Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Learning-based parameter prediction for quality control in three-dimensional medical image compression*

Yuxuan HOU[†1], Zhong REN[†‡1], Yubo TAO[1], Wei CHEN[2]

[1]*State Key Lab of CAD & CG, Zhejiang University, Hangzhou 310058, China*

[2]*The First Affiliated Hospital, Zhejiang University, Hangzhou 310003, China*

[†]E-mail: 3140104190@zju.edu.cn; renzhong@cad.zju.edu.cn

Received May 16, 2020; Revision accepted Oct. 8, 2020; Crosschecked Aug. 24, 2021

Abstract: Quality control is of vital importance in compressing three-dimensional (3D) medical imaging data. Optimal compression parameters need to be determined based on the specific quality requirement. In high efficiency video coding (HEVC), regarded as the state-of-the-art compression tool, the quantization parameter (QP) plays a dominant role in controlling quality. The direct application of a video-based scheme in predicting the ideal parameters for 3D medical image compression cannot guarantee satisfactory results. In this paper we propose a learning-based parameter prediction scheme to achieve efficient quality control. Its kernel is a support vector regression (SVR) based learning model that is capable of predicting the optimal QP from both video-based and structural image features extracted directly from raw data, avoiding time-consuming processes such as pre-encoding and iteration, which are often needed in existing techniques. Experimental results on several datasets verify that our approach outperforms current video-based quality control methods.

Key words: Medical image compression; High efficiency video coding (HEVC); Quality control; Learning-based
https://doi.org/10.1631/FITEE.2000234 **CLC number:** TP391

1 Introduction

To date, the volume of medical imaging data has dramatically grown with an increasing number of three-dimensional (3D) medical devices and growing healthcare demand. Due to bandwidth and storage limitations, it is vital to process medical images through lossy compression.

The medical image compression methods can be classified into two categories regarding the encoding process: classical methods and video-based methods. In the first category, tree-based methods, such as set partitioning in hierarchical trees (SPIHT) (Said and Pearlman, 1996), can achieve a fair compression ratio, but they suffer from distortion due to the lack of spatial information. Other methods such as the Joint Photographic Expert Group (JPEG) standard reduce distortion but have low performance on large-size medical data. In the second category video compression techniques are used, as video and 3D medical data share the same data modality. The state-of-the-art model in video compression is high efficiency video coding (HEVC), which achieves size reduction in the range of 50% for equal perceptual quality with its predecessor H.264/AVC, providing new possibilities for the compression of 3D medical images. Though HEVC is computationally expensive, it could be accelerated by specially designed hardware encoders in graphic processing units (GPUs). As such, the Digital Imaging and Communications in Medicine (DICOM) group has proposed to add HEVC to the official standard of communication in medical imaging (Liu et al., 2017).

Quality control (distortion control) is important to ensure correct diagnosis in medical applications.

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 61890954)

ORCID: Yuxuan HOU, https://orcid.org/0000-0002-0880-6418; Zhong REN, https://orcid.org/0000-0002-6798-3035

It is also essential in video compression due to the quality constraints of video storage. Note that the quantization parameter (QP) used in HEVC is vital for video quality control, as it is highly associated with the quantization step size in the encoding process. Thus, compression towards a target quality constraint can be formulated as a distortion-quantization (D-Q) model, with distortion presented in the form of peak signal-to-noise ratio (PSNR) (Huynh-Thu and Ghanbari, 2008). Many studies have analyzed the D-Q models for video compression. In the earliest studies, simple linear D-Q models (Ma S et al., 2005; Wang HL and Kwong, 2008) were proposed. However, they have low performance due to the lack of generality. Later, the distribution of transform coefficients in video coding is used to construct the D-Q models (Kamaci et al., 2005; Kwon et al., 2007; Ma SW et al., 2012; Pan and Chen, 2016). Their performances are bounded, however, because the real distribution does not always follow the supposed mathematical distribution. Wu and Su (2013) proposed models based on making several encoding/decoding passes iteratively. Dinh et al. (2018) and Santamaria et al. (2018) proposed models based on massive observations on a large number of benchmark video sequences. They yielded reasonable estimations of the D-Q curve at the cost of a large amount of time and a large storage space.

Even though a large number of studies have targeted D-Q models in video compression, the direct application of video-based quality control methods to medical image compression cannot guarantee satisfactory results. Two main challenges exist: first, video-based algorithms concentrate on motion vectors in video frames rather than on the 3D structural information, which is important in medical images; second, these methods need much time and space, while medical applications are typically time-critical or space-bounded, especially in real-time telemedicine systems. Moreover, their model parameters are hard to determine because of the changing characteristics of data.

