelerates their success with the resurgence of neural networks. The development of neural network-based image/video compression and related representative techniques will be discussed in the sections that follow.

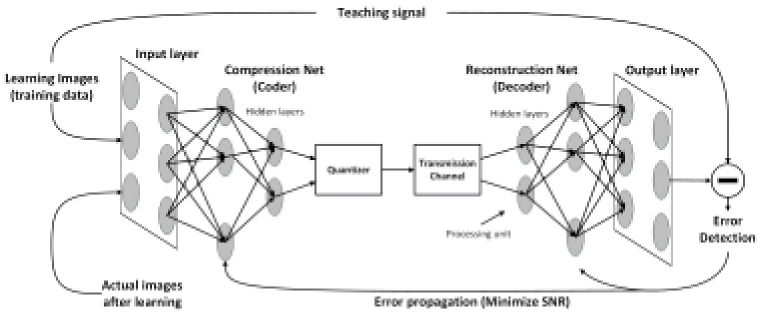

Fig. 5. The Neural Network based image codec [38].

III. PROGRESS OF NEURAL NETWORK BASED IMAGE COMPRESSION

In this section, we introduce image compression using machine learning methods, especially from the perspective of neural networks, which has been around since the late 1980s [35]. The Multilayer Perceptron (MLP), Random Neural Network (RNN), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) are among the neural network techniques covered in this portion. In the final segment, we'll discuss how generative adversarial networks have recently advanced image coding techniques (GAN).

A. Multi-layer Perceptron based Image Coding

MLP [36] is made up of multiple hidden layers of neurons, an input layer of neurons (or nodes, units), and a final layer of output neurons. Each neuron i's output hi inside the MLP is an abbreviation for Multi-Level Programming.

Theoretical research has shown that an MLP with more than one hidden layer can accurately approximate any continuous computable function to any precision [37]. This property provides proof for situations such as data compression and dimension reduction. The aim of using MLP to compress images is to create unitary transformations for the entire spatial data set.

Chua and Lin introduced an end-to-end image compression method in 1988, based on high parallelism and a powerful compact representation of a neural network [35], which could be useful as a model of human brain-like coding functions. They devised the standard image compression steps, i.e. as an integrated optimization problem to minimise the following cost function, the unitary transform of spatial domain image data, quantization of transform coefficients, and binary coding of quantized coefficients

Back propagation was used to train a completely connected neural network with 16 hidden units to compress each 808 patch of an image in 1989 [39]. This technique, however, set the neural network parameters for a fixed number of binary codes, making it difficult to adjust to a variable compression ratio in the ideal state. To compress the input image, Sonehara et al. proposed training a dimension reduction neural network, with quantization and entropy coding as separate modules [38]. The auto-encoder bottleneck structure is used in the design of the dimension reduction neural network (Fig. 5).

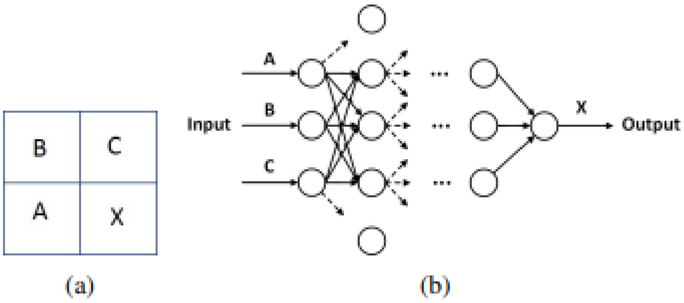

Fig. 6. Illustration of MLP-based predictive coding for image compression [42].

However, the adaptivity of the aforementioned algorithms is calculated by manually changing the number of hidden neurons rather than using networks with more layers and complex connections, which could limit MLP's compression efficiency [41]. To address this problem, an MLP-based predictive image coding algorithm [42] was investigated, which took advantage of spatial background information. To produce the non-linear predictor of the bottom-right pixel X in Fig. 6(a), the spatial information to the left and above (points A, B, and C, each small block orresponds to one pixel in Fig. 6(a)) was used (a). The MLP predictor has three input nodes, 30 hidden nodes, and one output node, as shown in Fig. 6. (b), The MLP model is trained by minimising the mean square errors between the initial and expected signals using the back propagation algorithm [32]. According to their findings, the MLP-based non-linear predictor improves error entropy from 4.7 to 3.9 bits per pixel (bpp) as compared to the linear predictor.

The hierarchical neural network and its Nested Training Algorithm (NTA) were proposed for machine learning in 1996. picture compression [44], which resulted in a significant reduction in training time. [45], [46] provide more information on MLP-based image compression techniques that increase compression efficiency by developing different link structures.

B. Random Neural Network based Image Coding

In 1989, a new form of random neural network [47] was introduced. The above-mentioned MLP-based methods, in which signals are in the spatial domain and optimised by the gradient backpropagation process, perform differently than the random neural network. The signals in a random neural network are transmitted as unit amplitude spikes. Positive signals represent excitatory signals and negative signals represent inhibition in the communication between these neurons, which is modelled as a Poisson mechanism. In [47], some theoretical findings for analysing the behaviour of random neural networks were presented. To update the parameters, a "backpropagation" style training method is used, which involves the solution of n linear and n non-linear equations each time with a new input-output pair.

Some researchers looked at combining the random neural network with image compression and came up with some interesting findings. The random neural network was first used in an image compression task by Gelenbe et al. [48]. A feedforward encoder/decoder random neural network with one intermediate layer is used in the architecture. The first layer, for example, accepts an image as input, the last layer produces a reconstructed image, and the intermediate layer produces compressed bits. Cramer et al. expand on their previous work in [48] by developing an adaptive block-by-block random neural network compression/decompression system [49]. There are several different neural compression networks: C1, ... ,CL which are intended to achieve varying levels of compression Each of these networks compresses the block simultaneously, and the networks are chosen based on the consistency of the decompressed data. Hai enhanced image compression efficiency even further by incorporating a random neural network into the wavelet domain [50].

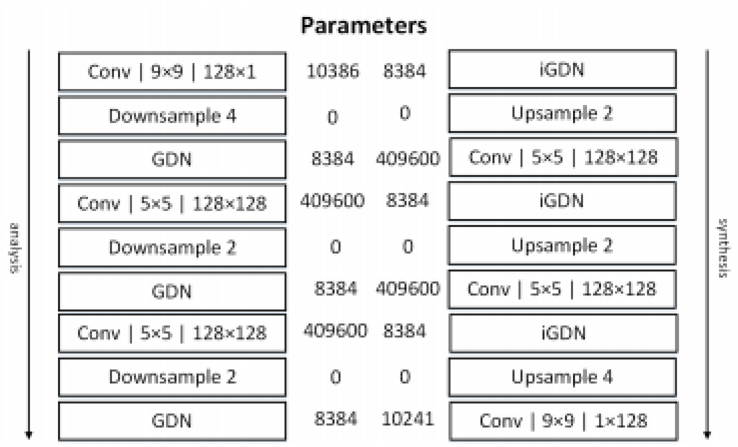

Fig. 7. The parameterized architecture of the CNN based end-to-end image compression proposed in [52].

