Review of Post-processing Methods Based on Image/Video Compression Artifact Removal

Abstract

We review post-processing algorithms for single image and video compression artifact removal in detail and evaluate the algorithms using commonly used datasets. Compression artifact is a visual quality problem caused by high-frequency information quantization after lossy compression of image/video. Post-processing algorithms are most widely used in the field of artifact removal, which can effectively improve visual quality and speed up computer processing. In this paper, we systematically summarize the evolution and framework of artifact removal methods from two perspectives of traditional and deep algorithms, and highlight the main improvements and research significance of these methods. Regarding the direction of deep learning that we focus on, we start with two aspects: JPEG-based single image processing algorithms and HEVC-based video processing algorithms. We summarize the evolution of the model from all aspects, including the improvement of the network structure, the utilization of transform domain information and the loss function. In addition, we use common datasets and evaluation metrics to compare and analyze key deep learning algorithms. This paper summarizes the algorithm framework for the subsequent research on related artifact removal, and puts forward development suggestions for related fields.

Keywords: JPEG, HEVC, Artifact Removal, Post-processing

1 Introduction

With the accelerated development of wireless networks, the requirement for image and video transmission increases greatly. Captured images and videos are often compressed to reduce information redundancy, in order to minimize storage capacity and transmission bandwidth. Joint Photographic Experts Group (JPEG) [41] and HEVC (High Efficiency Video Coding) [42], as well as other existing lossy compression algorithms, have been extensively explored to achieve this goal. They are effective in reducing bitrate costs but unavoidably lead to unsatisfying visual artifacts, such as blocking artifacts, ringing artifacts, etc. These artifacts not only lead to degraded viewing, but also further degrade the performance of downstream computer vision systems. Therefore, compression artifact removal is a critical post-processing task.

Current artifact removal algorithms can be mainly classified into traditional algorithms and deep learning-based methods. According to the literature [103], the existing traditional algorithms are divided into deblock-oriented, and recovery-oriented methods, and how to deal with time artifacts in video. The deblock-oriented methods focus on eliminating blocking and ringing artifacts and mainly use digital filters. Early preliminary work was done on filtering based on specific regions such as edges to remove simple artifacts, and attempts were made on various transform domains. For example, a shape-adaptive DCT filtering approach has been proposed by Foi et al. [10] to reduce compression artifacts. The recovery-oriented methods, on the other hand, treat the compression operation as some kind of distortion and put forward the corresponding recovery algorithms. In order to eliminate artifacts, these methods propose to learn the inverse mapping of the compression degradation. These methods mainly include the projections onto convex sets (POCS) [40], [45]-[47], Markov random field (MRF) [48]-[51], Sparse Representation [52]-[55], regression tree fields (RTF) [56]-[58] and so on. Temporal artifacts are caused by the video encoder's inability to similarly encode the same spatial region in the preceding and following frames[102]. The existing methods are generally based on various filters to complete video processing.

In this paper, algorithms based on deep learning are mainly considered JPEG and HEVC encoding. JPEG is widely used as a natural image compression standard. HEVC represents the highest standard for natural image and video compression. These two standards use a block-by-block compression scheme to convert blocks into conversion coefficients. These DCT coefficients are then coarsely quantified to eliminate high-frequency information to save storage space. The quantization step is the major cause of artifacts, which leads to the loss of high-frequency details. After being quantized, the boundaries between adjacent blocks become non-contiguous. Therefore, there is a blocking effect. Blurring is induced by the loss of high-frequency information. In the region containing the sharp edge, the ringing artifact becomes visible.

The original deep learning technique was first used for image artifact removal in 2015 [17]. With the deepening of research, the improvement of JPEG compression algorithm is mainly reflected in the improvement of network structure, such as dense residual network [18], the use of transform domain information[29]-[30], novel loss functions, and so on.

In 2013, HEVC officially became an internationally recognized video coding standard. At the same time, this indicates that video compression technology is to a new level. HEVC has introduced two intra-loop filtering methods to minimize artifacts, namely the deblocking filter (DF) [59] and the sampling adaptive offset (SAO) [60]. The intra-loop filter can sufficiently reduce many artifacts, especially the blocking artifacts. In previous years, most of the research on JPEG compression methods focused on blocking artifacts. Therefore, the removal of artifacts for JPEG is not effective in HEVC compression. In addition, although filters can improve the quality and compression performance of each frame, the introduction of filters will reduce the quality of compressed video.

Since the introduction of HEVC, researchers first tried to perform cyclic filtering to reduce compression artifacts. [61]-[63] proposes an alternative to the intra-loop filter. However, this modifies the HEVC encoder. Therefore, enhancing the quality of encoded video is impractical [104].

In recent years, some post-processing methods based on learning have been proposed. Since the video compression algorithm adopts different encoding modes, namely I/P/B frames, various strategies are adopted to process every frame. According to the literature [105], based on the number of input frames, existing methods can be classified into single-frame based approaches and multi-frame based approaches.

The single-frame based approaches aim to enhance the quality of each frame [64]-[69],[70]. Although these approaches play a certain role in image and video enhancement, the spatio-temporal relevance between adjacent frames is ignored and thus the capability is limited. The MFQE method used multiple frames as input for the first time and achieved better performance. Later, more multi-frame based approaches were proposed [72]-[80]. There are also approaches that capture space-time information from nearby frames in other ways. For example,[82],[83] adopts deep Kalman filtering method,[78],[84] adopts branch ConvLSTM and variable convolution method.

This review covers several papers in recent years on the post-processing of image and video compression for artifact removal. The main contributions of this paper are summarized as follows:

1. This review summarizes the post-processing method of image and video compression and artifact removal.

2. This review summarizes the treatment of different compression methods by traditional methods and deep learning methods. Finally, the performance of these methods is evaluated from both qualitative and quantitative analysis.

3. This review presents some potential challenges and points out directions of post-processing methods for image and video compression and artifact removal.

The rest of this review is organized as followed.

Section 2 discusses the traditional methods of image and video compression to remove artifacts. Section 3 introduces the deep learning methods for natural image/video compression and artifact removal post-processing technology. Section 4 provides a comparison of the effects of these artifact removal methods. Section 5 discusses the potential challenges and further research directions in this area. Section 6 gives the author's conclusions.

2 Traditional Algorithms

2.1 Deblock-oriented Algorithms

Deblock-oriented algorithms mainly include various digital filters and can be roughly divided into spatial domain and frequency domain processing.

2.1.1 Spatial Domain Processing

In the spatial domain, most of the existing researches have proposed various adaptive filters based on specific regions. The initial spatial filtering algorithm selects the filter type by the boundary features of image blocks. Filtering types include horizontal and vertical filtering in one dimension or spatial adaptive filtering in two dimensions.

The simplest deblocking method is that Reeve et al. [1] applied a 3×3 Gaussian filter to the pixels at the boundary of the image block. Later, Minami and Zakhor[3] attempted to adaptively infer the parameters of the filtering operation by reducing the expected value of the MSDS.

The above method is to use different deblocking strategies in each area to prevent excessive blurring of edges. Other spatial algorithms include post filtering used in the shift window of the image block [12], nonlinear spatial variable filtering [2], adaptive non-local mean filtering [4], adaptive bilateral filtering [11], etc.

2.1.2 Frequency Domain Processing

Working in the transform domain, the transform coefficients of each image block are adjusted directly, taking into account the artifacts caused by the rough quantization.

The ideas of this kind of method include using the characteristics of human visual system such as masking effect[5], using the correlation between the boundary pixel values of adjacent blocks[13, 14], and taking into account the pixel difference in different regions of the adaptive algorithm[15, 16].

At the same time, existing studies have been explored in a variety of different transformation domains. Chen et al.[5]used an adaptive weighting mechanism to filter the block DCT coefficients. Kim[6] and Wu[7] et al. process in the wavelet domain. The former uses the directional filter and the latter uses the soft threshold to deblock. Xu et al. [8] computed block activity according to transformation coefficients in the Hadamard transform (DHT) domain to adaptively classify different regions and filter them.

In addition, Lee et al. [9] reduced artifacts by predicting images with a linear regression model in the transform domain after using low-pass filtering. Foi et al. [10] use point-by-point shape-adaptive DCT (SA-DCT) for denoising, which is considered to be an effective method in the field of deblocking.

2.2 Recovery-oriented Algorithms

2.2.1 Projections Onto Convex Sets (POCS)

Zakhor et al. [46] introduce signals with known transformation coefficients and quantization coefficients as constraints. Based on convex set projection theory, [46] propose a new iterative block reduction technique. The algorithm in [46] uses POCS [40] to prove its convergence, but the smooth set used by the algorithm is not strictly defined. Reeves et al. [45] discuss this and explain the algorithm in [46] by the constrained minimization for providing the theoretical basis. Y. Yang et al. [47] propose two regularization reconstruction methods to reduce blocking artifacts, and strictly proved the convergence of these two algorithms.

2.2.2 Markov Random Field (MRF)

Thomas P. O 'Rourke [49] et al. transform the image decompression problem into an ill-posed inverse problem and form a well-posed construction algorithm by using the random regularization technique. In [49], the original image is modeled as huber

Markov random field (MRF) and adjust the coded image by referring to the model in the quantization constraint set (QCS). Based on [49], Meier et al [48] proposed a post-processing algorithm based on MRF. It effectively eliminates the blocking artifacts and does not introduce new artifacts. Robertson and Stevenson[50] describe statistically the spatial-domain quantization noise in the image compressed by DCT and find that the correlation Gaussian noise model is more accurate. Sun and Cham[51] model the distortion of coding and the original image as additive Gaussian noise and Markov random fields respectivelye.