To resolve these problems, we propose a learning-based quantization parameter prediction scheme for quality control in 3D medical imaging data compression. The kernel of this scheme is a support vector regression (SVR) (Schölkopf et al., 2000) based learning model, capable of predicting the optimal QP from video-based and structural image features extracted directly from the raw data, without time-consuming processes such as pre-encoding or iteration. Experimental results for a variety of datasets indicate that our proposed approach can generate more precise predictions in the application of medical image compression than conventional D-Q models. Results also show that this method obtains good results for time requirement.

2 Related works

2.1 Quality control models in non-video-based medical image compression

Many efforts have been made to achieve a proper compression ratio while satisfying specific quality constraints of medical image compression. Miaou and Chen (2004) combined the distortion-constrained codebook replenishment (DCCR) mechanism and the SPIHT technique while presenting an iterative search algorithm to reach the target image quality. Lazzerini et al. (2010) generated a family of optimal quantization tables that produce different trade-offs between the image compression ratio and quality in the JPEG encoding process. These methods employ iterative search at high time cost, but they are instructive for quality control in video-based medical image compression.

2.2 D-Q models in video compression

With regard to model characteristics, the D-Q models for video compression can be classified into two classes: empirical models and analytical models. The empirical models have simple forms. A linear rate-distortion model between PSNR and QP was introduced in Ma S et al. (2005). Wang HL and Kwong (2008) proposed a similar model where the mean squared error (MSE) is proportional to the quantization step size. A major drawback of empirical methods is that they cannot handle outliers, resulting in a loss of precision. Analytical models are based on certain mathematical assumptions. For example, one such model uses the Cauchy density function to fit the distribution of the discrete cosine transform (DCT) coefficients and derive an approximated D-Q step model, which yields a more accurate result than its predecessors (Kamaci et al., 2005). Another two models use the sum of absolute transform differences

(SATD) based on the Hadamard transform (Pratt et al., 1969) as an important parameter for D-Q modeling (Kwon et al., 2007; Ma SW et al., 2012). Even though analytical models outperform their empirical counterparts, their basic assumptions are not always consistent with the distributions of the DCT coefficients, thus failing to obtain high precision.

A novel learning-based approach applies the SVR learning method to approximate the relationship among QP, SATD, and PSNR (Pan and Chen, 2016). However, it considers only two features (PSNR and SATD), resulting in a lack of generality. Recently, multi-stage methods have been presented to improve the performance of D-Q models. For instance, one such model extracts feature vectors to describe the relationship between D and Q (Wu and Su, 2013). The initial weights of features are determined by extended simulations, while the weights are refined frame by frame. Nonetheless, the iterative optimization process in this model makes it time consuming. A convolutional neural network named VQANet was introduced by Dinh et al. (2018) to form the D-Q model, which yields good results. Due to the application of 3D convolutional layers, its computation and storage complexities are high. Besides, structural information may be lost because the images are clipped into small blocks before they are fed into the network. Another neural network based method (Santamaria et al., 2018) uses a two-dimensional convolutional network with skip connections to predict distortion. However, it suffers from a loss of structural information due to the similar image clipping process.

2.3 Learning-based regression models in medical imaging

Several learning-based methods have been proposed in medical imaging, including K-means, SVR, decision trees, and neural networks (Wang SJ and Summers, 2012). Of these, the dominant regression methods used are neural networks and SVR. Neural networks are suitable for search in a dense space due to the limit of the number of parameters, while SVR based methods are good at solving sparse linear space problems because of the linear property of SVR. For example, the SVR model outperforms the general neural network model for content-based image retrieval in digital mammography (El-Naqa et al., 2004). Moreover, SVR was used to enhance the conventional D-Q model using SATD (Ma SW et al., 2012), with good results achieved (Pan and Chen, 2016). In the aforementioned cases, SVR is preferred as it is robust and effective in processing data with a sparse representation. Thus, SVR is a promising method for solving the D-Q model of medical imaging with generality and robustness in time-critical cases.