C. Convolutional Neural Network based Coding

CNN recently outperformed conventional algorithms in high-level computer vision tasks such as image recognition and object detection by a large margin [51]. It also performs admirably for a variety of low-level computer vision tasks, such as super-resolution and compression artefact elimination. The convolution operation is used by CNN to describe the association between adjacent pixels, and the cascaded convolution operations are well-suited to the hierarchical statistical properties of natural images. Furthermore, the convolution operations introduce local receptive fields and mutual weights, which reduce CNN's trainable parameters. As a result, the probability of over-fitting is greatly reduced. Many studies have been conducted to investigate the feasibility of CNN-based lossy image compression, inspired by the powerful representation of CNN for images.

In 2016 [52], [53], Balle et al. published an end-to-end optimised CNN architecture for image compression under the scalar quantization assumption. Figure 7 shows the structure, which is made up of two modules: encoder and decoder analysis and synthesis transforms. Since the synthesis transform is the inverse operation of the analysis transform, all of the parameters in all three stages, h, c, and e, will be optimised in an end-to-end manner using the rate-distortion objective function. Balle et al. used an additive i.i.d uniform noise to simulate the quantizer in the CNN training protocol to deal with the zero derivatives caused by quantization. The stochastic gradient descent approach to the optimization problem is now possible. PSNR and MSSSIM metrics show that this approach outperforms JPEG2000. Furthermore, Balle and his colleagues expanded this model by using scale hyper priors for entropy estimation [54], which resulted in HEVC-like objective coding results. Minnen et al. improved the background model of entropy coding for end-to-end image compression [55], outperforming HEVC intra coding. Since the autoregressive portion is not easily parallelizable, both hardware-end support and energy-efficiency analysis should be further investigated for potential functional utility. Zhou et al. increase image compression efficiency by using a pyramidal feature fusion structure at the encoder and a CNN-based post-processing filter at the decoder [56]. Other end-to-end image compression work involving quantization and entropy coding can be found in [57], [58], and CNN prediction-based image compression in [59].

D. Recurrent Neural Network based Coding

RNN is a type of neural network that has memory to store recent behaviours, unlike the CNN architecture described above. RNN memory modules, in particular, have links to themselves that relay transformed information from previous executions. RNN adjusts the actions of the current forward mechanism to adjust to the current context by using the stored information. To solve the decayed error backflow's insufficiency, Hochreiter et al. suggested the Long Short-Term Memory (LSTM) [60]. More advanced improvements, such as the Gated Recurrent Unit (GRU) [61], are proposed to simplify recurrent evolution processes while maintaining the recurrent network's efficiency in specific tasks [62]. RNN, like CNN, struggles to spread the gradients of the rate estimation for image compression tasks.

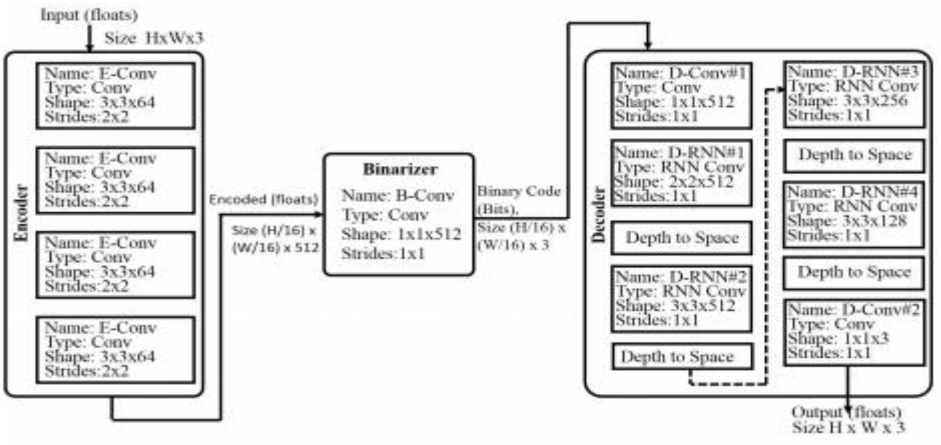

Fig. 8. A single iteration of our shared RNN architecture [63].

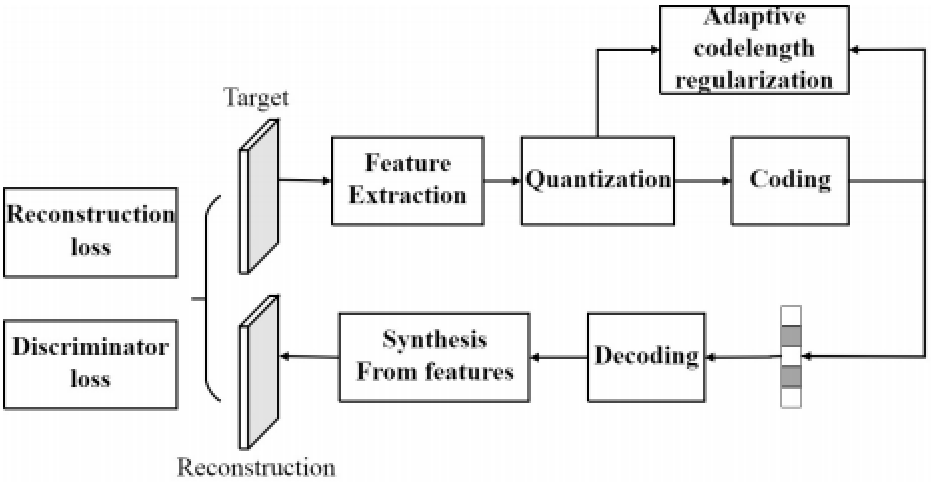

Fig. 9. The overall architecture of GAN based image compression [65]

Toderici et al. proposed the first RNN-based image compression scheme [63], which used a scaled-additive coding framework to limit the number of coding bits rather than the approximation of rate estimation in CNN [52]. In more detail, the proposed approach in [63] is a multi-iteration compression architecture that supports progressive variational bitrate compression..

E. Generative Adversarial Network based Coding

One of the most appealing advancements in deep neural network application is the generational adversarial network. GAN optimises both the generator and discriminator network models at the same time. Discriminator employs a deep neural network to determine whether or not the samples were created by the generator. Simultaneously, the generator is taught how to beat the discriminator and generate samples that pass inspection. Adversarial loss has the benefit of assisting the generator in improving the subjective quality of images while still being adaptable to various tasks. Some research works in the image compression challenge based on the perceptual consistency of the decoded images and used GAN to improve it.

Rippel and Bourdev proposed an integrated and well optimised GAN based image compression in 2017 [65], which not only achieves incredible compression ratio improvements but can also run in real-time by leveraging the huge parallel computing cores of GPU. The input image is compressed into a very compact feature space by networks, as shown in Fig. 9. The decoded image is reconstructed from the features using the generative network. The introduction of adversarial loss, which greatly improves the subjective quality of the restored image, is the most noticeable difference between GAN-based image compression and those of CNN or RNN-based schemes. The generative network and adversarial network are trained together to significantly improve the generative model's efficiency. On generic images of all quality standards, the GAN-based approach in [65] achieves substantial compression ratio improvement, generating compressed files 2.5 times smaller than JPEG and JPEG2000, 2 times smaller than WebP, and 1.7 times smaller than BPG. . The consistency is calculated using MS-SSIM in this case, but the process is still ineffective when using the PSNR metric. The advancements in GAN-based view have inspired me. With the sampled background views in LF [66], the light field (LF) image compression may achieve substantial coding gain with generating the missing views. Specifically, The semantics of the original content are more consistent with the contents produced by GAN than the specific textures. We can see the material difference in particular textures when enlarging the reconstructed images.