2.2.3 Sparse Representation

Farinella and Battiato[52] propose a novel method of blocking artifacts removal by using structural sparse coding model selection (SSMS). For the first time, the sparse representation has been shown to reduce blocking artifacts effectively. Based on [52], Jung et al. [53] use the K-singular value decomposition (K-SVD) algorithm to obtain a generic dictionary. The dictionary is from a set of training images and is beneficial to remove blocking artifacts from JPEG compressed images. Chang et al. [54] find a sparse representation on the dictionary which is attached to the recovered image. In [53] and [54], a set of training images is required and the JPEG images to be processed must have known compression factors. For solving this problem, Chiou et al. [55] propose a new image/video deblocking framework. Through sparse representation, deblocking is expressed as a morphological component analysis based image decomposition problem.

2.2.4 Regression Tree Fields (RTF)

Jancsary et al. [57] first propose regression tree fields (RTF), a random field model for perfect conditions of image labeling problems. Based on [57], they introduce a new image recovery framework [56] and extended it to make it feasible to optimize any differentiable loss function. New techniques of image denoising and JPEG deblocking are developed.

Schmidt et al. [58] deduced a discriminant model based on the RTF [56]. By using this discriminant prediction cascade, the problem of direct regression of appropriate parameters from input images can be overcome. Extending [58], other image restoration problems can be solved.

2.3 Time Domain Artifact Processing Algorithms

Many compressed video artifact removal efforts [1]-[3] have focused on reducing blocking and ringing artifacts. However, in decompressed videos after high intensity compression, the temporal rendering in the specified or assigned bandwidth is not good enough and time domain artifacts are still evident.

Temporal artifacts generally originate from the obvious discontinuities between the inner frame and the adjacent inter-prediction frames. For this reason, it can also be referred to as static region fluctuations[47] and flickering artifacts[**?**]. Fan et al. [96] found that videos compressed in the full encoding mode produce more severe time domain artifacts compared to the periodic intra-encoding mode.

To reduce the time artifact, some researchers modified the encoder [96]-[98], and also tried the pre-filtering/post-filtering method.

In the post-processing method, [99] proposes to adopt an interframe based on a motion-compensated version generated from its previous interframe and used to filter the interframe. [100] proposes a fuzzy filter that adapts to the activity of samples and the relative position between samples. Maximum a posteriori (MAP) method is also applied to video noise reduction [101] while implementing the Huber time filter for reducing fluctuation artifacts. [102] proposes a robust statistical time filter (RSTF) that depends on both spatial and temporal neighborhoods.

3 Deep Learning Algorithms

Since JPEG and HEVC are the two most widely studied coding methods in the field of image/video compression. Therefore, in the following discussion of deep learning methods, we will mainly consider the above two compression methods separately.

3.1 Image Artifact Removal Based on JPEG Compression Method

The proposal of ARCNN, a three-layer neural network based on CNN, makes deep learning algorithms begin to be applied to the field of artifact removal. According to [106], subsequent improvements in JPEG compression artifact reduction mainly focus on three important aspects: network structure improvement, transform domain information utilization, and novel loss function. Below, we review the exploration of deep learning from the above three points. All the algorithms are summarized in Table 1.

3.1.1 Network Structure Improvement

Many high-level networks are built to mine more complex structural relationships. In this way, more network architectures are proposed and applied to artifact processing, such as the CNN paradigm, encoder-decoder network, and GAN.

Among them, feedforward denoising CNN (DnCNN) in [65] has a deeper network architecture. Through the residual learning strategy, [65] can deal with the unknown noise level of Gaussian denoising and JPEG image deblocking. Encoder-decoder Network (RED-Net)[60] and CAS-CNN[61] utilize the fully convolutional coding structure, which symmetrically skip the convolutional and deconvolution layers. [21] proposes trainable nonlinear reaction-diffusion networks (TNRD), which update filters and influence functions based on losses.

The above network cannot meet the long-term dependence problem because of its deep architecture. [22] proposes a persistent memory network (MemNet), which adds recurrent units and gate units to store state information in order to enhance the connections between state layers.

In addition, there are some models that introduce traditional prior knowledge and constraints, including multi-scale constraints, sparsity, and wavelet signal structure. Depth residuals autoencoder at [25] uses prior knowledge of Residual-in-Residual

Method	Published	Inference Model	Priors/Side Information	Basic Idea
ARCNN	ICCV-2015 [7]	Three-layer CNN	/	The first work introducing deep models to the topic
DnCNN	TIP-2017 [28]	CNNwith residual learing and batch normalization	/	The combination of residual learing, batch normalization,and Adam optimization
RED-net	NIPS-2016 [29]	Encoder-decoder with skip connections	/	An encoder-decoder with symmetric skip connections
CAS-CNN	LJCNN-2017 [30]	Encoder-decoder with skip connections	Multi-scale losses	An encoder-decoder constrained by multi-scale losses
TNRD	TPAM1-2017 [27]	Trainable nonlinear diffusion model	/	The proposed nonlinear diffusion model unrolling into a deep network
MemNet	ICCV-2017 [32]	DenseNet architecture Memoryblock	Multi-supervision Long-term memory	The network is stacked by memory blocks, consisting of a recursive unit and a gate unit, to leamn explicit persistent memories
Autoencoder for depth residuals	IEEE Access-2019	Residual-in-Residual Dense Blocks (RRDB)	Prior knowledge of the JPEG compression pipeline	The performed JPEG restoration in two phases using two different autoencoders
DCSC	ICCV-2019	Classic learned iterative shrinkage-threshold algorithm	/	The method integrates model-based convolutional sparse coding with a learning-based deep neural network
G-CAR	2022	The cascaded residual encoder-decoder networks	/	The method first predicts QF by analyzing luminance patches with high activity.
ReWaGAN	2019	Improved GAN	/	The first to use an end-to-end generative adversarial network
WCDGAN	IEEE Access-2022	Improved GAN	/	The method has three main ingredients of mixed convolution, weakly connected dense block
D' Model	CVPR-2016	Leamed iterative shrinkage and thresholding with DCT layers	DCT domain constraint Sparsity constraint	The proposed nonlinear diffusion model unrolling into a deep network
DMCNN	ICIP-2018	A two-branch auto-encoder with dilated convolution	DCT domain constraint Multi-scale loss	It integrates the dual domain architecture(DCT and spatial domains),DCT loss and multi-scale loss
DDCN	ECCV-2016 [9]	A two-branch CNN	Range of DCT coeficients	A two-branch CNN works in pixel and DCT domains and finally aggregates their information
MWCNN	CVPRW-2018	Encoder-decoder with Wavelet transforms	Wavelet signal structure	Wavelet transforms are introduced into CNN architecture
DPW-SDNet	CVPRW-2018[35]	A two-branch CNN with Wavelet transforms	Dual domains Wavelet signal structure	A two-branch CNN is constructed to make use of both redudenancy in pixel and frequency domains
One-to-Many Network	CVPR-2017 [31]	ResNet Shift-and-average strategy	Perceptual loss Adversarial loss JPEG loss	A ResNet takes input as random noise and a compressed image, and its output is constrained by three losses

Table 1: Overview of key models for JPEG compression artifact removal.

Dense Blocks (RRDB) and JPEG compression pipes to perform JPEG recovery using two different autoencoders for luminance channel and chrominance channel.

The network structure can also be blind and generic, effectively restoring images of any compression level. For example, the adaptive Blind Depth Convolutional Sparse Coding (DCSC) network [24] uses sparse coding and dilated convolutional layers to extract multi-scale features from images. To enhance generality, (G-CAR) [43] utilizes a rectifier network and a modified Quality Factor (QF) estimator to predict the QF value and adaptively reconstruct the image in the encoder-decoder network.

Generative Adversarial Network (GAN) is also applied in artifact processing [26][27], such as rectified wasserstein GAN (ReWaGAN) [28] is based on WGAN, adding paired constraint improvement training. WCDGAN[44] proposes a generative adversarial network with weakly connected dense blocks, embedded with a hybrid attention mechanism and perceptual loss.

3.1.2 Transform Domain Information Utilization

Because of the redundant information that is ignored in the JPEG encoder, image conversions such as DCT and DWT can be embedded to promote the work of JPEG artifact removal. Deep bidomain (D^3) [29] introduces the idea of dual-domain for the first time, and proposes a deep model based on the prior knowledge of sparse coding and JPEG compression.

In addition, there is the dual-domain CNN (DMCNN) [30] which uses an autoencoder-style architecture with dilated convolution and multi-scale loss to expand the receptive field. DDCN[31] adopts the strategy of Adam and residual learning to build a deep dual-domain feedforward architecture. MWCNN[32] was based on the improved U-Net network, and Wavelet transform was introduced into the contraction subnet and the expansion subnet to balance performance and efficiency. In DPW-SDNet[33], two-branch CNN deals with pixel and recovery in discrete wavelet domains.

3.1.3 Novel Loss Function

To achieve good visual effects, many works focus on proposing various complete and efficient loss functions. These include adversarial loss[34, 35], perceptual loss[36, 37], edge emphasis loss[38], loss function based on JPEG compression principle[39] and structural similarity loss[93]. For example, one-to-many network [20] uses per-pixel l2 loss between output and true value to train one-to-one mapping, and uses perceptual loss and natural loss to enhance image quality.

