Early Exit Or Not: Resource-Efficient Blind Quality Enhancement for Compressed Images

Qunliang Xing[1 [0000−0002−3007−716X]], Mai Xu[1,2 [0000−0002−0277−3301]], Tianyi Li[1 [0000−0001−7038−7798]], and Zhenyu Guan[1 [0000−0002−3959−338X]⋆]

[1] School of Electronic and Information Engineering, Beihang University
[2] Hangzhou Innovation Institute, Beihang University
{xingql,maixu,tianyili,guanzhenyu}@buaa.edu.cn

Abstract. Lossy image compression is pervasively conducted to save communication bandwidth, resulting in undesirable compression artifacts. Recently, extensive approaches have been proposed to reduce image compression artifacts at the decoder side; however, they require a series of architecture-identical models to process images with different quality, which are inefficient and resource-consuming. Besides, it is common in practice that compressed images are with unknown quality and it is intractable for existing approaches to select a suitable model for blind quality enhancement. In this paper, we propose a resource-efficient blind quality enhancement (RBQE) approach for compressed images. Specifically, our approach blindly and progressively enhances the quality of compressed images through a dynamic deep neural network (DNN), in which an early-exit strategy is embedded. Then, our approach can automatically decide to terminate or continue enhancement according to the assessed quality of enhanced images. Consequently, slight artifacts can be removed in a simpler and faster process, while the severe artifacts can be further removed in a more elaborate process. Extensive experiments demonstrate that our RBQE approach achieves state-of-the-art performance in terms of both blind quality enhancement and resource efficiency.

Keywords: blind quality enhancement · compressed images · resource-efficient · early-exit

1 Introduction

We are embracing an era of visual data explosion. According to Cisco mobile traffic forecast [4], the amount of mobile visual data is predicted to grow nearly 10-fold from 2017 to 2022. To overcome the bandwidth-hungry bottleneck caused by a deluge of visual data, lossy image compression, such as JPEG [40], JPEG 2000 [28] and HEVC-MSP [37], has been pervasively used. However, compressed images inevitably suffer from compression artifacts, such as blocky effects, ringing effects and blurring, which severely degrade the Quality of Experience (QoE) [35, 39] and the performance of high-level vision tasks [17, 48].

⋆ Corresponding author: Mai Xu.

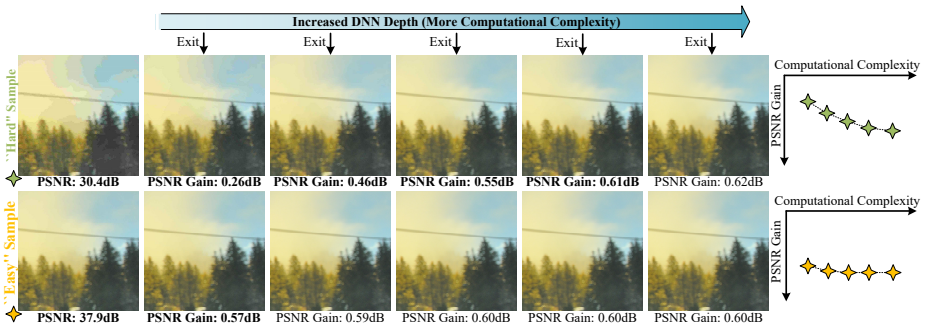

Fig. 1. Examples of quality enhancement on "easy" and "hard" samples, along with increased computational complexity.

For enhancing the quality of compressed images, many approaches [8, 13, 42, 22, 46, 16, 47, 12] have been proposed. Their basic idea is that one model needs to be trained for enhancing compressed images with similar quality reflected by a particular value of Quantization Parameter (QP) [37], and then a series of architecture-identical models need to be trained for enhancing compressed images with different quality. For example, [42, 46, 12] train 5 deep models to handle compressed images with QP = 22, 27, 32, 37 and 42. There are three main drawbacks to these approaches. (1) QP cannot faithfully reflect image quality, and thus it is intractable to manually select a suitable model based on QP value. (2) These approaches consume large computational resources during the training stage since many architecture-identical models need to be trained. (3) Compressed images with different quality are enhanced with the same computational complexity, such that these approaches impose excessive computational costs on "easy" samples (high-quality compressed images) but lack sufficient computation on "hard" samples (low-quality compressed images). Intuitively, the quality enhancement of images with different quality can be partly shared in a single framework, such that the joint computational costs can be reduced. More importantly, slight artifacts should be removed in a simpler and faster process, while the severe artifacts need to be further removed through a more elaborate process. Therefore, an ideal framework should automatically conduct a simple or elaborate enhancement process by distinguishing "easy" and "hard" samples, as a blind quality enhancement task.

In this paper, we propose a resource-efficient blind quality enhancement (RBQE) approach for compressed images. Specifically, we first prove that there exist "easy"/"hard" samples for quality enhancement on compressed images. We demonstrate that "easy" samples are those with slight compression artifacts, while "hard" samples are those with severe artifacts. Then, a novel dynamic deep neural network (DNN) is designed, which progressively enhances the quality of compressed image, assesses the enhanced image quality, and automatically decides whether to terminate (early exit) or continue the enhancement. The quality assessment and early-exit decision are managed by a Tchebichef moments-

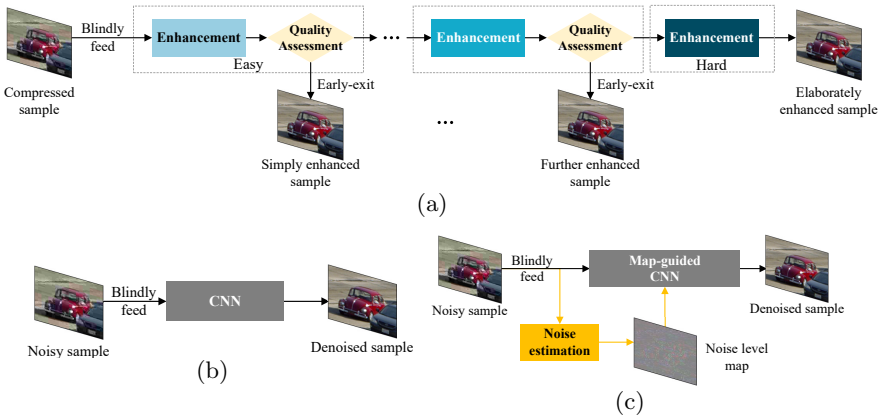

Fig. 2. Proposed resource-efficient blind quality enhancement paradigm (a) vs. two traditional blind denoising paradigms (b) and (c). Our paradigm dynamically processes samples with early exits for "easy" samples, while traditional paradigms (b) and (c) statically process images with equal computational costs on both "easy" and "hard" samples.

based Image Quality Assessment Module (IQAM), which is strongly sensitive to compression artifacts. Finally, our RBQE approach can perform "easy to hard" quality enhancement in an end-to-end manner. This way, images with slight compression artifacts can be simply and rapidly enhanced, while those with severe artifacts need to be further enhanced. Some examples are shown in Fig. 1. Also, experimental results verify that our RBQE approach achieves state-of-the-art performance for blind quality enhancement in both efficiency and efficacy.