In addition, Gregor et al. applied the image compression task to a homogeneous deep generative convolutional model called DRAW [67]. Gregor et al., unlike previous studies, aimed for conceptual compression by producing as much image semantic information as possible [68]. A GAN-based architecture for extreme image compression with bitrates below 0.1 bpp is investigated in depth, allowing for various levels of content generation [69]. At the moment, GAN-based compression works well in narrow-domain images like faces, but further research is needed to develop models for natural images in general.

IV. ADVANCEMENT OF VIDEO CODING WITH NEURAL NETWORKS

The use of the state-of-the-art video coding standard HEVC to analyse deep learning-based video coding has been a hot topic in recent years. Almost all of the HEVC modules have been investigated and developed by using different deep learning models. From the five key modules in HEV, we will review the creation of video coding works with deep learning models in this section .i.e., intra-prediction, inter-prediction, quantization, entropy coding, and loop filtering are all examples of intra-prediction techniques. Finally, we will discuss a number of novel video coding paradigms that are distinct from hybrid video coding system for coding.

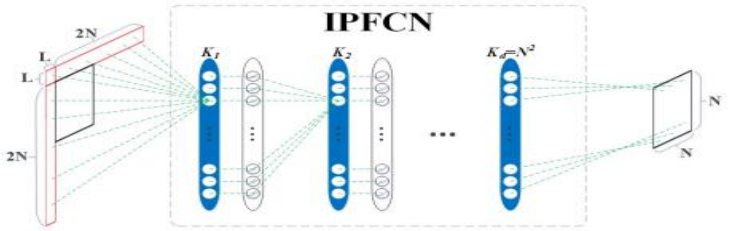

Fig. 10. The network structure of IPFCN [70].

A. Intra Prediction Techniques using Neural Networks

Li et al. proposed a new intra prediction mode using completely connected network (IPFCN) [70] instead of CNN to increase the accuracy of best HEVC intra prediction. competes with the 35 HEVC intra prediction modes that are currently available. IPFCN, like IPCNN, takes contextual feedback from adjacent multiple reference lines of reconstructed pixels. However, the current block's prediction version from HEVC intra prediction is not used. The IPFCN structure is shown in Figure 10, which is an intra prediction mapping from reconstructed neighbouring pixels to the current block. With the exception of theEach connected layer is followed by a nonlinear activation layer, which employs the parametric rectified linear unit (PReLU). Each pixel is represented by a node in the output layer.

[71] proposes a CNN-based chroma intra prediction that uses both the reconstructed luma block and neighbouring chrom blocks to boost intra chroma prediction efficiency. Pfaff et al. suggested a more high-efficiency intra prediction network under JEM programme in [72], and the simplification version's running time only increased by 74% and 38% for intra encoding and decoding processes, respectively, with a 2.26 percent BD-rate saving. Li et al. investigated CNN-based down/upsampling techniques as a new intra prediction mode for HEVC [73], and its extension for inter frame prediction is proposed in [74]. Unlike previous image-level down/upsampling techniques [75], [76], Li et al. developed a down/upsampling approach at the CTU level, as shown in Fig.11. Each CTU is first down-sampled into a low-resolution version, which is then coded using the HEVC intra coding method in down/up-sampling mode. To preserve the original resolution of the restored low resolution CTU, upsampling is used. When the entire frame has been restored, a second stage upsampling CNN network is applied to eliminate the boundary objects. Then there's the upsampling in the second level. CNN has access to all of the down/upsampling CTUs in the region. A flag is signalled into the bitstream to indicate whether down/upsampling is turned on to ensure coding quality. The flag is set in this case based on the encoder's rate distortion optimization.

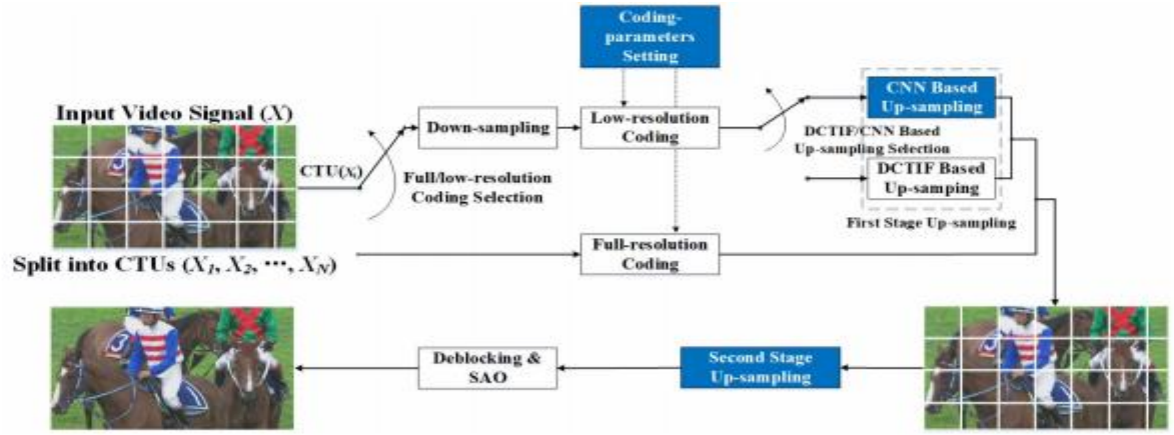

Fig. 11. The framework of neural network based intra prediction using upsampling [73].

This work achieves significant coding benefit, particularly at low bitrate scenarios, thanks to the high efficiency of CNN-based upsampling techniques, saving about 5.5 percent bitrate on average. HEVC stands for High Efficiency Video Coding. The bitrate saving for QPs (=22, 27, 32, 37) used in common HEVC test conditions is only 0.7 percent for the luma portion due to the limitations of the up-sampling algorithm.

We introduced a dual-network dependent super-resolution strategy by bridging the low-resolution image and upsampling network using an enhancement network to minimise the effects of compression noise on upsampling CNN [78]. Compressed noise reduction is the goal of the enhancement network. High-quality input is fed into the upsampling network. The proposed method improves coding performance at low bitrate scenarios, particularly for ultra high resolution videos, when compared to a single upsampling network. In the year 2019, Li et al. developed a compact representation CNN model to boost the super-resolution CNN-based compression system by limiting information loss in low-resolution images [79]. Other CNN-based intra coding strategies can be found in [80], [81], with [80] introducing the CTU level CNN enhancement model for intra coding. in addition [81] introduces RNN-based intra prediction using adjacent reconstructed samples.