3.2 Image Artifact Removal Based on HEVC Compression Method

According to [107], there are two ways to reduce the compression artifacts of HEVC, i.e. perform circular filtering and post-processing. Intra-loop filtering is performed in the encoding loop, and post-processing tools are used only at the decoder side for better reconstruction quality.

For intra-loop filtering, Han et al. [62] propose a high-performance in-loop filter placed after the original in-loop filter in HEVC. Many learning-based intra-loop filters have been proposed. Park and Kim[61] propose a CNN as an alternative to the SAO filter in HEVC. Huang et al. [63] remove DBF and SAO filters in HEVC and propose a CNN-based frame-level filtering method based on an FQE-CNN neural network. However, all of the above methods require modifications to the HEVC encoder. Therefore, it is almost impossible to improve the quality of the encoded video. In the following content, we focus on the post-processing algorithm to reduce the HEVC compression artifact and summarize it in Table 2.

3.2.1 Single Frame Video Processing

Earlier studies treat individual frames of video as images and focus on improving the quality of the image [64]-[69], [70]. Based on ARCNN[17], Dai et al [69] introduce a VRCNN for artifact removal of HEVC decoded frames. However, the internal compressed image must be recovered with DF and SAO removed. Yang et al. [66] propose DS-CNN, which can enhance the HEVC compression frame without considering the encoding mode. MPRGAN [70] uses the generative adversarial network to solve the video post-processing task that includes scalable structures for reducing computing resource footprint. For higher HEVC compression video quality at the decoder side, DCAD[64] proposes a CNN-based depth auto-decoder, which contains ten convolutional layers. However, [64] can not reduce the artifacts of I and P/B frames, because

Method	Published	Inference Model	Basic Idea
VRCNN	MMM-2017 [36]	Vaniable-filter-size Residue-learning	The designed CNN owns vanable filter size to learn the
DCAD	2017 Data Compression Conference	10 convolutional layers aoto decoder	a fully end-to-end feed forward network,generate the higher quality
Decoder-side Scalable Convolutional Neural	ICME-2017	CNN with two sub-networks	computational complexity is adjustable to the changing computational resources
quality enhancement convolutional neural network(QE-CNN)	TCSVT-2019	CNN with two sub-networks	the first CNN-based method designed to learn the distortion features of intra-and inter-coding
multi-level progressive refinement network via an adversarial training approach(MPRGAN)	ICASSP-2018	composed of two CNNs, the enhancement network and the discriminative network.	using multi-level progressive refinement network to make trade-off between enhanced quality and computational
Multi-Frame CNN(MFQE 1.0)	CVPR-2018	CNN with two sub-networks and a SVM-based PQF detector	reduce the compression artifacts of non-PQFs(Peak Quality Frames) by making use of the neighboring PQFs.
Multi-Frame Quality Enhancement (MFQE 2.0)	TPAMI-2019	CNN with two sub-networks and a BiLSTM-based PQF detector	improve the former network architecture
Learning-based Multi-frame Video Enhancement(LMVE)	ICIP-2019	two network for HF,MF and LF separately	use high-quality and moderate-quality frames to enhance the low-quality
SDTS-based network	ICIP-2019	network combine with the motion compensation, multi-frame fusion mode, and	use Res Slice Blockto extract useful information progressively
MFQE+	VCIP-2021	add Hierarchical Residual blocks to MFQE	Video quality enhancement without alignment
Quality-Gated Convolutional Long Short-Term Memory (QG-	ICME-2019	spatial-temporal structure adopts the bi-directional ConvLSTM structure	proposed learning the forget and input gates in ConvLSTM from quality
Spatio-Temporal Deformable Convolution for Compressed Video Quality Enhancement	AAAI-2020	composed of a Spatio-Temporal Deformable Fu-sion (STDF) module and a Quality Enhancement (QE) module.	incorporates a novel STDF scheme to aggre-gate temporal information.
Coarse-to-Fine Spatio-Temporal Information Fusion (CF-STIF)	IEEE Signal Processing Letters	composed of The Coarse Fusion Module, Multi-Level Residual Fusion Module, Deformable Convolution Fusion Module, and The	combining 3D convolutional layers with the proposed MLRF to get better deformable offsets
Recursive Fusion and Deformable Spatiotemporal Attention for Video Compression Artifact Reduction	ACM Multimedia 2021	composed of the Spatio-Temporal Feature Fusion (STFF) and the Quality Enhancement Module	propose a recursive fusion module to exploit more spa-tiotemporal information

Table 2: Overview of key models for HEVC compression artifact removal.

the difference between in-code and inter-code distortion is not considered. DS-CNN[66] proposes two subnets, namely DS-CNN-I and DS-CNN-B, to reduce the artifacts of I-frames and P/B frames respectively.

Although the above methods can be used for image and video enhancement, they all ignore the similarity of adjacent frames in the video. Moreover, the one-frame approach does not take advantage of temporal information in the video.

In addition, multiple artifacts can be reduced by DF and SAO sufficiently. Therefore, applying the existing deblocking CNN to every frame, the PSNR of the HEVC/AVC decoded video improve can not be improved significantly and flicker can not be alleviated, since the frames are processed independently. It becomes more difficult to gain through post-processing based on a single frame. For reducing inter-frame redundancy effectively, many multi-frame based methods are proposed.

3.2.2 Multiframe Video Processing

In 2018, the first multi-frame quality enhancement (MFQE) method [71] was proposed. [71] achieves better performance than the single-frame method. Later, more multi-frame based methods keep coming[72]-[89], [94].

[80] uses recurrent neural networks to retain more information. [78] proposes a multifunctional video coding (VVC) quality enhancement network for compressed video, combining spatial detail and temporal structure (SDTS).

In addition, [82], [83] model video restoration as a Kalman filtering process to exploit the previous frames' spatio-temporal information. [78] use non-local ConvLSTM while [79] use deformable convolution to acquire the dependence between multiple adjacent frames.

In the above multi-frame based methods, the spatio-temporal dependence of adjacent frames is utilized for detail reconstruction. According to [79], these methods can be broadly classified into two groups: aligning-based or unaligned.

For the unaligned method, LSTM[94] models frame dependencies, then aggregate prior information for building higher quality frames, e.g. MGANet[85] and NL-ConvLSTM[84]. FastMSDD[86] propose a fast multi-scale depth decoder for HEVC. Because of the implicit use of similar information, the results of this approach are not very satisfactory. [75] proposes an unaligned method (MFQE+) for multi-frame video quality enhancement and gets good results.

For the alignment-based method, [71] does alignment based on optical flow. However, inaccurate pixel correspondences cannot be dealt with effectively because the optical flow is estimated for motion compensation. The deformable convolution-based alignment method can reduce the problem to some extent. Follow this thread, the STDF scheme [78] and the recursive fusion and deformable spatio-temporal attention (RFDA) method [80] achieve the most advanced performance. However, the predicted alignment offset is inaccurate in [78] because Only spatio-temporal information of 2D convolutions is extracted. To predict better deformable offset, [79] proposes coarse-Fine spatio-temporal information fusion (CF-STIF). Following this technical line, Peng et al. [88] propose a cyclic deformation fusion method.

However, the information propagated from adjacent frames is not necessarily correct, and to control the information flow, an efficient fusion operation is requisite. Shi et al. [89] designed a loop network to maintain time consistency of compressed video quality enhancement. Scale and offset factors were introduced to redesign feature alignment in Basicvrs ++[90]

4 Experiment

4.1 JPEG

4.1.1 Datasets

To validate the performance of the proposed method in reducing JPEG compression artifacts, two well-known datasets, LIVE1[93] and BSDS500[91] are commonly used. The LIVE1 dataset contains 29 original images and 779 distorted images, with a total of 5 distortion types. The BSD500 is a dataset provided by the computer vision group at Berkeley University that can be used for image segmentation and object edge detection. The dataset contains 200 training images, 100 validation images, and 200 test images Quantitative evaluation is performed on 29 images from the LIVE 1 and 200 test images from the BSDS500.

4.1.2 Evaluation Metrics

PSNR, PSNR-B (which focuses the assessment on blocking artifacts), and SSIM metrics are the globally adopted metrics for assessing image quality in artifact removal tasks. Image quality information in terms of noise and perceived quality is provided by the PSNR and PSNR-B metrics, and PSNR-B further takes into account blocking artifacts; The SSIM index is a metric used to describe the similarity of the quality of structures and edges contained in the image. For these three metrics, larger PSNR and PSNR-B, or SSIM closer to 1, indicates that the structure and content in the reconstructed image are closer to those in the original image.

4.1.3 Objective Evaluation

On the BSDS500 and LIVE 1 datasets, we evaluated 16 representatives of the latest algorithms: Artifact removal CNN(AR-CNN)[17], deep dual domain JPEG artifact removal D^3[29], dual domain Convolutional Network (DDCN)[31], CAS-CNN[19], trained nonlinear reactive diffusion (TNRD)[21], de-noising CNN(DnCNN)[65], one-to-many network Baseline[20], persistent memory network (MemNet)[22], Multilevel wavelet -CNN(MWCNN)[32], two-domain Multi-scale CNN(DMCNN)[30], two-pixel - wavelet domain depth CNN(DPW-SDNet)[33], scalable convolutional Neural Network (S-NET)[23], Deep convolutional sparse coding (DCSC) [24], artifact removal GAN (Argan-mse [27], ARGAN[27]), deep Residual encoder (Deep Residual) [25], Multi-domain Residual Codec Network (G-CAR) [43], Weakly connected Dense Generative Adversation Network (WCDGAN) [44]. The results are from the reports of corresponding papers.