To the best of knowledge, our approach is a first attempt to manage quality enhancement of compressed images in a resource-efficient manner. To sum up, the contributions are as follows:

(1) We prove that "easy"/"hard" samples exist in quality enhancement, as the theoretical foundation of our approach.
(2) We propose the RBQE approach with a simple yet effective dynamic DNN architecture, which processes "easy to hard" paradigm for blind quality enhancement.
(3) We develop a Tchebichef moments-based IQAM, workable for early-exit determination in our dynamic DNN structure.

2 Related Work

2.1 Quality Enhancement for Compressed Images

Due to the astonishing development of Convolutional Neural Networks (CNNs) [36, 34, 9] and large-scale image datasets [7], several CNN-based quality enhancement

approaches have been successfully applied to JPEG-compressed images. Dong et al. [8] proposed a shallow four-layer Artifacts Reduction Convolutional Neural Network (AR-CNN), which is the pioneer of CNN-based quality enhancement of JPEG-compressed images. Later, Deep Dual-Domain (D3) approach [43] and Deep Dual-domain Convolutional neural Network (DDCN) [13] were proposed for JPEG artifacts removal, which are motivated by dual-domain sparse coding and utilize the quantization prior of JPEG compression. DnCNN [49] is a milestone for reducing both Additive White Gaussian Noise (AWGN) and JPEG artifacts. It is a 20-layer deep network employing residual learning [15] and batch normalization [19], which can yield better results than Block-Matching and 3-D filtering (BM3D) approach [5]. It also achieves blind denoising by mixing and sampling training data randomly with different levels of noise.

Most recently, extensive works have been devoted to the latest video/image coding standard, HEVC/HEVC-MSP [37, 31, 2, 32, 26, 25, 45]. Due to the elaborate coding strategies of HEVC, the approaches for JPEG-compressed images [8, 43, 13, 49], especially those utilizing the prior of JPEG compression [43, 13], cannot be directly used for quality enhancement of HEVC-compressed images. In fact, HEVC [37] codec already incorporates the in-loop filters, which consist of Deblocking Filter (DF) [33] and Sample Adaptive Offset (SAO) filter [10], to suppress blocky effects and ringing effects. However, these handcrafted filters are far from optimum, resulting in still visible artifacts in compressed images. To alleviate this issue, Wang et al. [42] proposed the DCAD approach, which is the first attempt for CNN-based non-blind quality enhancement of HEVC-compressed images. Later, Yang et al. [46] proposed a novel QE-CNN for quality enhancement of images compressed by HEVC-MSP. Unfortunately, they are all non-blind approaches, typically requiring QP information before quality enhancement.

2.2 Blind Denoising for Images

In this section, we briefly review the CNN-based blind denoiser, as the closest field of blind quality enhancement of compressed images. The existing approaches for CNN-based blind denoising can be roughly summarized into two paradigms based on the mechanism of noise level estimation, as shown in Fig. 2 (b) and (c). The first paradigm implicitly estimates the noise level. To achieve blind denoising, images with various levels of noise are mixed and randomly sampled during training [49, 38]. Unfortunately, the performance is always far from optimum, as stated in [50, 14]. It degrades severely when there is a mismatch of noise levels between training and test data. The second paradigm explicitly estimates the noise level. It sets a noise level estimation sub-net before a non-blind denoising sub-net. For example, [14] generates a noise level map to guide the subsequent non-blind denoising. This paradigm can always yield better results than the first paradigm, yet it is not suitable for quality enhancement of compressed artifacts, mainly due to two reasons. (1) The generated noise level map cannot well represent the level of compression artifacts. The compression artifacts are much more complex than generic noise since it is always assumed to be signal-independent and white [43]. (2) Both “easy” and “hard” samples are processed in the same

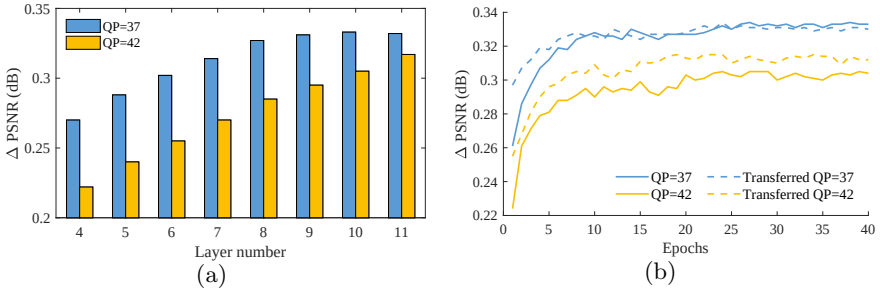

Fig. 3. (a) Average improved peak signal-to-noise ratio (ΔPSNR) of vanilla CNNs over the test set. (b) Average ΔPSNR curves alongside increased epochs, for vanilla CNN models and their transferred models over the validation set during the training stage.

deep architecture consuming equal computational resources, resulting in low efficiency. In this paper, we provide a brand-new paradigm for image reconstruction (as shown in Fig. 2 (a)) and exemplify it by our proposed RBQE on quality enhancement of compressed images. It is worth mentioning that our brand-new paradigm also has the potential for the blind denoising task.