B. Neural Network based Inter Prediction

Inter prediction is realised in hybrid video coding by motion estimation on previously coded frames against the current frame, and in HEVC, motion estimation precision is up to quarter-pixel, with the value determined through interpolation, e.g., discrete cosine transform dependent interpolation filter (DCTIF) [17]. Intuitively, the better the coding accuracy, the more similar the inter predicted block and the current block are. This is because there are less prediction residuals left. Huo et al. proposed CNN-based motion compensation refinement (CNNMCR) as a simple method [82] to increase inter prediction performance by using the current variable-filter-size residue-learning CNN (VRCNN) [83]. The motion compensated prediction and its neighbouring reconstructed blocks are fed into the CNNMCR by VRCNN, which is trained by minimising the mean square errors between the input and its corresponding original signal. In reality, CNNMCR's improved inter prediction quality is due to the built network's ability to reduce compression noise and the boundary artifacts improve inter prediction quality. Yan et al. proposed a Fractionalpixel Reference generation CNN (FRCNN) to predict fractional pixels, recognising the importance of fractional-pixel motion compensation in inter prediction [85]. This research differs from previous interpolation or super-resolution studies that predicted pixel sizes. high-resolution picture values, FRCNN, on the other hand, is used to create fractional-pixels from a reference frame in order to reach the current coding frame. As a result, fractional-pixel generation is expressed as a regression problem with the loss function,

As a result, for each fractional-pixel location, a separate CNN is educated. A training example for three half-pixel CNN models is shown in Fig.12. The theory of FRCNN is essentially the same as that of adaptive learning. The parameters of interpolation filters [86], which are derived by minimising prediction errors at fractional-pixel positions and must be transmitted to the decoder side, are derived by minimising prediction errors at fractional-pixel positions.

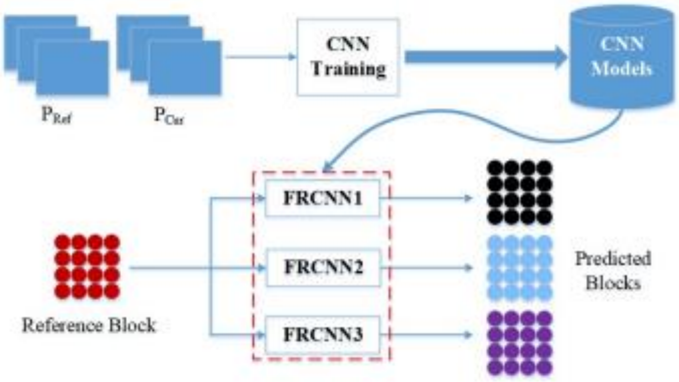

Fig. 12. Framework of the FRCNN in [85]

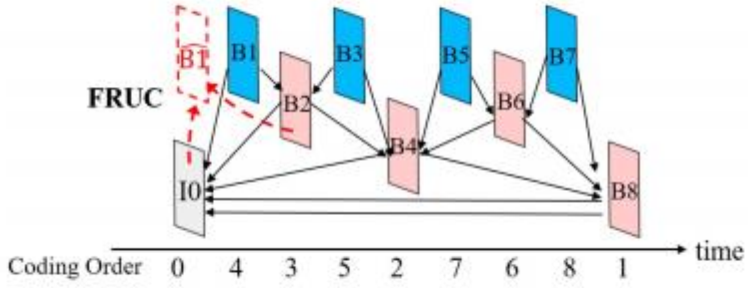

Fig. 13. Illustration for the proposed DVRF mode in [87].

Rather than improving fractional-pixel prediction efficiency, we use CNN-based frame rate up conversion (FRUC) techniques to investigate inter prediction block generation. In the CTU stage, a CNN-based FRUC method is used. [87], [88] proposed to build a virtual reference frame FV irtual, which is used as a new reference frame and is called direct virtual reference frame (DVRF). As shown in Figure 13, the current coding block can use the co-located block in FV irtual as an inter prediction block without having to transfer it. vectors are a type of data. The network is fed with the two closest bidirectional reference frames in the reference list, using the state-of-the-art deep FRUC method Adaptive Separable Convolution [89]., this approach achieves very promising compression efficiency, saving around 4.6 percent bitrate compared to HM-16.9 and 0.7 percent bitrate compared to JEM7.1 [90]. JEM (Joint Exploration Model) is the reference programme for the JVET group, which is working on the exploration of next-generation video coding standards, and is focused on the HEVC reference model.

C. Neural Network based Quantization and Entropy Coding for Video Coding

Quantization and entropy coding are the lossy and lossless compression procedures of video coding, respectively. Due to its low computational and memory costs, the scalar quantization approach has dominated hybrid video coding frameworks. This uniform scalar quantization does not match the features of the human visual system, and it is not conducive to improving perceptual efficiency. Alam et al. suggested a two-stage quantization strategy based on neural networks in [95]. A CNN called VNet-2 is used in the first phase to predict local visibility. The VNet-2 consists of 894 trainable parameters in three layers, namely convolution, subsampling, and complete link, each of which includes one feature map, and the threshold CT for HEVC distortions of individual video frames.

In comparison to HEVC, the proposed adaptive quantization strategy will save on average 11% bitrate for the luma channel while maintaining the same perceptual efficiency. structural similarity index (SSIM) [96]. Song et al. enhanced the CABAC output on compressing the syntax elements of 35 intra prediction modes by leveraging the prediction efficiency of CNN. Instead of using handcrafted background models, CNN was used to explicitly estimate the probability distribution of intra modes [97]. The network architecture is based on LeNet-5 [98], which was suggested by LeCun et al. Puri et al. used a CNN to predict the best transform index probability distribution from quantized coefficient blocks, and then used the probability to binarize the transform index. To improve entropy coding accuracy, the transform index uses variable length instead of fixed length coding [99].

D. Neural Network based Loop Filtering

Since H.263+ [100], the loop filtering module has been included in video coding standards, and several different loop filters [21]–[23], [27], [28] have been suggested. Many CNN-based loop filters have recently been developed to eliminate compression artefacts, inspired by CNN's success in the image/video restoration field. When compared to other video coding modules, they are much easier to execute the end-to-end instruction. [101] Zhang et al. suggested a Loop filtering in HEVC using a residual highway convolutional neural network (RHCNN). We improved the efficiency of the CNN-based loop filter and designed the spatial-temporal residue network by exploiting the coherence of the spatial and temporal adaptations. Loop filter based on (STResNet) [102].

Furthermore, in [103], we introduced a content-aware CNN-based loop filter that improved the filtering efficiency even more. Multiple CNN models are iteratively trained according to their filtering output for a reconstructed frame, as described in [104], and a corresponding discriminative network is also trained to aid in the selection of the optimal filter in the test stage to eliminate coding overheads. Under the HEVC common test condition (CTC), the proposed multi-model CNN filters achieve major improvements over HEVC with/without ALF. Even with the GeForce GTX TITAN X GPU, They proposed an effective solution for memory-efficient CNN-based loop filters in [105]. By simply padding the scalar QPs into a matrix with the same size of input frames or patches, they combined QPs as an input fed into the CNN training stage. [106] proposes a residual prediction CNN model for in-loop filters, and [107] proposes a multi-scale CNN model for in-loop filters. To delete compression artefacts, Dong et al. suggested an end-to-end CNN [108] that is learned in a supervised manner. The CNN architecture is derived from the SRCNN [109] super-resolution network by adding one or more "feature enhancement" layers after the first layer of SRCNN to clean up the noisy features. Yang et al. suggested a multi-frame quality enhancement neural network for compressed video that uses neighbouring high-quality frames to improve low-quality frames. A support vector is shown here. In compressed video, a machine-based detector is used to locate peak quality frames [111]. In the area of multiview plus depth video coding, CNN-based quality enhancement also performs admirably. CNN models were created by Zhu et al. to improve 3D video coding efficiency by post-processing synthesised views [112]. More works [113], [114] used more complex structures to boost compressed images.