The baseline we chose covered most of the representative methods. [17] is the first CNN-based deblocking algorithm. Subsequently,[65],[20]-[25],[27],[43]-[44] improved the network structure, [29]-[33] makes use of transform domain information.

We report the evaluation of different JPEG quality factors of 10, 20, and 40, in PSNR, PSNR-B, and SSIM. All the training and evaluation procedures are processed on grayscale images (Y channel in YCbCr space).

As can be seen from Table 3, DAR[25] obtained the most advanced methodology improvements in PSNR and SSIM on the LIVE1 dataset. Meanwhile, the G-CAR framework [43] consistently performs best on perceived quality (PSNR-B) that takes blocking artifacts into account. Specifically, G-CAR results are better than DnCNN [65] for unknown QF. DAR[25] consistently obtains the best or suboptimal PSNR, PSNR-B, and SSIM on the BSDS500 dataset.

4.1.4 Subjective Evaluation

In this section, we select six important algorithms for qualitative comparison and present visual examples in Figure 1.

Dataset	LIVE1			BSDS500		
Quality	10	20	40	10	20	40
JPEG	27.77|25.33|0.791	30.07|27.57|0.868	32.35|29.96|0.917	27.80|25.10|0.788	30.05|27.22|0.867	32.89|–|0.926
AR–CNN[26]	29.13|28.74|0.823	31.40|30.69|0.890	33.63|33.12|0.931	28.56|28.54|0.783	30.42|30.39|0.852	33.55|32.80|0.930
DDCN[41]	–	–	–	29.59|29.18|0.838	31.88|31.10|0.900	34.27|33.15|0.939
D3[39]	29.96|29.45|0.823	32.21|31.35|0.890	–	–	–	–
CAS–CNN[29]	29.44|29.19|0.833	31.70|30.88|0.895	34.10|33.68|0.937	–	–	–
TNRD[31]	29.15|28.88|0.811	31.46|31.04|0.877	–	28.43|28.30|0.781	30.35|30.16|0.854	33.73|32.79|0.932
DnCNN[27]	29.19|28.91|0.812	31.59|31.08|0.880	34.07|33.24|0.904	28.84|28.44|0.783	31.05|30.29|0.857	–
Baseline[30]	–	–	–	29.56|29.10|0.835	31.89|31.04|0.898	–
MemNet[32]	29.45|–|0.819	31.83|–|0.885	–	29.50|28.60|0.835	31.34|29.84|0.889	33.23|31.04|0.928
MWCNN[42]	29.37|28.85|0.832	31.58|30.83|0.891	34.17|33.33|0.936	–	–	–
DMCNN[40]	29.73|29.55|0.842	32.09|31.32|0.905	–	29.67|–|0.840	31.98|–|0.904	–
S–NET[33]	29.87|–|0.847	32.26|–|0.907	34.61|–|0.942	29.82|–|0.844	32.15|–|0.905	34.45|–|0.941
DPW–SDNet[43]	29.53|29.13|0.821	31.90|31.27|0.885	34.30|33.44|0.928	–	–	–
DCSC[34]	29.17|29.17|0.815	31.48|31.47|0.880	–	28.81|28.79|0.784	30.96|30.92|0.858	–
ARGAN–MSE[37]	29.47|29.13|0.833	31.81|31.29|0.897	34.17|33.42|0.937	29.05|28.64|0.806	31.23|30.49|0.877	33.45|32.34|0.923
ARGAN[37]	27.65|27.63|0.777	29.99|29,69|0.864	31.64|31.17|–	27.31|27.31|0.740	28.48|29.03|0.841	30.98|30.16|0.884
DRA[35]	**29.98**|29.61|**0.851**	**32.34**|31.76|**0.908**	**34.78**|33.96|**0.944**	**29.92**|**29.41**|**0.841**	**32.23**|**31.39**|**0.906**	**34.61**|**33.34**|**0.943**
WCDGAN[64]	29.56|–|0.823	31.88|–|0.890	–	–	–	–
G–CAR[63]	29.75|**29.71**|0.829	32.07|**32.00**|0.890	34.46|**34.30**|0.931	–	–	–

Table 3: Average results of PSNR, PNSR-B, and SSIM on LIVE1 and BSDS500

Figure 1 displays the input jpeg distorted image and the deblocking results from the above algorithm. The figure also contains the ground truth gray image for reference. From the figure, we find that G-CAR[43] performs better than ARCNN[17], DnCNN[65] and TNRD[21]. When striped lines are present in the image, G-CAR[43] handles clearly better than MWCNN[32] and DMCNN[30].

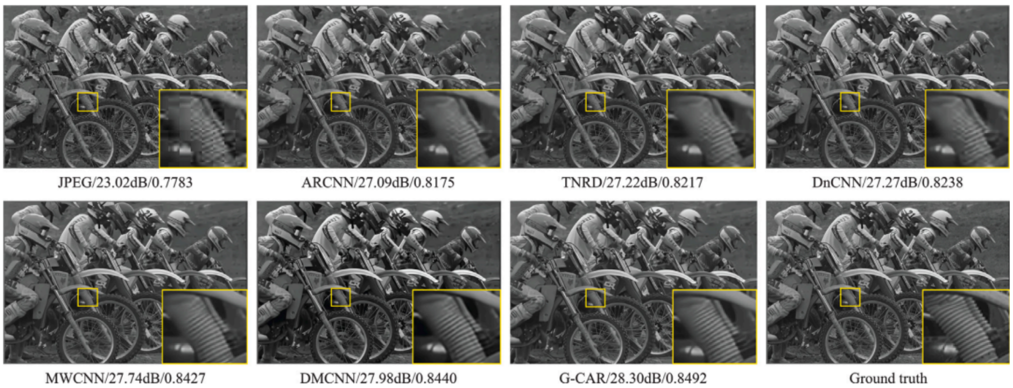

Figure 1: Performance of different deblocking methods for bicycle images using QF = 10 compressed from LIVE1 dataset. Below each recovered image, its PSNR-B and SSIM values are marked.

4.2 HEVC

4.2.1 Datasets

In our HEVC-based model comparison, we refer to the MFQE dataset, which is composed of 126 video sequences from JCT-VC, VQEG, and Xiph.org. The test set includes class A: Traffic and PeopleOnStreet; class B: Kimono, ParkScene, BQTerrace, Cactus, and BasketballDrive; class C: RaceHorses, BQMall, PartyScene, and BasketballDrill; class D: RaceHorses, BQSquare, BlowingBubbles and BasketballPass; class E: FourPeople, Johnny and KristenAndSara.

4.2.2 Evaluation Metrics

To evaluate the model effect, we also used PSNR and SSIM as quantitative metrics on the Y channel (YUV / YCbCr color space) to assess the quality enhancement performance.

4.2.3 Objective Evaluation

On the above dataset, we have chosen to compare the single-image artifact removal processing model, single-frame video processing model, and multi-frame video processing model. The more representative algorithms are selected in their respective fields.

Single-image artifact removal model: Artifact Removal CNN(AR-CNN)[17], denoising CNN(DnCNN)[65]

Single-frame video processing model: VRCNN)[69], DCAD)[64], DS-CNN[66]

Multi-frame video processing model: MFQE 1.0[71], MGANet[85], SDTS(SF)[78], SDTS(MC)[78], MFQE 2.0[72], STDF-R3[78], STDF-R3L[78], CF-STIF[79], RFDA[80].

All the video sequences are processed using HEVC Low Delay-P (LDP) configuration, and QPS of 22,27,32,37, and 42 were respectively tested.

As we can see from Table 4, STDF-R3 and RFDA are the winners for the complete reference metrics, outperforming almost all existing methods on the 5 QPs, verifying validity and superiority. Moreover, overall, due to the exploitation of spatio-temporal information, the single-frame video processing-based approach outperforms the single-image based framework, and the multi-frame based model outperforms all single-frame based approaches. These results have a high level of consistency across the different test sequences.

4.2.4 Subjective Evaluation

We selected 5 important methods from the above HEVC model for qualitative comparison and presented some comparisons in Fig.2. We used a single-image processing model (AR-CNN, DnCNN) and a multi-frame processing model (MFQE 2.0, STDF-R3, RFDA).