3 Proposed Approach

In this section, we propose our RBQE approach for blind quality enhancement. Specifically, we solve three challenging problems that are crucial to the resource-efficient paradigm of our approach. (1) Which samples are "simple"/"hard" in quality enhancement? (to be discussed in Sec. 3.1) (2) How to design a dynamic network for progressive enhancement? (to be discussed in Sec. 3.2) (3) How to measure compression artifacts of enhanced compressed images for early exits? (to be discussed in Sec. 3.3)

3.1 Motivation

Our RBQE approach is motivated by the following two propositions. **Proposition 1**: "Easy" samples (i.e., high-quality compressed images) can be simply enhanced, while "hard" samples (i.e., low-quality compressed images) should be further enhanced. **Proposition 2**: The quality enhancement process with different computational complexity can be jointly optimized in a single network through an "easy to hard" manner, rather than a "hard to easy" manner.

Proof of Proposition 1. We construct a series of vanilla CNNs with different depths and feed them with "easy" and "hard" samples, respectively. Specifically, a series of vanilla CNNs with the layer number from 4 to 11 are constructed. Each layer includes $64\times3\times3$ filters, except for the last layer with $1\times3\times3$ filter. Beside, ReLU [30] activation and global residual learning [15] are adopted. The training,

validation and test sets (including 400, 100 and 100 raw images, respectively) are randomly selected from Raw Image Database (RAISE) [6] without overlapping. They are all compressed by HM16.5[3] under intra-coding configuration [37] with QP = 37 and 42 for obtaining "easy" and "hard" samples, respectively. Then, we train the vanilla CNNs with the "easy" samples, and then obtain converged models "QP = 37". Similarly, we train the CNNs with the "hard" samples and then obtain converged models "QP = 42". As shown in Fig. 3 (a), the performance of QP = 42 models improves significantly with the increase of layer numbers, while the performance of QP = 37 models gradually becomes saturated once the layer number excesses 9. Therefore, it is possible to enhance the "easy" samples with a simpler architecture and fewer computational resources, while further enhancing the "hard" samples in a more elaborate process.

Proof of Proposition 2. The advantage of the "easy to hard" strategy has been pointed out in neuro-computation [11]. Here, we investigate its efficacy on image quality enhancement through the experiments of transfer learning. If the filters learned from "easy" samples can be transferred to enhance "hard" samples more successfully than the opposite manner, then our proposition can be proved. Here, we construct 2 identical vanilla CNNs with 10 convolutional layers. The other settings conform to the above. We train these 2 models with the training sets of images compressed at QP = 37 and 42, respectively, and accordingly these 2 models are called "QP = 37" and "QP = 42". After convergence, they exchange their parameters for the first 4 layers and restart training with their own training sets. Note that the exchanged parameters are frozen during the training stage. We name the model transferred from QP = 42 to QP = 37 as "transferred QP = 37" and the model transferred from QP = 37 to QP = 42 as "transferred QP = 42". Fig. 3 (b) shows the validation-epoch curves of the original 2 models and their transferred models. As shown in this figure, the transferred QP = 42 model improves the performance of the QP = 42 model, while the transferred QP = 37 model slightly degrades the performance of the QP = 37 model. Consequently, the joint simple and elaborate enhancement process should be conducted in an "easy to hard" manner rather than a "hard to easy" manner. Besides, the experimental results of Section 4 show that the simple and elaborate enhancement process can be jointly optimized in a single network. In summary, proposition 2 can be proved. The above propositions can be also validated by JPEG-compressed images, as detailed in the supplementary material.

Given the above two propositions, we propose our RBQE approach for resource-efficient quality enhancement of compressed images in an "easy to hard" manner.

3.2 Dynamic DNN Architecture with Early-exit Strategy

Notations. In this section, we present the DNN architecture of the proposed RBQE approach for resource-efficient quality enhancement. We first introduce

[3] HM16.5 is the latest HEVC reference software.

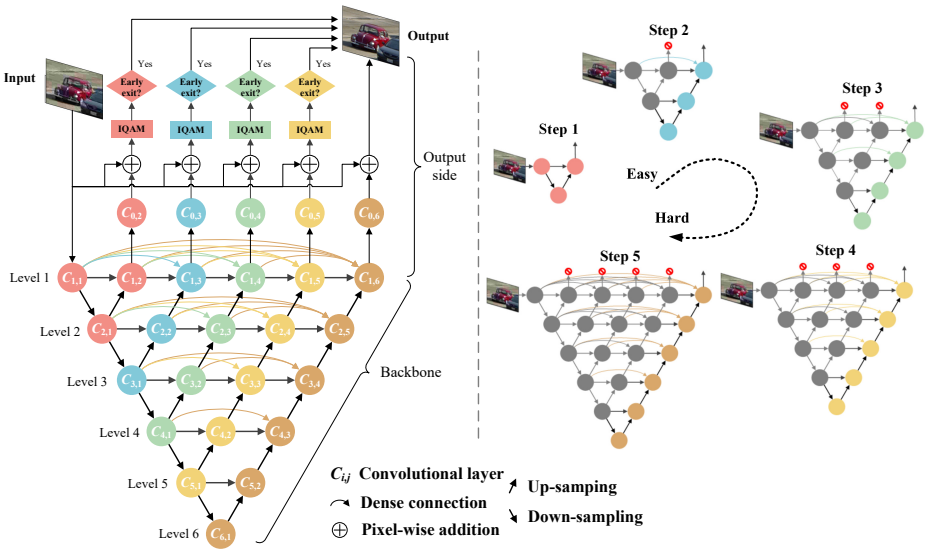

Fig. 4. Dynamic DNN architecture and early-exit strategy of our RBQE approach. The computations of gray objects (arrays and circles) are accomplished in the previous step and inherited in the current step.

the notations for our RBQE approach. The input sample is denoted by $\mathbf{S}_{\text{in}}$. The convolutional layer is denoted by $C_{i,j}$, where i denotes the level and j denotes the index of the convolutional layer on the same level. In addition, I is the total number of levels. Accordingly, the feature maps generated from $C_{i,j}$ are denoted by $\mathbf{F}_{i,j}$. The enhancement residuals are denoted by $\{\mathbf{R}_j\}_{j=2}^{I}$. Accordingly, the output enhanced samples are denoted by $\{\mathbf{S}_{\text{out},j}\}_{j=2}^{I}$.