E. New Video Coding Frameworks Based on Neural Network

Chen et al. proposed Deep Coder, a CNN network that combined multiple CNN networks to achieve comparable perceptual consistency to a low-profiled x264 encoder [115]. Deep Coder is a programming language that allows you to create complex programmes. The intra prediction is obtained from motion estimation on previous frames and is implemented using a neural network to produce a feature map, denoted as fMap. By incorporating the idea of Voxel CNN and exploring spatial-temporal coherence to efficiently perform video coding, Chen et al. proposed a truly learning-based video coding system. Within a learning network, predictive coding [116]. Srivastava et al. proposed to use the Long Short Term Memory (LSTM) Encoder-Decoder method to predict future frames of generative models [117], which was inspired by the prediction for future frames of generative models [117]. In [118], learn video representations that can be used to predict potential video frames. The LSTM Auto encoder Model and the LSTM Future Predictor Model are the two main models, both of which are based on recurrent neural networks. . Unlike Ranzato's [117] work, which could only predict one future frame, this model can predict a long future series. Based on the results of the experiments with 16 natural video frames as data, The model is capable of reconstructing these 16 frames as well as predicting the next 13 frames.

V. OPTIMIZATION TECHNIQUES FOR IMAGE AND VIDEO COMPRESSION

HEVC, a cutting-edge video coding standard, achieves optimum compression efficiency by exhaustively traversing all feasible coding modes and partitions to decide the best one. Rate-distortion costs are used to determine the best coding parameters. By predicting the optimal coding parameters and avoiding excessive RD calculations, the computational costs can be drastically reduced. On the basis of neural networks, fast mode-decision algorithms for the coding unit (CU) and prediction unit (PU) are proposed, which are not only parallel-friendly but also efficient. But it's also simple to design VLSI [119], [120]. To decide if the mode decision should be terminated early, Xu et al. used both CNN and LSTM to predict the entire CTU partition structure [121].

VI. CONCLUSIONS AND OUTLOOK

We believe that the advantages of neural networks in image and video compression are threefold, based on our study. First, neural networks outperform signal processing-based models in terms of content adaptively because network parameters are extracted from a large amount of real-world data, while models in the state of the art coding standards are handcrafted based on image and video prior information. Second, larger receptive fields are commonly used in neural network models, which not only use neighbouring information but can also boost performance. coding efficiency by leveraging samples from a long distance, but conventional coding tools only used nearby samples and it's difficult to use samples from a long distance. Third, since the neural network can accurately represent both texture and function, it can be used for both human and computer vision research. However, the current state of affairs Only high compression efficiency for human view tasks is pursued by coding standards.

Image and video compression aims to reduce the size of visual signals while maintaining high quality, and is becoming increasingly relevant in the era of large visual data. In The neural network-based image and video compression techniques, especially the recent deep learning-based image and video compression techniques, have been reviewed in this paper.

We believe that deep learning-based image/video compression can play a larger role in representing and delivering images and videos with higher quality and lower bitrates in the future, and that the following issues must be addressed. to be investigated further:

Picture and video compression with semantic fidelity. With the rapid advancement in computer vision techniques and the exponential growth of photographs and videos, visual signal receivers are no longer limited to humans. computer vision algorithms, as well as the visual system.

For the compression mission, rate-distortion (RD) optimization driven neural network training and adaptive switching. Traditional image and video compression relies on the rate-distortion principle, but it has limitations. Present neural network-based compression tasks haven't been thoroughly investigated.

Develop for a realistic image and video codec that is memory and computation efficient. The burdens in computation and memory are the most significant impediment to the implementation of deep learning-based image and video compression. Larger neural networks with more layers and nodes are normally considered to achieve high efficiency.

We attempted to develop a groundbreaking visual signal representation system to elegantly help both human vision viewing and computer vision processing for semantically friendly image and video compression. We suggested the hierarchical approach because of the lightweight and relevance of features for visual semantic descriptors, such as CNN features. [122] compresses the function descriptors and visual content together to reflect a visual signal. In [123], we looked into the novel visual signal representation structure in conjunction with a deep learning-based end-to-end image compression platform that can perform further image comprehension tasks directly from the compression domain. The reason for this approach is that the neural network architectures currently used for learned compression are not suitable for this task (in particular the encoders). The previous work [124] suggested a complexity-distortion optimization formulation for the video coding problem under power constraints, which can be improved upon. Combined with computational costs and video compression efficiency, CNN model compression optimization has been expanded.

Moreover, network-based end-to-end optimization techniques are more adaptable than hand-crafted methods, and they can be used to solve a variety of problems. The network can be quickly programmed or tuned, which gives it a lot of potential for future image and video compression problems, as well as other artificial intelligence problems.

REFERENCES

1. [1] D. A. Huffman, “A method for the construction of minimumredundancy codes,” Proceedings of the IRE, vol. 40, no. 9, pp. 1098– 1101, 1952.
2. [2] S. Golomb, “Run-length encodings (Corresp.),” IEEE Trans. on information theory, vol. 12, no. 3, pp. 399–401, 1966.
3. [3] I. H. Witten, R. M. Neal, and J. G. Cleary, “Arithmetic coding for data compression,” Communications of the ACM, vol. 30, no. 6, pp. 520–540, 1987.
4. [4] H. Andrews and W. Pratt, “Fourier transform coding of images,” in Proc. Hawaii Int. Conf. System Sciences, 1968, pp. 677–679.
5. [5] W. K. Pratt, J. Kane, and H. C. Andrews, “Hadamard transform image coding,” Proceedings of the IEEE, vol. 57, no. 1, pp. 58–68, 1969.
6. [6] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transform,” IEEE Trans. on Computers, vol. 100, no. 1, pp. 90–93, 1974.
7. [7] C. Harrison, “Experiments with linear prediction in television,” Bell System Technical Journal, vol. 31, no. 4, pp. 764–783, 1952.
8. [8] G. K. Wallace, “Overview of the JPEG (ISO/CCITT) still image compression standard,” in Image Processing Algorithms and Techniques, vol. 1244. International Society for Optics and Photonics, 1990, pp. 220–234.
9. [9] C. Christopoulos, A. Skodras, and T. Ebrahimi, “The JPEG2000 still image coding system: an overview,” IEEE trans. on consumer electronics, vol. 46, no. 4, pp. 1103–1127, 2000.
10. [10] D. Taubman, “High performance scalable image compression with EBCOT,” IEEE Trans. on image processing, vol. 9, no. 7, pp. 1158– 1170, 2000.
11. [11] Y. Taki, M. Hatori, and S. Tanaka, “Interframe coding that follows the motion,” Proc. Institute of Electronics and Communication Engineers Jpn. Annu. Conv.(IECEJ), p. 1263, 1974.