As shown in Figure 2, There are significant compression artifacts in the patch after compression (in the second column), for example, lack of human details, and the

<table>
<tr><th colspan="3"></th><th colspan="2">Image QE Methods</th><th colspan="3">Video QE Methods–Single frame</th><th colspan="9">Video QE Methods–multiple frame</th></tr>
<tr><th>QP</th><th colspan="2">Approach</th><th>AR–CNN</th><th>DnCNN</th><th>VRCNN</th><th>DCAD</th><th>DS–CNN</th><th>MFQE 1.0</th><th>MGANet</th><th>SDTS(SF)</th><th>SDTS(MC)</th><th>MFQE 2.0</th><th>STDF–R3</th><th>STDF–R3L</th><th>CF–STIF</th><th>RFDA</th></tr>
<tr><td rowspan="21">37</td><td rowspan="2">A</td><td>Traffic</td><td>0.24|0.47</td><td>0.24|0.57</td><td></td><td>0.31|0.67</td><td>0.29|0.60</td><td>0.50|0.90</td><td>0.50|–</td><td></td><td></td><td>0.59|1.02</td><td>0.65|1.04</td><td>0.73|1.15</td><td>0.78|1.26</td><td>0.80|1.28</td></tr>
<tr><td>PeopleOnStreet</td><td>0.35|0.75</td><td>0.41|0.82</td><td></td><td>0.50|0.95</td><td>0.42|0.85</td><td>0.80|1.37</td><td>0.91|–</td><td></td><td></td><td>0.92|1.57</td><td>1.18|1.82</td><td>1.25|1.96</td><td>1.40|2.16</td><td>1.44|2.22</td></tr>
<tr><td rowspan="5">B</td><td>Kimono</td><td>0.22|0.65</td><td>0.24|0.75</td><td>0.07|–</td><td>0.28|0.78</td><td>0.25|0.75</td><td>0.50|1.13</td><td>0.84|–</td><td>0.18|–</td><td>0.27|–</td><td>0.55|1.18</td><td>0.77|1.47</td><td>0.85|1.61</td><td>0.99|1.81</td><td>1.02|1.86</td></tr>
<tr><td>ParkScene</td><td>0.14|0.38</td><td>0.14|0.50</td><td>0.10|–</td><td>0.16|0.50</td><td>0.15|0.50</td><td>0.39|1.03</td><td>0.51|–</td><td>0.25|–</td><td>0.37|–</td><td>0.46|1.23</td><td>0.54|1.32</td><td>0.59|1.47</td><td>0.65|1.60</td><td>0.64|1.58</td></tr>
<tr><td>Cactus</td><td>0.19|0.38</td><td>0.20|0.48</td><td>0.08|–</td><td>0.26|0.58</td><td>0.24|0.58</td><td>0.44|0.88</td><td>0.55|–</td><td>0.22|–</td><td>0.28|–</td><td>0.50|1.00</td><td>0.70|1.23</td><td>0.77|1.38</td><td>0.83|1.50</td><td>0.83|1.49</td></tr>
<tr><td>BQTerrace</td><td>0.20|0.28</td><td>0.20|0.38</td><td>0.12|–</td><td>0.28|0.50</td><td>0.26|0.48</td><td>0.27|0.48</td><td>0.37|–</td><td>–0.06|–</td><td>–0.04|–</td><td>0.40|0.67</td><td>0.58|0.93</td><td>0.63|1.06</td><td>0.71|1.17</td><td>0.65|1.06</td></tr>
<tr><td>BasketballDrive</td><td>0.23|0.55</td><td>0.25|0.58</td><td>0.05|–</td><td>0.31|0.68</td><td>0.28|0.65</td><td>0.41|0.80</td><td>0.44|–</td><td>0.24|–</td><td>0.29|–</td><td>0.47|0.83</td><td>0.66|1.07</td><td>0.75|1.23</td><td>0.88|1.43</td><td>0.87|1.40</td></tr>
<tr><td rowspan="4">C</td><td>RaceHorses</td><td>0.22|0.43</td><td>0.25|0.65</td><td>0.07|–</td><td>0.28|0.65</td><td>0.27|0.63</td><td>0.34|0.55</td><td>0.52|–</td><td>0.07|–</td><td>0.14|–</td><td>0.39|0.80</td><td>0.48|1.09</td><td>0.55|1.35</td><td>0.59|1.54</td><td>0.48|1.23</td></tr>
<tr><td>BQMall</td><td>0.28|0.68</td><td>0.28|0.68</td><td>0.10|–</td><td>0.34|0.88</td><td>0.33|0.80</td><td>0.51|1.03</td><td>0.43|–</td><td>0.34|–</td><td>0.38|–</td><td>0.62 |1.20</td><td>0.90|1.61</td><td>0.99|1.80</td><td>1.11|1.97</td><td>1.09|1.97</td></tr>
<tr><td>PartyScene</td><td>0.11|0.38</td><td>0.13|0.48</td><td>0.06|–</td><td>0.16|0.48</td><td>0.17|0.58</td><td>0.22|0.73</td><td>0.34|–</td><td>0.26|–</td><td>0.30|–</td><td>0.36|1.18</td><td>0.60|1.60</td><td>0.68|1.94</td><td>0.75|2.10</td><td>0.66|1.88</td></tr>
<tr><td>BasketballDrill</td><td>0.25|0.58</td><td>0.33|0.68</td><td>0.07|–</td><td>0.39|0.78</td><td>0.35|0.68</td><td>0.48|0.90</td><td>0.49|–</td><td>0.17|–</td><td>0.20|–</td><td>0.58|1.20</td><td>0.70|1.26</td><td>0.79|1.49</td><td>0.94|1.73</td><td>0.88|1.67</td></tr>
<tr><td rowspan="4">D</td><td>RaceHorses</td><td>0.27 |0.55</td><td>0.31|0.73</td><td>0.14|–</td><td>0.34|0.83</td><td>0.32|0.75</td><td>0.51|1.13</td><td>0.81|–</td><td>0.21|–</td><td>0.36|–</td><td>0.59|1.43</td><td>0.73|1.75</td><td>0.83|2.08</td><td>0.59|1.54</td><td>0.85|2.11</td></tr>
<tr><td>BQSquare</td><td>0.08|0.08</td><td>0.13|0.18</td><td>0.19|–</td><td>0.20|0.38</td><td>0.20|0.38</td><td>–0.01|0.15</td><td>0.35|–</td><td>0.41|–</td><td>0.45|–</td><td>0.34|0.65</td><td>0.91|1.13</td><td>0.94|1.25</td><td>1.05|1.42</td><td>1.05|1.39</td></tr>
<tr><td>BlowingBubbles</td><td>0.16|0.35</td><td>0.18|0.58</td><td>0.08|–</td><td>0.22|0.65</td><td>0.23|0.68</td><td>0.39|1.20</td><td>0.43|–</td><td>0.30|–</td><td>0.38|–</td><td>0.53|1.70</td><td>0.68|1.96</td><td>0.74|2.26</td><td>0.78|2.38</td><td>0.78|2.40</td></tr>
<tr><td>BasketballPass</td><td>0.26|0.58</td><td>0.31|0.75</td><td>0.13|–</td><td>0.35|0.85</td><td>0.34|0.78</td><td>0.63|1.38</td><td>0.57|–</td><td>0.39|–</td><td>0.49|–</td><td>0.73|1.55</td><td>0.95|1.82</td><td>1.08|2.12</td><td>1.20|2.39</td><td>1.12|2.23</td></tr>
<tr><td rowspan="3">E</td><td>FourPeople</td><td>0.37|0.50</td><td>0.39|0.60</td><td>0.22|–</td><td>0.51|0.78</td><td>0.46|0.70</td><td>0.66|0.85</td><td>0.66|–</td><td>0.46|–</td><td>0.48|–</td><td>0.73|0.95</td><td>0.92|1.07</td><td>0.94|1.17</td><td>1.04|1.27</td><td>1.13|1.36</td></tr>
<tr><td>Johnny</td><td>0.25|0.10</td><td>0.32|0.40</td><td>0.20|–</td><td>0.41|0.50</td><td>0.38|0.40</td><td>0.55|0.55</td><td>0.58|–</td><td>0.40|–</td><td>0.43|–</td><td>0.60|0.68</td><td>0.69|0.73</td><td>0.81|0.88</td><td>0.88|0.97</td><td>0.90|0.94</td></tr>
<tr><td>KristenAndSara</td><td>0.41|0.50</td><td>0.42|0.60</td><td>0.19|–</td><td>0.52|0.70</td><td>0.48|0.60</td><td>0.66|0.75</td><td>0.79|–</td><td>0.43|–</td><td>0.46|–</td><td>0.75|0.85</td><td>0.94|0.89</td><td>0.97|0.96</td><td>1.10|1.08</td><td>1.19|1.15</td></tr>
<tr><td colspan="2">Average</td><td>0.23|0.45</td><td>0.26|0.58</td><td>0.12|–</td><td>0.32|0.67</td><td>0.30|0.63</td><td>0.46|0.88</td><td></td><td>0.27|–</td><td>0.33|–</td><td>0.56|1.09</td><td>0.75|1.32</td><td>0.83|1.51</td><td>0.92|1.67</td><td>0.91|1.62</td></tr>
<tr><td>42</td><td colspan="2">Average</td><td>0.29|0.96</td><td>0.22|0.77</td><td></td><td>0.32|1.09</td><td>0.31|1.01</td><td>0.44|1.30</td><td>0.57|–</td><td></td><td></td><td>0.59|1.65</td><td></td><td></td><td>0.89|2.32</td><td>0.82|2.20</td></tr>
<tr><td>32</td><td colspan="2">Average</td><td>0.18|0.19</td><td>0.26|0.35</td><td>0.10|–</td><td>0.32|0.44</td><td>0.27|0.38</td><td>0.43|0.58</td><td>0.60|–</td><td>0.27|–</td><td>0.29|–</td><td>0.516|0.68</td><td>0.73|0.87</td><td>0.86|1.04</td><td>0.95|1.18</td><td>0.87|1.07</td></tr>
<tr><td>27</td><td colspan="2">Average</td><td>0.18|0.14</td><td>0.27|0.24</td><td></td><td>0.32|0.30</td><td>0.27|0.23</td><td>0.40|0.34</td><td>0.56|–</td><td></td><td></td><td>0.49|0.42</td><td>0.67|0.53</td><td>0.72|0.57</td><td>0.94|0.78</td><td>0.82|0.68</td></tr>
<tr><td>22</td><td colspan="2">Average</td><td>0.14|0.08</td><td>0.29|0.18</td><td></td><td>0.31|0.19</td><td>0.25|0.15</td><td>0.31|0.19</td><td></td><td></td><td></td><td>0.46|0.27</td><td>0.57|0.30</td><td>0.63|0.34</td><td>0.85|0.46</td><td>0.76|0.42</td></tr>
</table>

Table 4: Average PSNR and SSIM results on the MFQE dataset.