Architecture. To better illustrate the architecture of RBQE, we separate the backbone and the output side of RBQE, as shown in the left half of Fig. 4. In this figure, we take RBQE with 6 levels as an example. The backbone of RBQE is a progressive UNet-based structure. Convolutional layers $C_{1,1}$ and $C_{2,1}$ can be seen as the encoding path of the smallest 2-level UNet, while $\{C_{i,1}\}_{i=1}^{6}$ are the encoding path of the largest 6-level UNet. Therefore, the backbone of RBQE can be considered as a compact combination of 5 different-level UNets. In the backbone of RBQE, the input sample is first fed into the convolutional layer $C_{1,1}$. After that, the feature maps generated by $C_{1,1}$ (i.e., $\mathbf{F}_{1,1}$) are progressively down-sampled and convoluted by $\{C_{i,1}, i = 2, 3, ..., 6\}$. This way, we obtain feature maps $\{\mathbf{F}_{i,1}\}_{i=1}^{6}$ at 6 different levels, the size of which progressively becomes smaller from level 1 to 6. In accordance with the encoder-decoder architecture of the UNet approach, $\{\mathbf{F}_{i,1}\}_{i=1}^{6}$ are then progressively up-sampled and convoluted until level 1. Moreover, based on the progressive UNet structure, we adopt dense connections [18] at each level. For example, at level 1, $\mathbf{F}_{1,1}$ are directly

fed into the subsequent convolutional layers at the same level: $\{C_{1,j}\}_{j=2}^{6}$. The adoption of dense connection does not only encourage the reuse of encoded low-level fine-grained features by decoders, but also largely decreases the number of parameters, leading to a lightweight structure for RBQE.

At the output side, the obtained feature maps $\{\mathbf{F}_{1,j}\}_{j=2}^{6}$ are further convoluted by independent convolutional layers $\{C_{0,j}\}_{j=2}^{6}$, respectively. In this step, we obtain the enhancement residuals: $\{\mathbf{R}_j\}_{j=2}^{6}$. For each residual $\mathbf{R}_j$, it is then added into the input sample $\mathbf{S}_{\text{in}}$ for calculating the enhanced image $\mathbf{S}_{\text{out},j}$:

$$\mathbf{S}_{\text{out},j} = \mathbf{S}_{\text{in}} + \mathbf{R}_j. \tag{1}$$

To assess the quality of enhanced image $\mathbf{S}_{\text{out},j}$, we feed it into IQAM, which is to be presented in Section 3.3.

The backbone of RBQE is motivated by [51], which extends the UNet architecture to a wide UNet for medical image segmentation. Here, we advance the wide UNet in the following aspects: (1) The wide UNet adopts deep supervision [21] directly for the feature maps $\{\mathbf{F}_{1,j}\}_{j=2}^{5}$.Here, we further process the output feature maps $\{\mathbf{F}_{1,j}\}_{j=2}^{I}$ independently through the convolutional layers in the output side $\{C_{0,j}\}_{j=2}^{I}$. This process can alleviate the interference between outputs, while slightly increase the computational costs. (2) The work of [51] manually selects one of the 4 different-level UNet-based structures in the test stage, based on the requirement for speed and accuracy. Here, we incorporate IQAM into RBQE and provide early exits in the test stage. Therefore, all UNet-based structures are progressively and automatically selected to generate the output. The early-exit strategy and proposed IQAM are presented in the following.

Early-exit Strategy. Now we explain the early-exit strategy of RBQE. Similarly, we take the RBQE structure with 6 levels as an example. The backbone of RBQE can be ablated progressively into 5 different-level UNet-based structures, as depicted in the right half of Fig. 4. For example, the smallest UNet-based structure with 2 levels consists of 3 convolutional layers: $C_{1,1}$, $C_{2,1}$ and $C_{1,2}$. In addition to these 3 layers, the 3-level UNet-based structure includes 3 more convolutional layers: $C_{3,1}$, $C_{2,2}$ and $C_{1,3}$. Similarly, we can identify the layers of the remaining 3 UNet-based structures. Note that the interval activation layers are omitted for simplicity. We denote the parameters of the i-level UNet-based structure by θ_i. This way, the output enhanced samples $\{\mathbf{S}_{\text{out},j}\}_{j=2}^{6}$ can be formulated as:

$$\mathbf{S}_{\text{out},j} = \mathbf{S}_{\text{in}} + \mathbf{R}_j(\theta_j),\ j = 2, 3, ..., 6. \tag{2}$$

In the test stage, $\{\mathbf{S}_{\text{out},j}\}_{j=2}^{6}$ are obtained and assessed progressively. That is, we first obtain $\mathbf{S}_{\text{out},2}$ and send it to IQAM. If $\mathbf{S}_{\text{out},2}$ is assessed to be qualified as the output, the quality enhancement process is terminated. Otherwise, we further obtain $\mathbf{S}_{\text{out},3}$ and assess its quality through IQAM. The same procedure applies to $\mathbf{S}_{\text{out},4}$ and $\mathbf{S}_{\text{out},5}$. If $\{\mathbf{S}_{\text{out},j}\}_{j=2}^{5}$ are all rejected by IQAM, $\mathbf{S}_{\text{out},6}$ is output

without assessment. This way, we successfully perform the early-exit strategy for "easy" samples, which are expected to output in the early stage.

3.3 Image Quality Assessment for Enhanced Images

In this section, we introduce IQAM for blind quality assessment and automatic early-exit decision. Most existing blind denoising approaches (e.g., [49, 42, 46, 14]) ignore the characteristics of compression artifacts; however, these characteristics are important to assess the compression artifacts. Motivated by [23], this paper considers two dominant factors that degrade the quality of enhanced compressed images: (1) blurring in the textured area and (2) blocky effects in the smooth area.