12. [12] A. Netravali and J. Stuller, “Motion-Compensated Transform Coding,” Bell System Technical Journal, vol. 58, no. 7, pp. 1703–1718, 1979.
13. [13] C. Reader, “History of Video Compression (Draft),” document JVTD068, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), 2002.
14. [14] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Trans. on circuits and systems for video technology, vol. 13, no. 7, pp. 560–576, 2003.
15. [15] “AVS working group website,” http://www.avs.org.cn, Accessed Aug. 2018.
16. [16] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” IEEE Trans. on circuits and systems for video technology, vol. 22, no. 12, pp. 1649– 1668, 2012.
17. [17] H. Lv, R. Wang, X. Xie, H. Jia, and W. Gao, “A comparison of fractional-pel interpolation filters in HEVC and H. 264/AVC,” in Visual Communications and Image Processing (VCIP), 2012, pp. 1–6.
18. [18] J. Lainema, F. Bossen, W.-J. Han, J. Min, and K. Ugur, “Intra coding of the HEVC standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1792–1801, 2012.
19. [19] J.-L. Lin, Y.-W. Chen, Y.-W. Huang, and S.-M. Lei, “Motion vector coding in the HEVC standard,” IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 6, pp. 957–968, 2013.
20. [20] M. Naccari and F. Pereira, “Adaptive bilateral filter for improved inloop filtering in the emerging high efficiency video coding standard,” in IEEE Picture Coding Symposium (PCS), 2012, pp. 397–400.
21. [21] X. Zhang, R. Xiong, W. Lin, J. Zhang, S. Wang, S. Ma, and W. Gao, “Low-rank-based nonlocal adaptive loop filter for high-efficiency video compression,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 27, no. 10, pp. 2177–2188, 2017.
22. [22] S. Ma, X. Zhang, J. Zhang, C. Jia, S. Wang, and W. Gao, “Nonlocal in-loop filter: The way toward next-generation video coding?” IEEE MultiMedia, vol. 23, no. 2, pp. 16–26, 2016.
23. [23] C.-Y. Tsai, C.-Y. Chen, T. Yamakage, I. S. Chong, Y.-W. Huang, C.-M. Fu, T. Itoh, T. Watanabe, T. Chujoh, M. Karczewicz et al., “Adaptive Loop Filtering for Video Coding,” IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 6, pp. 934–945, 2013.
24. [24] X. Zhang, R. Xiong, S. Ma, and W. Gao, “Adaptive loop filter with temporal prediction,” in IEEE Picture Coding Symposium (PCS), 2012, pp. 437–440.
25. [25] X. Zhang, S. Wang, Y. Zhang, W. Lin, S. Ma, and W. Gao, “HighEfficiency Image Coding via Near-Optimal Filtering,” IEEE Signal Processing Letters, vol. 24, no. 9, pp. 1403–1407, 2017.
26. [26] P. List, A. Joch, J. Lainema, G. Bjontegaard, and M. Karczewicz, “Adaptive deblocking filter,” IEEE Trans. on circuits and systems for video technology, vol. 13, no. 7, pp. 614–619, 2003.
27. [27] A. Norkin, G. Bjontegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. Van der Auwera, “HEVC deblocking filter,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1746–1754, 2012.
28. [28] C.-M. Fu, E. Alshina, A. Alshin, Y.-W. Huang, C.-Y. Chen, C.-Y. Tsai, C.-W. Hsu, S.-M. Lei, J.-H. Park, and W.-J. Han, “Sample AdaptiveOffset in the HEVC Standard,” IEEE Trans. on Circuits and Systems for Video technology, vol. 22, no. 12, pp. 1755–1764, 2012.
29. [29] G. E. Hinton, “Learning translation invariant recognition in a massively parallel networks,” in International Conference on Parallel Architectures and Languages Europe. Springer, 1987, pp. 1–13.
30. [30] F. Rosenblatt, “Principles of neurodynamics,” 1962. [31] P. Werbos, “New Tools for Prediction and Analysis in the Behavioral Sciences,” Ph. D. dissertation, Harvard University, 1974.
31. [32] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” Nature, vol. 323, no. 6088, p. 533, 1986.
32. [33] Y. Le Cun, O. Matan, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. Jacket, and H. S. Baird, “Handwritten zip code recognition with multilayer networks,” in Proceedings 10th International Conference on Pattern Recognition. IEEE, 1990, pp. 35–40.
33. [34] G. K. Wallace, “The JPEG still picture compression standard,” Communications of the ACM, vol. 34, no. 4, pp. 30–44, 1991.
34. [35] L. Chua and T. Lin, “A neural network approach to transform image coding,” International Journal of Circuit Theory and Applications, vol. 16, no. 3, pp. 317–324, 1988.

35. [36] M. W. Gardner and S. Dorling, “Artificial neural networks (the multilayer perceptron)-a review of applications in the atmospheric sciences,” Atmospheric environment, vol. 32, no. 14-15, pp. 2627–2636, 1998.
36. [37] R. J. Schalkoff, Artificial neural networks. McGraw-Hill New York, 1997, vol. 1.
37. [38] N. Sonehara, M. Kawato, S. Miyake, and K. Nakane, “Image data compression using a neural network model,” in Proc. IJCNN, vol. 2, 1989, pp. 35–41.
38. [39] P. Munro and D. Zipser, “Image compression by back propagation: an example of extensional programming,” Models of cognition: rev. of cognitive science, vol. 1, no. 208, p. 1, 1989.
39. [40] G. Sicuranza, G. Romponi, and S. Marsi, “Artificial neural network for image compression,” Electronics letters, vol. 26, no. 7, pp. 477–479, 1990.
40. [41] R. D. Dony and S. Haykin, “Neural network approaches to image compression,” Proceedings of the IEEE, vol. 83, no. 2, pp. 288–303, 1995.
41. [42] S. Dianat, N. Nasrabadi, and S. Venkataraman, “A non-linear predictor for differential pulse-code encoder (DPCM) using artificial neural networks,” in International Conference on Acoustics, Speech, and Signal Processing, ICASSP 1991, pp. 2793–2796.
42. [43] C. Manikopoulos, “Neural network approach to DPCM system design for image coding,” IEE Proceedings I (Communications, Speech and Vision), vol. 139, no. 5, pp. 501–507, 1992.
43. [44] A. Namphol, S. H. Chin, and M. Arozullah, “Image compression with a hierarchical neural network,” IEEE Trans. on Aerospace and Electronic Systems, vol. 32, no. 1, pp. 326–338, 1996.
44. [45] J. G. Daugman, “Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression,” IEEE Trans. on acoustics, speech, and signal processing, vol. 36, no. 7, pp. 1169–1179, 1988.
45. [46] H. Abbas and M. Fahmy, “Neural model for Karhunen-Loeve transform with application to adaptive image compression,” IEE Proceedings I (Communications, Speech and Vision), vol. 140, no. 2, pp. 135–143, 1993.
46. [47] E. Gelenbe, “Random neural networks with negative and positive signals and product form solution,” Neural computation, vol. 1, no. 4, pp. 502–510, 1989.
47. [48] E. Gelenbe and M. Sungur, “Random network learning and image compression,” in IEEE International Conference on Neural Networks (ICNN), vol. 6, 1994, pp. 3996–3999.
48. [49] C. Cramer, E. Gelenbe, and I. Bakircioglu, “Video compression with random neural networks,” in Neural Networks for Identification, Control, Robotics, and Signal/Image Processing, International Workshop on. IEEE, 1996, pp. 476–484.
49. [50] F. Hai, K. F. Hussain, E. Gelenbe, and R. K. Guha, “Video compression with wavelets and random neural network approximations,” in Applications of Artificial Neural Networks in Image Processing VI, vol. 4305. International Society for Optics and Photonics, 2001, pp. 57–65.
50. [51] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature, vol. 521, no. 7553, p. 436, 2015.
51. [52] J. Balle, V. Laparra, and E. P. Simoncelli, “End-to-end optimized image ´ compression,” arXiv preprint arXiv:1611.01704, 2016.
52. [53] ——, “End-to-end optimization of nonlinear transform codes for perceptual quality,” in Picture Coding Symposium (PCS), 2016, pp. 1–5.
53. [54] J. Balle, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, “Vari- ´ ational image compression with a scale hyperprior,” in International Conference on Learning Representations, 2018.
54. [55] D. Minnen, J. Balle, and G. D. Toderici, “Joint autoregressive and ´ hierarchical priors for learned image compression,” in Advances in Neural Information Processing Systems 31. Curran Associates, Inc., 2018, pp. 10 771–10 780.
55. [56] L. Zhou, C. Cai, Y. Gao, S. Su, and J. Wu, “Variational Autoencoder for Low Bit-rate Image Compression,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 2617–2620.
56. [57] E. Agustsson, F. Mentzer, M. Tschannen, L. Cavigelli, R. Timofte, L. Benini, and L. V. Gool, “Soft-to-hard vector quantization for endto-end learning compressible representations,” in Advances in Neural Information Processing Systems, 2017, pp. 1141–1151.
57. [58] L. Theis, W. Shi, A. Cunningham, and F. Huszar, “Lossy im- ´ age compression with compressive autoencoders,” arXiv preprint arXiv:1703.00395, 2017.
58. [59] E. Ahanonu, M. Marcellin, and A. Bilgin, “Lossless Image Compression Using Reversible Integer Wavelet Transforms and Convolutional Neural Networks,” in IEEE Data Compression Conference, 2018.