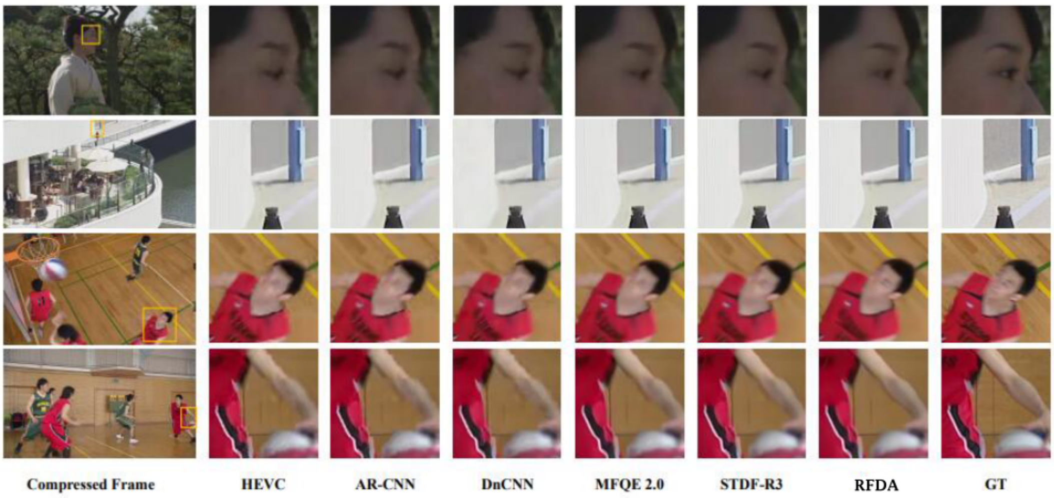

Figure 2: Qualitative comparison was made at QP =37.

single-frame approach cannot address temporal noise well, while MFQE 2.0 and STDF-R3 faces excessive smoothing. In contrast, RFDA recovers more details or texture compared with other approaches, especially in the boundary regions of fast-moving objects. The effect is relatively good.

5 Trends and Challenges

1. Recent good methods that have shown significant results focus on large numbers of parameters and high model accuracy. However, they all ignore the application effect

of actual scenarios, and there are limitations in the realization of lightweight models in actual applications. Lightweight, small sample, compact and efficient model waiting to be mined.

2. Image and video compression methods are also improving, and the latest codec, Universal Video Coding (VVC), has appeared today. In the face of more complex compression principles and structures, it is meaningful to develop more efficient and general artifact removal methods for specific compression.

3. With deep learning developing constantly, there are more and more methods for artifact removal of multi-frame video compression. However, due to the late start, the spatio-temporal information of video hasn't been widely used, so there is still room for exploration in the fields of network structure, terminal information integration, and loss function. In addition, most of the current video quality evaluations still use PSNR and SSIM as objective metrics. Future breakthrough work may further propose improved perceptual quality metrics to enhance the quality of experience (QoE).

4. Artifact removal is only one branch of the field of image/video compression reconstruction. Due to the unified goal of improving the visual quality of images, excellent models in various fields can be scaled to some lower-level visual tasks related to video, including restoration, super-resolution, and frame synthesis, to achieve high-efficiency temporal information fusion.

6 Conclusion

This paper presents a comprehensive review of artifact removal methods for image/video compression, including traditional methods and deep learning-based methods. According to the principle, traditional artifact removal can be classified into deblock-based filtering methods and recovery-based inspection estimation methods. With the development of deep learning, excellent models for specific compression methods are constantly being mined. At present, the single image processing model based on JPEG has three improvement directions: the improvement of network structure, the utilization of transform domain information, and the loss function. The video processing model based on HEVC has evolved from single-frame processing to multi-frame processing, combined with temporal and spatial correlation and a variety of prior knowledge, and has achieved good results. In this paper, only two widely used compression algorithms are studied, and the generality of the algorithms will be expanded later.

References

[1] H. C. Reeve and J. S. Lim, "Reduction of blocking effect in image coding", Proc. IEEE ICASSP, vol. 8, pp. 1212-1215, 1983-Apr.

[2] B. Ramamurthi and A. Gersho, "Nonlinear space-variant postprocessing of block coded images", IEEE Trans. Acoust. Speech Signal Process., vol. 34, no. 5, pp. 1258-1268, Oct. 1986.

[3] S. Minami and A. Zakhor, "An optimization approach for removing blocking effects in transform coding," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 5, no. 2, pp. 74-82, April 1995, doi: 10.1109/76.388056.

[4] A. Buades, B. Coll and J. M. Morel, "A review of image denoising algorithms with a new one", SIAM J. Multiscale Model. Simul., vol. 4, no. 2, pp. 490-530, Jan. 2005.

[5] T. Chen, H. R. Wu and B. Qiu, "Adaptive postfiltering of transform confidents for the reduction of blocking artifacts", IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 594-602, May 2001.

[6] J. Kim, "Adaptive blocking artifact reduction using wavelet-based block analysis," in IEEE Transactions on Consumer Electronics, vol. 55, no. 2, pp. 933-940, May 2009, doi: 10.1109/TCE.2009.5174477.

[7] S. Wu, H. Yan and Z. Tan, "An efficient wavelet-based deblocking algorithm for highly compressed images," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 11, pp. 1193-1198, Nov. 2001, doi: 10.1109/76.964789.

[8] Xu, Jun, Shibao Zheng and Xiaokang Yang. "Adaptive video-blocking artifact removal in discrete Hadamard transform domain." Optical Engineering 45 (2006): 080501.

[9] K. Lee, D. S. Kim and T. Kim, "Regression-based prediction for blocking artifact reduction in JPEG-compressed images", IEEE Trans. Image Process., vol. 14, no. 1, pp. 36-48, Jan. 2005.

[10] A. Foi, V. Katkovnik and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images", IEEE Trans. Image Process., vol. 16, no. 5, pp. 1395-1411, May 2007.

[11] Francisco, Nelson C., Nuno M. M. Rodrigues, Eduardo Antônio Barros da Silva and Sérgio M. M. Faria. "A generic post-deblocking filter for block based image compression algorithms." Signal Process. Image Commun. 27 (2012): 985-997.

[12] G. Zhai, W. Zhang, X. Yang, W. Lin and Y. Xu, "Efficient Image Deblocking Based on Postfiltering in Shifted Windows," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 1, pp. 122-126, Jan. 2008, doi: 10.1109/TCSVT.2007.906942.

[13] S. Minami and A. Zakhor, "An optimization approach for removing blocking effects in transform coding," [1991] Proceedings. Data Compression Conference, Snowbird, UT, USA, 1991, pp. 442-, doi: 10.1109/DCC.1991.213319.

[14] Ying Luo and R. K. Ward, "Removing the blocking artifacts of block-based DCT compressed images," in IEEE Transactions on Image Processing, vol. 12, no. 7,

pp. 838-842, July 2003, doi: 10.1109/TIP.2003.814252.

[15] Shizhong Liu and A. C. Bovik, "Efficient DCT-domain blind measurement and reduction of blocking artifacts," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 12, pp. 1139-1149, Dec. 2002, doi: 10.1109/TCSVT.2002.806819.

[16] G. A. Triantafyllidis, D. Tzovaras and M. G. Strintzis, "Blocking artifact detection and reduction in compressed data," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 10, pp. 877-890, Oct. 2002, doi: 10.1109/TCSVT.2002.804880.

[17] C. Dong, Y. Deng, C. C. Loy and X. Tang, "Compression artifacts reduction by a deep convolutional network", Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 576-584, Dec. 2015.

[18] X. Mao, C. Shen and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections", Proc. NIPS, pp. 2802-2810, 2016.

[19] L. Cavigelli, P. Hager and L. Benini, "CAS-CNN: A deep convolutional neural network for image compression artifact suppression", Proc. IJCNN, pp. 752-759, 2017.

[20] J. Guo and H. Chao, "One-to-many network for visually pleasing compression artifacts reduction", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 3038-3047, Jul. 2017.

[21] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration", IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1256-1272, Jun. 2017.

[22] Y. Tai, J. Yang, X. Liu and C. Xu, "MemNet: A persistent memory network for image restoration", Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 4539-4547, Oct. 2017.

[23] Zheng B., Sun R., Tian X., Chen Y. S-Net: a scalable convolutional neural network for JPEG compression artifact reduction J. Electron. Imaging, 27 (4) (2018), Article 043037

[24] X. Fu, Z. -J. Zha, F. Wu, X. Ding and J. Paisley, "JPEG Artifacts Reduction via Deep Convolutional Sparse Coding," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 2501-2510, doi: 10.1109/ICCV.2019.00259.