Specifically, the enhanced image is first partitioned into non-overlapping patches. The patches should cover all potential compression block boundaries. Then, these patches are classified into smooth and textured ones according to their sum of squared non-DC Tchebichef moment (SSTM) values that measure the patch energy [29, 23]. We take a 4×4 patch as an example, of which Tchebichef moments can be denoted by $\mathbf{M}$:

$$\mathbf{M} = \begin{pmatrix} m_{00} & \cdots & m_{03} \\ \vdots & \ddots & \vdots \\ m_{30} & \cdots & m_{33} \end{pmatrix}. \tag{3}$$

If the patch is classified as a smooth one, we evaluate its score of blocky effects $\mathcal{Q}_{\mathrm{S}}$ by calculating the ratio of the summed absolute 3rd order moments to the SSTM value [24]:

$$e_{\mathrm{h}} = \frac{\sum_{i=0}^{3} |m_{i3}|}{\left(\sum_{i=0}^{3}\sum_{j=0}^{3} |m_{ij}|\right) - |m_{00}| + C}, \tag{4}$$

$$e_{\mathrm{v}} = \frac{\sum_{j=0}^{3} |m_{3j}|}{\left(\sum_{i=0}^{3}\sum_{j=0}^{3} |m_{ij}|\right) - |m_{00}| + C}, \tag{5}$$

$$\mathcal{Q}_{\mathrm{S}} = \log_{(1-T_e)}\left(1 - \frac{e_{\mathrm{v}} + e_{\mathrm{h}}}{2}\right), \tag{6}$$

where e_{v} and e_{h} measure the energy of vertical and horizontal blocky effects, respectively; C is a small constant to ensure numerical stability; T_e is a perception threshold. The average quality score of all smooth patches is denoted by $\bar{\mathcal{Q}}_{\mathrm{S}}$. If the patch is classified as a textured one, we first blur it using a Gaussian filter. Similarly, we obtain the Tchebichef moments of this blurred patch $\mathbf{M}'$. Then, we evaluate its blurring score $\mathcal{Q}_{\mathrm{T}}$ by calculating the similarity between $\mathbf{M}$ and

$\mathbf{M}'$:

$$\mathbf{S}(i,j) = \frac{2m_{ij}m'_{ij} + C}{(m_{ij})^2 + (m'_{ij})^2 + C},\ i,j = 0,1,2,3, \tag{7}$$

$$\mathcal{Q}_{\mathrm{T}} = 1 - \frac{1}{3 \times 3}\sum_{i=0}^{3}\sum_{j=0}^{3}\mathbf{S}(i,j), \tag{8}$$

where $\mathbf{S}(i,j)$ denotes the similarity between two moment matrices. The average quality score of all textured patches is denoted by $\bar{\mathcal{Q}}_{\mathrm{T}}$. The final quality score $\mathcal{Q}$ of the enhanced image is calculated as

$$\mathcal{Q} = (\bar{\mathcal{Q}}_{\mathrm{S}})^{\alpha} \cdot (\bar{\mathcal{Q}}_{\mathrm{T}})^{\beta}, \tag{9}$$

where α and β are the exponents balancing the relative importance between blurring and blocky effects. If $\mathcal{Q}$ exceeds a threshold $T_{\mathcal{Q}}$, the enhanced image is directly output at early exits of the enhancement process. Otherwise, the compressed image needs to be further enhanced by RBQE. Please refer to the supplementary material for additional details.

The advantages of IQAM are as follow: (1) IQAM is constructed based on Tchebichef moments [29], which are highly interpretable for evaluating blurring and blocky effects. (2) The quality score $\mathcal{Q}$ obtained by IQAM is positively and highly correlated to the evaluation metrics of objective image quality, e.g., PSNR and structural similarity (SSIM) index. See the supplementary material for the validation of such correlation, which is verified over 1,000 pairs of raw/compressed images. (3) With IQAM, we can balance the tradeoff between enhanced quality and efficiency by simply tuning threshold $T_{\mathcal{Q}}$.

3.4 Loss Function

For each output, we minimize the mean-squared error (MSE) between the input compressed image and output enhanced image:

$$\mathcal{L}_j(\theta_j) = \|\mathbf{S}_{\mathrm{out},j}(\theta_j) - \mathbf{S}_{\mathrm{in}}\|_2^2,\ j = 2,3,...,I. \tag{10}$$

Although MSE is known to have limited correlation with the perceptual quality of images [44], it can still yield high accuracy in terms of other metrics, such as PSNR and SSIM [14, 12]. The loss function of our RBQE approach (i.e., $\mathcal{L}_{\mathrm{RBQE}}$) can be formulated as the weighted combination of these MSE losses:

$$\mathcal{L}_{\mathrm{RBQE}} = \sum_{j=2}^{I} w_j \cdot \mathcal{L}_j(\theta_j), \tag{11}$$

where w_j denotes the weight of $\mathcal{L}_j(\theta_j)$. By minimizing the loss function, we can obtain the converged RBQE model that simultaneously enhances the quality of input compressed images with different quality in a resource-efficient manner.

4 Experiments

In this section, we present the experimental results to verify the performance of the proposed RBQE approach for resource-efficient blind quality enhancement. Since HEVC-MSP [37] is a state-of-the-art image codec and JPEG [40] is a widely used image codec, our experiments mainly focus on quality enhancement of both HEVC-MSP and JPEG images.

4.1 Dataset

The recent works have adopted large-scale image datasets such as BSDS500 [1] and ImageNet [7], which are widely used for image denoising, segmentation and other vision tasks. However, the images of these datasets are compressed by unknown codecs and compression settings, thus containing various unknown artifacts. To obtain "clean" data without any unknown artifact, we adopt the RAISE dataset, from which 3,000, 1,000 and 1,000 non-overlapping raw images are as the training, validation and test sets, respectively. These images are all center-cropped into 512×512 images. Then, we compress the cropped raw images by HEVC-MSP using HM16.5 under intra-coding configuration [37], with QP = 22, 27, 32, 37 and 42. Note that QPs ranging from 22 to 42 can reflect the dramatically varying quality of compressed images, also in accordance with existing works [42, 46, 12]. For JPEG, we use the JPEG encoder of Python Imaging Library (PIL) [27] to compress the cropped raw images with quality factor (QF) = 10, 20, 30, 40 and 50. Note that these QFs are also used in [49].