59. [60] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.
60. [61] K. Cho, B. Van Merrienboer, D. Bahdanau, and Y. Bengio, “On the ¨ properties of neural machine translation: Encoder-decoder approaches,” arXiv preprint arXiv:1409.1259, 2014.
61. [62] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” arXiv preprint arXiv:1412.3555, 2014.
62. [63] G. Toderici, D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor, and M. Covell, “Full Resolution Image Compression with Recurrent Neural Networks,” in CVPR, 2017, pp. 5435–5443.
63. [64] D. Minnen, G. Toderici, M. Covell, T. Chinen, N. Johnston, J. Shor, S. J. Hwang, D. Vincent, and S. Singh, “Spatially adaptive image compression using a tiled deep network,” arXiv preprint arXiv:1802.02629, 2018.
64. [65] O. Rippel and L. Bourdev, “Real-time adaptive image compression,” arXiv preprint arXiv:1705.05823, 2017.
65. [66] C. Jia, X. Zhang, S. Wang, S. Wang, S. Pu, and S. Ma, “Light field image compression using generative adversarial network based view synthesis,” IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2018.
66. [67] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra, “Draw: A recurrent neural network for image generation,” arXiv preprint arXiv:1502.04623, 2015.
67. [68] K. Gregor, F. Besse, D. J. Rezende, I. Danihelka, and D. Wierstra, “Towards conceptual compression,” in Advances In Neural Information Processing Systems, 2016, pp. 3549–3557.
68. [69] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. Van Gool, “Extreme Learned Image Compression with GANs,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 2587–2590.
69. [70] J. Li, B. Li, J. Xu, R. Xiong, and W. Gao, “Fully Connected NetworkBased Intra Prediction for Image Coding,” IEEE Trans. on Image Processing, 2018.
70. [71] Y. Li, L. Li, Z. Li, J. Yang, N. Xu, D. Liu, and H. Li, “A Hybrid Neural Network for Chroma Intra Prediction,” in 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018, pp. 1797–1801.
71. [72] J. Pfaff, P. Helle, D. Maniry, S. Kaltenstadler, B. Stallenberger, P. Merkle, M. Siekmann, H. Schwarz, D. Marpe, and T. Wiegan, “Intra prediction modes based on neural networks ,” in JVET-J0037. ISO/IEC JTC/SC 29/WG 11, Apr. 2018, pp. 1–14.
72. [73] Y. Li, D. Liu, H. Li, L. Li, F. Wu, H. Zhang, and H. Yang, “Convolutional neural network-based block up-sampling for intra frame coding,” IEEE Trans. on Circuits and Systems for Video Technology, 2017.
73. [74] J. Lin, D. Liu, H. Yang, H. Li, and F. Wu, “Convolutional Neural Network-Based Block Up-Sampling for HEVC,” IEEE Transactions on Circuits and Systems for Video Technology, 2018.
74. [75] R. Molina, A. Katsaggelos, L. Alvarez, and J. Mateos, “Toward a new video compression scheme using super-resolution,” in Visual Communications and Image Processing (VCIP), vol. 6077. International Society for Optics and Photonics, 2006, p. 607706.
75. [76] M. Shen, P. Xue, and C. Wang, “Down-sampling based video coding using super-resolution technique,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 21, no. 6, pp. 755–765, 2011.
76. [77] J. Pfaff, P. Helle, D. Maniry, S. Kaltenstadler, W. Samek, H. Schwarz, D. Marpe, and T. Wiegand, “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.
77. [78] L. Feng, X. Zhang, X. Zhang, S. Wang, R. Wang, and S. Ma, “A Dual-Network based Super-Resolution for Compressed High Definition Video,” in Pacific-Rim Conference on Multimedia. Springer, 2018, pp. 600–610.
78. [79] Y. Li, D. Liu, H. Li, L. Li, Z. Li, and F. Wu, “Learning a Convolutional Neural Network for Image Compact-Resolution,” IEEE Transactions on Image Processing, vol. 28, no. 3, pp. 1092–1107, 2019.
79. [80] Z.-T. Zhang, C.-H. Yeh, L.-W. Kang, and M.-H. Lin, “Efficient CTUbased intra frame coding for HEVC based on deep learning,” in AsiaPacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2017, pp. 661–664.
80. [81] Y. Hu, W. Yang, S. Xia, W.-H. Cheng, and J. Liu, “Enhanced Intra Prediction with Recurrent Neural Network in Video Coding,” in IEEE Data Compression Conference (DCC), 2018, pp. 413–413.