[25] S. Zini, S. Bianco and R. Schettini, "Deep Residual Autoencoder for Blind Universal JPEG Restoration," in IEEE Access, vol. 8, pp. 63283-63294, 2020, doi:

10.1109/ACCESS.2020.2984387.

[26] L. Galteri, L. Seidenari, M. Bertini, A. Del Bimbo, Deep generative adversarial compression artifact removal, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4826–4835.

[27] L. Galteri, L. Seidenari, M. Bertini and A. D. Bimbo, "Deep Universal Generative Adversarial Compression Artifact Removal," in IEEE Transactions on Multimedia, vol. 21, no. 8, pp. 2131-2145, Aug. 2019, doi: 10.1109/TMM.2019.2895280.

[28] H. Ma, D. Liu and F. Wu, "Rectified Wasserstein Generative Adversarial Networks for Perceptual Image Restoration," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 3, pp. 3648-3663, 1 March 2023, doi: 10.1109/TPAMI.2022.3185316.

[29] Z. Wang, D. Liu, S. Chang, Q. Ling, Y. Yang and T. S. Huang, "D3: Deep dual-domain based fast restoration of JPEG-compressed images", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2764-2772, Jun. 2016.

[30] X. Zhang, W. Yang, Y. Hu and J. Liu, "DMCNN: Dual-domain multi-scale convolutional neural network for compression artifacts removal", Proc. 25th IEEE Int. Conf. Image Process. (ICIP), pp. 390-394, Oct. 2018.

[31] Guo, Jun and Hongyang Chao. "Building Dual-Domain Representations for Compression Artifacts Reduction." European Conference on Computer Vision (2016).

[32] P. Liu, H. Zhang, K. Zhang, L. Lin, W. Zuo, Multi-level wavelet-CNN for image restoration, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 773–782.

[33] H. Chen, X. He, L. Qing, S. Xiong and T. Q. Nguyen, "DPW-SDNet: Dual pixel-wavelet domain deep CNNs for soft decoding of JPEG-compressed images", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), pp. 711-720, Jun. 2018.

[34] M. Mehralian and B. Karasfi, "RDCGAN: Unsupervised Representation Learning With Regularized Deep Convolutional Generative Adversarial Networks," 2018 9th Conference on Artificial Intelligence and Robotics and 2nd Asia-Pacific International Symposium, Kish Island, Iran, 2018, pp. 31-38, doi: 10.1109/AIAR.2018.8769811.

[35] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134.

[36] Dosovitskiy, Alexey and Thomas Brox. “Generating Images with Perceptual Similarity Metrics based on Deep Networks.” NIPS (2016).

[37] Johnson, Justin, Alexandre Alahi and Li Fei-Fei. “Perceptual Losses for Real-Time Style Transfer and Super-Resolution.” ArXiv abs/1603.08155 (2016): n. pag.

[38] Svoboda, Pavel, Michal Hradiš, David Barina and Pavel Zemčík. “Compression Artifacts Removal Using Convolutional Neural Networks.” ArXiv abs/1605.00366 (2016): n. pag.

[39] Yu, Fisher and Vladlen Koltun. “Multi-Scale Context Aggregation by Dilated Convolutions.” CoRR abs/1511.07122 (2015): n. pag.

[40] L. M. Bregman, “Finding the common point of convex sets’by the method of successive projection,” Dokl. Akad. Nauk SSSR, vol. 162, pp. 487-490, 1965.

[41] Gregory K Wallace“The jpeg Still picture compression standard”IEEE Transactions on Consumer Electronicsvol. 38 no. 1 pp. xviii-xxxiv 1992.

[42] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han and Thomas Wiegand, ”Overview of the high efficiency video coding (HEVC) standard”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, 2012.

[43] Zhang, Yi, Damon M. Chandler and Xuanqin Mou. “Multi-domain residual encoder-decoder networks for generalized compression artifact reduction.” J. Vis. Commun. Image Represent. 83 (2022): 103425.

[44] B. Xie, H. Zhang and C. Jung, ”WCDGAN: Weakly Connected Dense Generative Adversarial Network for Artifact Removal of Highly Compressed Images,” in IEEE Access, vol. 10, pp. 1637-1649, 2022, doi: 10.1109/ACCESS.2021.3138106.

[45] S. J. Reeves and S. L. Eddins, “Comments on “Iterative procedures for reduction of blocking effects in transform image coding,” IEEE Trans. Circuits and Syst. for video Technol., vol. 3, pp. 43940, Dec. 1993.

[46] A. Zakhor,“Iterative procedures for reduction of blocking effects in transform image coding”, IEEE Trans. on Circuits and Systems for Video Tech. vol. 2, no. 1, March 1992.

[47] Yongyi Yang, N. P. Galatsanos and A. K. Katsaggelos, ”Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images,” in IEEE Transactions on Circuits and Systems for Video Technology, vol. 3, no. 6, pp. 421-432, Dec. 1993, doi: 10.1109/76.260198.

[48] T. Meier, K. N. Ngan and G. Crebbin, ”Reduction of blocking artifacts in image and video coding,” in IEEE Transactions on Circuits and Systems for Video

Technology, vol. 9, no. 3, pp. 490-500, April 1999, doi: 10.1109/76.754778.

[49] T. P. O'Rourke and R. L. Stevenson, "Improved image decompression for reduced transform coding artifacts," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 5, no. 6, pp. 490-499, Dec. 1995, doi: 10.1109/76.475891.

[50] M. A. Robertson and R. L. Stevenson, "DCT quantization noise in compressed images," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 1, pp. 27-38, Jan. 2005, doi: 10.1109/TCSVT.2004.839995.

[51] D. Sun and W. -K. Cham, "Postprocessing of Low Bit-Rate Block DCT Coded Images Based on a Fields of Experts Prior," in IEEE Transactions on Image Processing, vol. 16, no. 11, pp. 2743-2751, Nov. 2007, doi: 10.1109/TIP.2007.904969.

[52] G.M. Farinella, S. Battiato, On the application of structured sparse model selection to JPEG compressed images, in: Proceedings of International Conference on Computational Color Imaging, 2011, pp. 137–151.

[53] C. Jung, L. Jiao, H. Qi and T. Sun, "Image deblocking via sparse representation", Signal Process. Image Commun., vol. 27, no. 6, pp. 663-677, 2012.

[54] H. Chang, M. K. Ng and T. Zeng, "Reducing artifacts in JPEG decompression via a learned dictionary", IEEE Trans. Signal Process., vol. 62, no. 3, pp. 718-728, Feb. 2014.

[55] Y. -W. Chiou, C. -H. Yeh, L. -W. Kang, C. -W. Lin and S. -J. Fan-Jiang, "Efficient image/video deblocking via sparse representation," 2012 Visual Communications and Image Processing, San Diego, CA, USA, 2012, pp. 1-6, doi: 10.1109/VCIP.2012.6410838.

[56] J. Jancsary, S. Nowozin and C. Rother, "Loss-specific training of non-parametric image restoration models: A new state of the art", Proc. Eur. Conf. Comput. Vis., pp. 112-125, 2012.

[57] J. Jancsary, S. Nowozin, T. Sharp and C. Rother, "Regression Tree Fields — An efficient, non-parametric approach to image labeling problems," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 2012, pp. 2376-2383, doi: 10.1109/CVPR.2012.6247950.

[58] U. Schmidt, J. Jancsary, S. Nowozin, S. Roth and C. Rother, "Cascades of Regression Tree Fields for Image Restoration," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 4, pp. 677-689, 1 April 2016, doi: 10.1109/TPAMI.2015.2441053.

[59] A. Norkin et al., "HEVC deblocking filter," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1746–1754, Dec. 2012.

[60] C.-M. Fu et al., “Sample adaptive offset in the HEVC standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1755–1764, Dec. 2012.

[61] W.-S. Park and M. Kim, ”CNN-based in-loop filtering for coding efficiency improvement”, IEEE Image Video and Multidimensional Signal Processing Workshop, pp. 1-5, 2016.

[62] Qinglong Han and WaiKuen Cham, ”High performance loop filter for HEVC”, ICIP, 2015.

[63] H. Huang, I. Schiopu and A. Munteanu, ”Frame-Wise CNN-Based Filtering for Intra-Frame Quality Enhancement of HEVC Videos,” in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 6, pp. 2100-2113, June 2021, doi: 10.1109/TCSVT.2020.3018230.

[64] T. Wang, M. Chen and H. Chao, ”A novel deep learning-based method of improving coding efficiency from the decoder-end for HEVC”, Proc. IEEE Data Compression Conf., pp. 410-419, 2017.

[65] K. Zhang, W. Zuo, Y. Chen, D. Meng and L. Zhang, ”Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising”, IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142-3155, Jul. 2017.

[66] R. Yang, M. Xu and Z. Wang, ”Decoder-side HEVC quality enhancement with scalable convolutional neural network”, Proc. IEEE Int. Conf. Multimedia Expo, pp. 817-822, 2017.

[67] R. Yang, M. Xu, T. Liu, Z. Wang and Z. Guan, ”Enhancing quality for HEVC compressed videos”, IEEE Trans. Circuits Syst. Video. Technol., vol. 29, no. 7, pp. 2039-2054, Jul. 2019.