4.2 Implementation Details

We set the number of levels $I = 6$ for the DNN architecture of RBQE. Then, $\{C_{i,1}\}_{i=1}^{6}$ are conducted by two successive $32 \times 3 \times 3$ convolutions. The other $C_{i,j}$ are conducted by two successive separable convolutions [3]. Note that each separable convolution consists of a depth-wise $k \times 3 \times 3$ convolution (k is the input channel number) and a point-wise $32 \times 1 \times 1$ convolution. The down-sampling is achieved through a $32 \times 3 \times 3$ convolution with the stride of 2, while the up-sampling is achieved through a transposed $32 \times 2 \times 2$ convolution with the stride of 2. For each group of feature maps $\mathbf{F}_{i,j}$, it is further processed by an efficient channel attention layer [41] before being feeding into other convolutional layers. Additionally, ReLU [30] nonlinearity activation is adopted between neighboring convolutions, except the successive depth-wise and point-wise convolutions within each separable convolution. For IQAM, we set $\alpha = 0.9$, $\beta = 0.1$, $C = 1e{-}8$ and $T_e = 0.05$ through a 1000-image validation. Additionally, as discussed in Sec. 4.3, $T_{\mathcal{Q}}$ is set to 0.89 and 0.74 for HEVC-MSP-compressed and JPEG-compressed images, respectively.

In the training stage, batches with QP from 22 to 42 are mixed and randomly sampled. In accordance with the "easy to hard" paradigm, we set $\{w_j\}_{j=2}^{6}$ to $\{2, 1, 1, 0.5, 0.5\}$ for QP = 22 or QF = 50, to $\{1, 2, 1, 0.5, 0.5\}$ for QP = 27 or QF = 40, to $\{0.5, 1, 2, 1, 0.5\}$ for QP = 32 or QF = 30, to $\{0.5, 0.5, 1, 2, 1\}$ for QP

Table 1. Average ΔPSNR (dB) over the HEVC-MSP and JPEG test sets.

HEVC-MSP						JPEG					
QP	CBDNet	DnCNN	DCAD	QE-CNN	RBQE	QF	CBDNet	DnCNN	DCAD	QE-CNN	RBQE
22	0.470	0.264	0.311	0.082	**0.604**	50	1.342	1.078	1.308	1.230	**1.552**
27	0.385	0.414	0.278	0.182	**0.487**	40	1.393	1.362	1.356	1.290	**1.582**
32	0.375	0.405	0.314	0.275	**0.464**	30	1.459	1.550	1.415	1.352	**1.626**
37	0.403	0.314	0.353	0.313	**0.494**	20	1.581	1.572	1.501	1.420	**1.713**
42	0.411	0.186	0.321	0.264	**0.504**	10	1.726	1.121	1.676	1.577	**1.920**
ave.	0.409	0.317	0.316	0.223	**0.510**	ave.	1.500	1.337	1.451	1.374	**1.678**

$= 37$ or QF $= 20$, and to $\{0.5, 0.5, 1, 1, 2\}$ for QP $= 42$ or QF $= 10$. This way, high-quality samples are encouraged to output at early exits, while low-quality samples are encouraged to output at late exits. We apply the Adam optimizer [20] with the initial learning rate lr $= 1e{-}4$ to minimize the loss function.

4.3 Evaluation

In this section, we validate the performance of our RBQE approach for the blind quality enhancement of compressed images. In our experiments, we compare our approach with 4 state-of-the-art approaches: DnCNN [49], CBDNet [14], QE-CNN [46] and DCAD [42]. Among them, QE-CNN and DCAD are the latest non-blind quality enhancement approaches for compressed images. For these non-blind approaches, the training batches of different QPs are mixed and randomly sampled in the training stage, such that they can also manage blind quality enhancement. Note that there is no blind approach for quality enhancement of compressed images. Thus, the state-of-the-art blind denoisers (i.e., DnCNN and CBDNet) are used for comparison, which are modified for blind quality enhancement by retraining over compressed images. For fair comparison, all compared approaches are retrained over our training set.

Evaluation on efficacy. To evaluate the efficacy of our approach, Table 1 presents the ΔPSNR results of our RBQE approach and other compared approaches over the images compressed by HEVC-MSP. As shown in this table, the proposed RBQE approach outperforms all other approaches in terms of ΔPSNR. Specifically, the average ΔPSNR of RBQE is 0.510 dB, which is 24.7% higher than that of the second-best CBDNet (0.409 dB), 60.9% higher than that of DnCNN (0.317 dB), 61.4% higher than that of DCAD (0.316 dB), and 128.7% higher than that of QE-CNN (0.223 dB). Similar results can be found in Table 1 for the quality enhancement of JPEG images.

Evaluation on efficiency. More importantly, the proposed RBQE approach is in a resource-efficient manner. To evaluate the efficiency of the RBQE approach, Fig. 5 shows the average consumed floating point operations (FLOPs)[4]

[4] Note that the definition of FLOPs follows [15, 18], i.e., the number of multiply-adds.

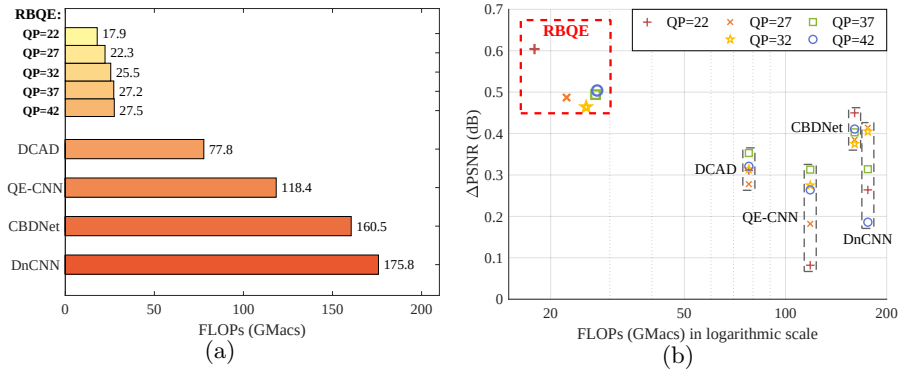

Fig. 5. (a) Average FLOPs (GMacs) over the HEVC-MSP test set. (b) Average FLOPs (GMacs) vs. improved peak signal-to-noise ratio (ΔPSNR), for blind quality enhancement by our RBQE and compared approaches over the HEVC-MSP test set.