81. [82] S. Huo, D. Liu, F. Wu, and H. Li, “Convolutional Neural NetworkBased Motion Compensation Refinement for Video Coding,” in International Symposium on Circuits and Systems (ISCAS). IEEE, 2018, pp. 1–4.
82. [83] Y. Dai, D. Liu, and F. Wu, “A convolutional neural network approach for post-processing in HEVC intra coding,” in International Conference on Multimedia Modeling. Springer, 2017, pp. 28–39.
83. [84] J. Liu, S. Xia, W. Yang, M. Li, and D. Liu, “One-for-all: Grouped variation network-based fractional interpolation in video coding,” IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2140–2151, 2019.
84. [85] N. Yan, D. Liu, H. Li, B. Li, L. Li, and F. Wu, “Convolutional Neural Network-Based Fractional-Pixel Motion Compensation,” IEEE Trans. on Circuits and Systems for Video Technology, 2018.
85. [86] Y. Vatis and J. Ostermann, “Adaptive interpolation filter for H.264/AVC,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 19, no. 2, pp. 179–192, 2009.
86. [87] L. Zhao, S. Wang, X. Zhang, S. Wang, S. Ma, and W. Gao, “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.
87. [88] ——, “Enhanced Motion-compensated Video Coding with Deep Virtual Reference Frame Generation,” submitted to IEEE Trans. on Image Processing, 2018.
88. [89] S. Niklaus, L. Mai, and F. Liu, “Video frame interpolation via adaptive separable convolution,” arXiv preprint arXiv:1708.01692, 2017.
89. [90] J. Chen, E. Alshina, G. J. Sullivan, J.-R. Ohm, and J. Boyce, “Algorithm Description of Joint Exploration Test Model 1,” in JVET-A1001. ISO/IEC JTC/SC 29/WG 11, Oct. 2015, pp. 1–48.
90. [91] Z. Zhao, S. Wang, S. Wang, X. Zhang, S. Ma, and J. Yang, “CNNBased Bi-Directional Motion Compensation for High Efficiency Video Coding,” in International Symposium on Circuits and Systems (ISCAS), 2018, pp. 1–4.
91. [92] ——, “Enhanced Bi-prediction with Convolutional Neural Network for High Efficiency Video Coding,” to be appear in IEEE Trans. on Circuits and Systems for Video Technology, 2018.
92. [93] H. Zhang, L. Song, Z. Luo, and X. Yang, “Learning a convolutional neural network for fractional interpolation in HEVC inter coding,” in Visual Communications and Image Processing (VCIP), 2017, pp. 1–4.
93. [94] N. Yan, D. Liu, H. Li, T. Xu, F. Wu, and B. Li, “Convolutional Neural Network-Based Invertible Half-Pixel Interpolation Filter for Video Coding,” in 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018, pp. 201–205.
94. [95] M. M. Alam, T. D. Nguyen, M. T. Hagan, and D. M. Chandler, “A perceptual quantization strategy for HEVC based on a convolutional neural network trained on natural images,” in Applications of Digital Image Processing XXXVIII, vol. 9599. International Society for Optics and Photonics, 2015, p. 959918.
95. [96] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. on image processing, vol. 13, no. 4, pp. 600–612, 2004.
96. [97] R. Song, D. Liu, H. Li, and F. Wu, “Neural network-based arithmetic coding of intra prediction modes in HEVC,” in Visual Communications and Image Processing (VCIP), 2017, pp. 1–4.
97. [98] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
98. [99] S. Puri, S. Lasserre, and P. Le Callet, “CNN-based transform index prediction in multiple transforms framework to assist entropy coding,” in Signal Processing Conference (EUSIPCO), European, 2017, pp. 798–802.
99. [100] G. Cote, B. Erol, M. Gallant, and F. Kossentini, “H.263+: Video coding at low bit rates,” IEEE Transactions on circuits and systems for video technology, vol. 8, no. 7, pp. 849–866, 1998.
100. [101] 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.
101. [102] C. Jia, S. Wang, X. Zhang, S. Wang, and S. Ma, “Spatial-temporal residue network based in-loop filter for video coding,” in Visual Communications and Image Processing (VCIP), 2017, pp. 1–4.
102. [103] C. Jia, S. Wang, X. Zhang, J. Liu, S. Pu, S. Wang, and S. Ma, “Content-Aware Convolutional Neural Network for In-loop Filtering in High Efficiency Video Coding,” Accepted by IEEE Trans. on Image Processing, 2019.
103. [104] X. Zhang, S. Wang, K. Gu, W. Lin, S. Ma, and W. Gao, “Just-noticeable difference-based perceptual optimization for JPEG compression,” IEEE Signal Processing Letters, vol. 24, no. 1, pp. 96–100, 2017.

104.[105] X. Song, J. Yao, L. Zhou, L. Wang, X. Wu, D. Xie, and S. Pu, “A practical convolutional neural network as loop filter for intra frame,” arXiv preprint arXiv:1805.06121, 2018.

105.[106] W.-S. Park and M. Kim, “CNN-based in-loop filtering for coding efficiency improvement,” in Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), 2016, pp. 1–5.

106.[107] J. Kang, S. Kim, and K. M. Lee, “Multi-modal/multi-scale convolutional neural network based in-loop filter design for next generation video codec,” in International Conference on Image Processing (ICIP), 2017, pp. 26–30.

107.[108] C. Dong, Y. Deng, C. Change Loy, and X. Tang, “Compression artifacts reduction by a deep convolutional network,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 576–584.

108.[109] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in European conference on computer vision. Springer, 2014, pp. 184–199.

109.[110] K. Li, B. Bare, and B. Yan, “An efficient deep convolutional neural networks model for compressed image deblocking,” in International Conference on Multimedia and Expo (ICME), 2017, pp. 1320–1325.

110.[111] R. Yang, M. Xu, Z. Wang, and T. Li, “Multi-Frame Quality Enhancement for Compressed Video,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6664–6673.

111.[112] L. Zhu, Y. Zhang, S. Wang, H. Yuan, S. Kwong, and H. H.-S. Ip, “Convolutional neural network-based synthesized view quality enhancement for 3d video coding,” IEEE Transactions on Image Processing, vol. 27, no. 11, pp. 5365–5377, 2018.

112.[113] L. Cavigelli, P. Hager, and L. Benini, “CAS-CNN: A deep convolutional neural network for image compression artifact suppression,” in International Joint Conference on Neural Networks (IJCNN). IEEE, 2017, pp. 752–759.

113.[114] B. Zheng, R. Sun, X. Tian, and Y. Chen, “S-Net: a scalable convolutional neural network for JPEG compression artifact reduction,” Journal of Electronic Imaging, vol. 27, no. 4, p. 043037, 2018.

114.[115] T. Chen, H. Liu, Q. Shen, T. Yue, X. Cao, and Z. Ma, “DeepCoder: A deep neural network based video compression,” in Visual Communications and Image Processing (VCIP), 2017 IEEE, pp. 1–4.

115.[116] Z. Chen, T. He, X. Jin, and F. Wu, “Learning for Video Compression,” arXiv preprint arXiv:1804.09869, 2018.

116.[117] M. Ranzato, A. Szlam, J. Bruna, M. Mathieu, R. Collobert, and S. Chopra, “Video (language) modeling: a baseline for generative models of natural videos,” arXiv preprint arXiv:1412.6604, 2014.

117.[118] N. Srivastava, E. Mansimov, and R. Salakhudinov, “Unsupervised learning of video representations using LSTMS,” in International conference on machine learning, 2015, pp. 843–852.

118.[119] Z. Liu, X. Yu, Y. Gao, S. Chen, X. Ji, and D. Wang, “CU partition mode decision for HEVC hardwired intra encoder using convolution neural network,” IEEE Trans. on Image Processing, vol. 25, no. 11, pp. 5088–5103, 2016.

119.[120] N. Song, Z. Liu, X. Ji, and D. Wang, “CNN oriented fast PU mode decision for HEVC hardwired intra encoder,” in IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2017, pp. 239–243.

120.[121] M. Xu, T. Li, Z. Wang, X. Deng, R. Yang, and Z. Guan, “Reducing Complexity of HEVC: A Deep Learning Approach,” IEEE Trans. on Image Processing, 2018.

121.[122] X. Zhang, S. Ma, S. Wang, X. Zhang, H. Sun, and W. Gao, “A joint compression scheme of video feature descriptors and visual content,” IEEE Trans. on Image Processing, vol. 26, no. 2, pp. 633–647, 2017.

122.[123] Y. Li, C. Jia, X. Zhang, S. Wang, S. Ma, and W. Gao, “Joint rate-distortion optimization for simultaneous texture and deep feature compression of facial images,” in IEEE International Conference on Multimedia Big Data (BigMM), 2018, pp. 334–341.

123.[124] L. Su, Y. Lu, F. Wu, S. Li, and W. Gao, “Complexity-constrained H.264 video encoding,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 19, no. 4, pp. 477–490, 2009.