[68] K. Li, B. Bare and B. Yan, ”An efficient deep convolutional neural networks model for compressed image deblocking”, Proc. IEEE Int. Conf. Multimedia Expo, pp. 1320-1325, 2017.

[69] Y. Dai, D. Liu and F. Wu, ”A convolutional neural network approach for post-processing in hevc intra coding”, International Conference on Multimedia Modeling, pp. 28-39, 2017.

[70] Z. Jin, P. An, C. Yang and L. Shen, ”Quality enhancement for intra frame coding via cnns: An adversarial approach”, IEEE International Conference on Acoustics Speech and Signal Processing, pp. 1368-1372, 2018.

[71] R. Yang, M. Xu, Z. Wang and T. Li, ”Multi-frame quality enhancement for compressed video”, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 6664-6673, 2018.

[72] Z. Guan, Q. Xing, M. Xu, R. Yang, T. Liu and Z. Wang, "MFQE 2.0: A New Approach for Multi-Frame Quality Enhancement on Compressed Video," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 3, pp. 949-963, 1 March 2021, doi: 10.1109/TPAMI.2019.2944806.

[73] J. Tong, X. Wu, D. Ding, Z. Zhu and Z. Liu, "Learning-based multi-frame video quality enhancement", Proc. IEEE Int. Conf. Image Process., pp. 929-933, 2019.

[74] X. Meng, X. Deng, S. Zhu and B. Zeng, "Enhancing quality for VVC compressed videos by jointly exploiting spatial details and temporal structure", Proc. IEEE Int. Conf. Image Process., pp. 1193-1197, 2019.

[75] D. Luo, M. Ye, S. Chen and X. Li, "Alignment-free video compression artifact reduction", Proc. IEEE Int. Conf. Vis. Commun. Image Process., pp. 1-5, 2021.

[76] J. Wang, X. Deng, M. Xu, C. Chen and Y. Song, "Multi-level wavelet-based generative adversarial network for perceptual quality enhancement of compressed video", Proc. Eur. Conf. Comput. Vis., pp. 405-421, 2020.

[77] R. Yang, X. Sun, M. Xu and W. Zeng, "Quality-gated convolutional LSTM for enhancing compressed video", Proc. IEEE Int. Conf. Multimedia Expo, pp. 532-537, 2019.

[78] J. Deng, L. Wang, S. Pu and C. Zhuo, "Spatio-temporal deformable convolution for compressed video quality enhancement", Proc. AAAI Conf. Artif. Intell., pp. 10696-10703, 2020.

[79] D. Luo, M. Ye, S. Li and X. Li, "Coarse-to-fine spatio-temporal information fusion for compressed video quality enhancement", IEEE Signal Process. Lett., vol. 29, pp. 543-547, 2022.

[80] M. Zhao, Y. Xu and S. Zhou, "Recursive fusion and deformable spatio-temporal attention for video compression artifact reduction", Proc. 29th ACM Int. Conf. Multimedia, pp. 5646-5654, 2021.

[81] Z. Wang, M. Ye, S. Li and X. Li, "Multi-Frame Compressed Video Quality Enhancement by Spatio-Temporal Information Balance," in IEEE Signal Processing Letters, vol. 30, pp. 105-109, 2023, doi: 10.1109/LSP.2023.3244711.

[82] Guo Lu, Wanli Ouyang, Dong Xu, Xiaoyun Zhang, Zhiyong Gao, and Ming-Ting Sun. 2018. Deep Kalman Filtering Network for Video Compression Artifact Reduction. In ECCV. 568–584.

[83] Guo Lu, Xiaoyun Zhang, Wanli Ouyang, Dong Xu, Li Chen, and Zhiyong Gao. 2019. Deep Non-Local Kalman Network for Video Compression Artifact Reduction. TIP 29 (2019), 1725–1737

[84] Yi Xu, Longwen Gao, Kai Tian, Shuigeng Zhou, and Huyang Sun. 2019. Non-local ConvLSTM for video compression artifact reduction. In ICCV. 7043–7052.

[85] X. Meng, X. Deng, S. Zhu, S. Liu, C. Wang, C. Chen, et al., "MGANet: A Robust Model for Quality Enhancement of Compressed Video", arXiv preprint arXiv:1811.09150, 2018.

[86] W. Xiao, H. He, T. Wang and H. Chao, "The interpretable fast multi-scale deep decoder for the standard hevc bitstreams", IEEE Trans. Multimedia, vol. 22, no. 7, pp. 1680-1691, 2020.

[87] X. Meng, X. Deng, S. Zhu, X. Zhang and B. Zeng, "A robust quality enhancement method based on joint spatial-temporal priors for video coding", IEEE Transactions on Circuits and Systems for Video Technology, 2020.

[88] L. Peng, A. Hamdulla, M. Ye, S. Li and H. Guo, "Recurrent Deformable Fusion for Compressed Video Artifact Reduction," 2022 IEEE International Symposium on Circuits and Systems (ISCAS), Austin, TX, USA, 2022, pp. 3175-3179, doi: 10.1109/ISCAS48785.2022.9937741.

[89] X. Shi, J. Lin, D. Jiang, C. Nian and J. Yin, "Recurrent Network with Enhanced Alignment and Attention-Guided Aggregation for Compressed Video Quality Enhancement," 2022 IEEE International Conference on Visual Communications and Image Processing (VCIP), Suzhou, China, 2022, pp. 1-5, doi: 10.1109/VCIP56404.2022.10008807.

[90] K. Chan et al., BasicVSR++: Improving video super-resolution with enhanced propagation and alignment, 2021.

[91] P. Arbelaez et al., "Contour detection and hierarchical image segmentation," IEEE Trans. Pattern Anal. Mach. Intell. 33(5), 898–916 (2011)

[92] C. Yim and A. C. Bovik, "Quality assessment of deblocked images", IEEE Trans. Image Process., vol. 20, no. 1, pp. 88-98, Jan. 2011.

[93] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", IEEE Trans. Image Process., vol. 13, no. 4, pp. 600-612, Apr. 2004.

[94] S. Hochreiter and J. Schmidhuber, "Long short-term memory", Neural Comput, vol. 9, no. 8, pp. 1735-1780, 1997.

[95] M. Yuen, H. R. Wu and K. R. Rao, "Coding artifacts and visual distortions" in Digital Video Image Quality and Perceptual Coding, FL:CRC Press, pp. 87-122, 2006.

[96] X. Fan, W. Gao, Y. Lu and D. Zhao, "Flicking reduction in all intra-frame coding", Proc. 5th Meeting Joint Video Team ISO/IEC JTC1/SC29/WG11 ITU-T SG16 Q.6 document JVT-E070.doc, 2002-Oct.

[97] S. S. Chun, J.-R. Kim and S. Sull, "Intra-prediction mode selection for flicker reduction in H.264/AVC", IEEE Trans. Consum. Electron., vol. 52, no. 4, pp. 1303-1310, Nov. 2006.

[98] K. Chono, Y. Senda and Y. Miyamoto, "Detented quantization to suppress flicker artifacts in periodically inserted intra-coded pictures in H.264 video coding", Proc. IEEE Int. Conf. Image Process. (ICIP), pp. 1713-1716, 2006.

[99] Y. Kuszpet, D. Kletsel, Y. Moshe and A. Levy, "Postprocessing for flicker reduction in H.264/AVC", Proc. 16th Picture Coding Symp. (PCS), 2007-Nov.

[100] D. T. V, T. Q. Nguyen, S. Yea and A. Vetro, "Adaptive fuzzy filtering for artifact reduction in compressed images and videos", IEEE Trans. Image Process., vol. 18, no. 6, pp. 1166-1178, Jun. 2009.

[101] R. L. Stevenson, "Reduction of coding artifacts in transform image coding", Proc. IEEE Int. Conf. Acoust. Speech Signal Process., vol. 5, pp. 401-404, 1993-Mar.

[102] Jie Xiang Yang and Hong Ren Wu, "Robust filtering technique for reduction of temporal fluctuation in H.264 video sequences," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 20, no. 3, pp. 458-462, Mar. 2010.

[103] Y. Kim et al., "A Pseudo-Blind Convolutional Neural Network for the Reduction of Compression Artifacts," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 4, pp. 1121-1135, April 2020, doi: 10.1109/TCSVT.2019.2901919.

[104] J. W. Soh et al., "Reduction of Video Compression Artifacts Based on Deep Temporal Networks," in IEEE Access, vol. 6, pp. 63094-63106, 2018, doi: 10.1109/ACCESS.2018.2876864.

[105] X. Han, W. Zhang and J. Pu, "Exploring Spatio-temporal Relationships for Improving Compressed Video Quality," 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 2022, pp. 400-406, doi: 10.1109/ICPR56361.2022.9956306.

[106] J. Liu, D. Liu, W. Yang, S. Xia, X. Zhang and Y. Dai, "A Comprehensive Benchmark for Single Image Compression Artifact Reduction," in IEEE Transactions on Image Processing, vol. 29, pp. 7845-7860, 2020, doi: 10.1109/TIP.2020.3007828.

[107] Y. Dai, D. Liu, Z. -J. Zha and F. Wu, "A CNN-Based In-Loop Filter with CU Classification for HEVC," 2018 IEEE Visual Communications and Image Processing (VCIP), Taichung, Taiwan, 2018, pp. 1-4, doi: 10.1109/VCIP.2018.8698616.