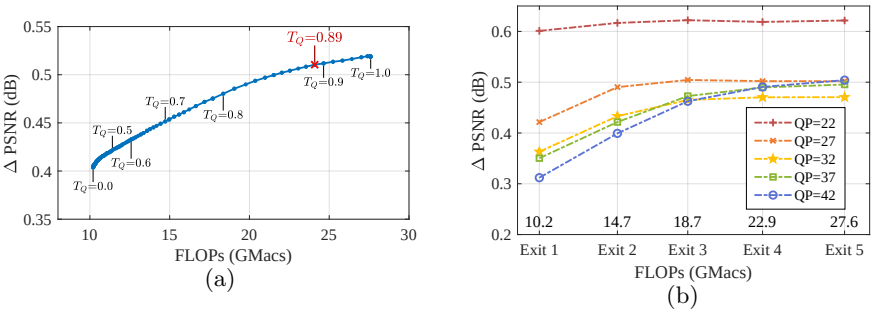

Fig. 6. (a) Average ΔPSNR and FLOPs under a series T_Q on HEVC test set. (b) Ablation results of the early-exit strategy.

by our RBQE and other compared approaches. Note that the results of Fig. 5 are averaged over all images in our test set. As can be seen in this figure, RBQE consumes only 27.5 GMacs for the "hardest" samples, i.e., the images compressed at QP = 42 and 17.9 GMacs for the "easiest" samples, i.e., the images compressed at QP = 22. In contrast, DCAD, QE-CNN, CBDNet and DnCNN consume constantly 77.8, 118.4, 160.5 and 175.8 GMacs for all samples that are either "easy" or "hard" samples compressed at 5 different QPs. Similar results can also be found for the JPEG test set, as reported in the supplementary material. In summary, our RBQE approach achieves the highest ΔPSNR results, while consuming minimal computational resources especially for "easy" samples.

Tradeoff between efficacy and efficiency. As aforementioned, we can simply control the tradeoff between efficacy and efficiency by tuning T_Q. As shown in Fig. 6 (a), the average Δ PSNR improves along with the increased consumed

FLOPs by enlarging $T_{\mathcal{Q}}$. In this paper, we choose $T_{\mathcal{Q}} = 0.89$ for HEVC-MSP-compressed images, since the improvement of average Δ PSNR gradually becomes saturated, especially when $T_{\mathcal{Q}} > 0.89$. Due to the similar reason, we choose $T_{\mathcal{Q}} = 0.74$ for JPEG-compressed images. In a word, the tradeoff between efficacy and efficiency of quality enhancement can be easily controlled in our RBQE approach.

Ablation studies. To verify the effectiveness of the early-exit structure of our RBQE approach, we progressively ablate the 5 outermost decoding paths. Specifically, for the HEVC-MSP images compressed at QP = 22, we force their enhancement process to be terminated at 5 different exits (i.e., ignoring the automatic decision by IQAM), respectively, and then we obtain the brown curve in Fig. 6 (b). Similarly, we can obtain the other 4 curves. As shown in this figure, "simplest" (i.e., QP = 22) samples can achieve ΔPSNR = 0.601 dB at the first exit, which is only 0.02 dB lower than that at the last exit. However, the expense is 270% FLOPs when outputting those samples at the last exit instead of the first one. In the opposite, the ΔPSNR of "hardest" (i.e., QP = 42) samples output from the last exit is 0.192 dB higher than that from the first exit. Therefore, "easy" samples can be simply enhanced while slightly sacrificing quality enhancement performance; meanwhile, more resources provided to "hard" samples can result in significantly higher ΔPSNR. This is in accordance with our motivation and also demonstrates the effectiveness of the early exits proposed in our RBQE approach.

5 Conclusions

In this paper, the RBQE approach has been proposed with a simple yet effective DNN structure to blindly enhance the quality of compressed images in a resource-efficient manner. Different from the traditional quality enhancement approaches, the proposed RBQE approach progressively enhances the quality of compressed images, which assesses the enhanced quality and then automatically terminates the enhancement process according to the assessed quality. To achieve this, our RBQE approach incorporates the early-exit strategy into a UNet-based structure, such that compressed images can be enhanced in an "easy to hard" manner. This way, "easy" samples can be simply enhanced and output at the early exits, while "hard" samples can be further enhanced and output at the late exits. Finally, we conducted extensive experiments on enhancing HEVC-compressed and JPEG-compressed images, and the experimental results validated that our proposed RBQE approach consistently outperforms the state-of-the-art quality enhancement approaches, while consuming minimal computational resources.

Acknowledgment This work was supported by the NSFC under Project 61876013, Project 61922009, and Project 61573037.

References

1. Arbeláez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(5), 898–916 (May 2011). https://doi.org/10.1109/TPAMI.2010.161
2. Cai, Q., Song, L., Li, G., Ling, N.: Lossy and lossless intra coding performance evaluation: HEVC, H. 264/AVC, JPEG 2000 and JPEG LS. In: Asia Pacific Signal and Information Processing Association Annual Summit and Conference. pp. 1–9. IEEE (2012)
3. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1251–1258 (2017)
4. Cisco Systems, I.: Cisco visual networking index: Global mobile data traffic forecast update, 2017-2022 white paper, https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html
5. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Transactions on Image Processing (TIP) **16**(8), 2080–2095 (2007)
6. Dang-Nguyen, D.T., Pasquini, C., Conotter, V., Boato, G.: Raise: A raw images dataset for digital image forensics. In: The 6th ACM Multimedia Systems Conference. pp. 219–224. ACM (2015)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 248–255. Ieee (2009)
8. Dong, C., Deng, Y., Change Loy, C., Tang, X.: Compression artifacts reduction by a deep convolutional network. In: IEEE International Conference on Computer Vision (ICCV). pp. 576–584 (2015)
9. Fan, Z., Wu, H., Fu, X., Huang, Y., Ding, X.: Residual-guide network for single image deraining. In: Proceedings of the 26th ACM international conference on Multimedia. pp. 1751–1759 (2018)
10. Fu, C.M., Alshina, E., Alshin, A., Huang, Y.W., Chen, C.Y., Tsai, C.Y., Hsu, C.W., Lei, S.M., Park, J.H., Han, W.J.: Sample adaptive offset in the HEVC standard. IEEE Transactions on Circuits and Systems for Video technology (TCSVT) **22**(12), 1755–1764 (2012)
11. Gluck, M.A., Myers, C.E.: Hippocampal mediation of stimulus representation: A computational theory. Hippocampus **3**(4), 491–516 (1993)
12. Guan, Z., Xing, Q., Xu, M., Yang, R., Liu, T., Wang, Z.: MFQE 2.0: A new approach for multi-frame quality enhancement on compressed video. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) pp. 1–1 (2019). https://doi.org/10.1109/TPAMI.2019.2944806
13. Guo, J., Chao, H.: Building dual-domain representations for compression artifacts reduction. In: European Conference on Computer Vision (ECCV). pp. 628–644. Springer (2016)
14. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1712–1722 (2019)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)

16. He, X., Hu, Q., Zhang, X., Zhang, C., Lin, W., Han, X.: Enhancing HEVC compressed videos with a partition-masked convolutional neural network. In: IEEE International Conference on Image Processing (ICIP). pp. 216–220. IEEE (2018)
17. Hennings-Yeomans, P.H., Baker, S., Kumar, B.V.: Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1–8. IEEE (2008)
18. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4700–4708 (2017)
19. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
20. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
21. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets. In: Artificial Intelligence and Statistics. pp. 562–570 (2015)
22. Li, K., Bare, B., Yan, B.: An efficient deep convolutional neural networks model for compressed image deblocking. In: IEEE International Conference on Multimedia and Expo (ICME). pp. 1320–1325. IEEE (2017)
23. Li, L., Zhou, Y., Lin, W., Wu, J., Zhang, X., Chen, B.: No-reference quality assessment of deblocked images. Neurocomputing **177**, 572–584 (2016)
24. Li, L., Zhu, H., Yang, G., Qian, J.: Referenceless measure of blocking artifacts by Tchebichef kernel analysis. IEEE Signal Processing Letters **21**(1), 122–125 (2013)
25. Li, S., Xu, M., Ren, Y., Wang, Z.: Closed-form optimization on saliency-guided image compression for HEVC-MSP. IEEE Transactions on Multimedia (TMM) **20**(1), 155–170 (2017)
26. Liu, Y., Hamidouche, W., Déforges, O., Lui, Y., Dforges, O.: Intra Coding Performance Comparison of HEVC, H.264/AVC, Motion-JPEG2000 and JPEGXR Encoders. Research report, IETR/INSA Rennes (Sep 2018), https://hal.archives-ouvertes.fr/hal-01876856
27. Lundh, F.: Python imaging library (PIL), http://www.pythonware.com/products/pil
28. Marcellin, M.W., Gormish, M.J., Bilgin, A., Boliek, M.P.: An overview of JPEG-2000. In: Data Compression Conference (DCC). pp. 523–541. IEEE (2000)
29. Mukundan, R., Ong, S., Lee, P.A.: Image analysis by Tchebichef moments. IEEE Transactions on Image Processing (TIP) **10**(9), 1357–1364 (2001)
30. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: The 27th International Conference on Machine Learning (ICML). pp. 807–814 (2010)
31. Nguyen, T., Marpe, D.: Performance analysis of HEVC-based intra coding for still image compression. In: Picture Coding Symposium (PCS). pp. 233–236. IEEE (2012)
32. Nguyen, T., Marpe, D.: Objective performance evaluation of the HEVC main still picture profile. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) **25**(5), 790–797 (2014)
33. Norkin, A., Bjontegaard, G., Fuldseth, A., Narroschke, M., Ikeda, M., Andersson, K., Zhou, M., Van der Auwera, G.: HEVC deblocking filter. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) **22**(12), 1746–1754 (2012)
34. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: A better and simpler baseline. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp. 3937–3946 (2019)

35. Seshadrinathan, K., Soundararajan, R., Bovik, A.C., Cormack, L.K.: Study of subjective and objective quality assessment of video. IEEE Transactions on Image Processing (TIP) **19**(6), 1427–1441 (2010)
36. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
37. Sullivan, G.J., Ohm, J.R., Han, W.J., Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) **22**(12), 1649–1668 (2012)
38. Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: IEEE International Conference on Computer Vision (ICCV). pp. 4539–4547 (2017)
39. Tan, T.K., Weerakkody, R., Mrak, M., Ramzan, N., Baroncini, V., Ohm, J.R., Sullivan, G.J.: Video quality evaluation methodology and verification testing of HEVC compression performance. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) **26**(1), 76–90 (2015)
40. Wallace, G.K.: The JPEG still picture compression standard. IEEE Transactions on Consumer Electronics (TCE) **38**(1), xviii–xxxiv (Feb 1992). https://doi.org/10.1109/30.125072
41. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-Net: Efficient channel attention for deep convolutional neural networks. arXiv preprint arXiv:1910.03151 (2019)
42. Wang, T., Chen, M., Chao, H.: A novel deep learning-based method of improving coding efficiency from the decoder-end for HEVC. In: Data Compression Conference (DCC). pp. 410–419. IEEE (2017)
43. Wang, Z., Liu, D., Chang, S., Ling, Q., Yang, Y., Huang, T.S.: D3: Deep dual-domain based fast restoration of JPEG-compressed images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2764–2772 (2016)
44. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing (TIP) **13**(4), 600–612 (2004)
45. Xu, M., Li, T., Wang, Z., Deng, X., Yang, R., Guan, Z.: Reducing complexity of hevc: A deep learning approach. IEEE Transactions on Image Processing (TIP) **27**(10), 5044–5059 (2018)
46. Yang, R., Xu, M., Liu, T., Wang, Z., Guan, Z.: Enhancing quality for HEVC compressed videos. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) (2018)
47. Yang, R., Xu, M., Wang, Z., Li, T.: Multi-frame quality enhancement for compressed video. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6664–6673 (2018)
48. Zhang, H., Yang, J., Zhang, Y., Nasrabadi, N.M., Huang, T.S.: Close the loop: Joint blind image restoration and recognition with sparse representation prior. In: IEEE International Conference on Computer Vision (ICCV). pp. 770–777. IEEE (2011)
49. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing (TIP) **26**(7), 3142–3155 (2017)
50. Zhang, K., Zuo, W., Zhang, L.: FFDNet: Toward a fast and flexible solution for CNN-based image denoising. IEEE Transactions on Image Processing (TIP) **27**(9), 4608–4622 (2018)

51. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: UNet++: A nested U-Net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11. Springer (2018)