3 Our proposed method

The optimal parameter estimation problem can be formed as follows:

$$f(V, \mathrm{PSNR}_{\mathrm{target}}) \rightarrow \mathrm{QP}_{\mathrm{opt}}, \tag{1}$$

where V denotes the volume data, $\mathrm{PSNR}_{\mathrm{target}}$ the target distortion, and $\mathrm{QP}_{\mathrm{opt}}$ the optimal QP for the compression to meet the target distortion. Note that both PSNR and the compressed file size will go down with the rise of QP as the quantization step size becomes larger. Thus, the optimal QP is the largest QP that meets the target PSNR constraints.

Our approach extracts the essential features of the video compression of 3D medical images and employs a learning-based model for solving the D-Q problem based on SVR.

3.1 Pipeline

As shown in Fig. 1, our proposed approach consists of two stages: training and testing. In the training stage, we use the brute-force method to find the optimal QP of all target PSNRs for all training data, and then put the $<\mathrm{PSNR}_{\mathrm{target}}, \boldsymbol{x}, \mathrm{QP}_{\mathrm{opt}}>$ tuple into the regression model, where $\boldsymbol{x}$ denotes the feature vector. In the testing stage, the SVR method produces an optimal QP using the target PSNR and the features extracted from the test medical data. As mentioned in Section 1, the video-based features important for video compression and the structural features important for medical image compression are combined. This combination makes our D-Q model capable of handling the video-based compression of medical images. What is more, this is achieved more efficiently by the SVR method than by conventional D-Q models.

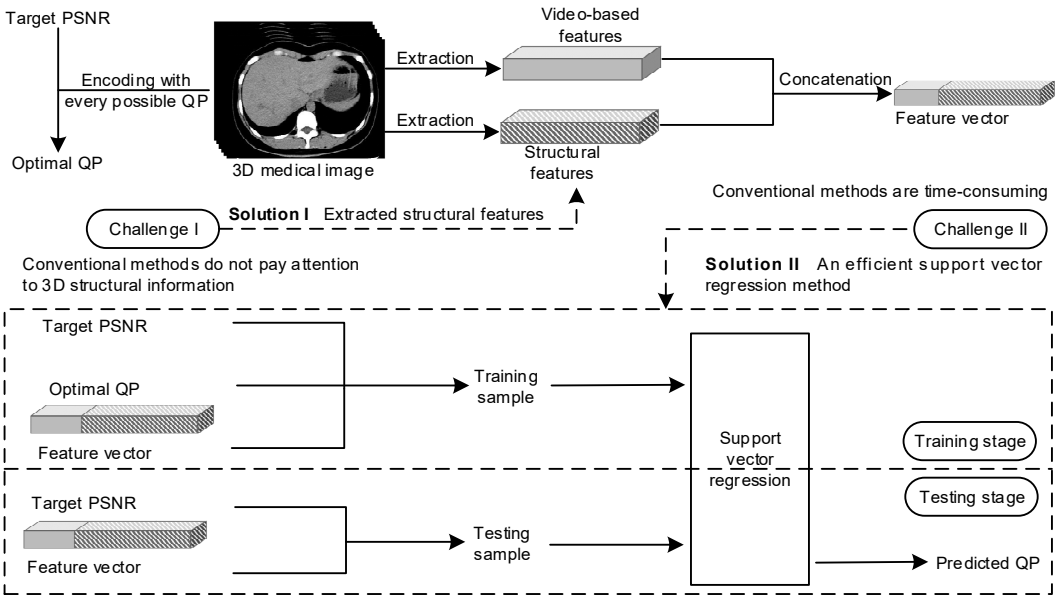

Fig. 1 The pipeline of our work

3.2 Feature extraction

The selected features, i.e., video-based features and structural features, are extracted to represent the entire volume data.

3.2.1 Video-based features

One of the most important video-based features is the number of video frames, i.e., the number of slices in the volume:

$$x_1 = x_{\text{Z-slice}} = n. \tag{2}$$

Another feature is the ratio of zero-valued voxels. In HEVC, coding is based on the quantization of the difference (in other words, residual) between adjacent frames. There is no difference between blank regions, which can be encoded efficiently. For a similar reason, the mean difference of the adjacent slices in the 3D medical image is computed. These two features can be written as

$$x_2 = x_{\text{zero-ratio}} = \sum\nolimits_k 1(V[k]) \Big/ \sum\nolimits_k, \tag{3}$$

$$x_3 = x_{\text{mean-difference}} = \frac{1}{n}\sum_{i=1}^{n-1}\sum_{k} |\, S_i[k] - S_{i+1}[k] \,|, \tag{4}$$

where k denotes the voxel index of the volume, $V[k]$ the voxel intensity, $1(\cdot)$ the indicator function, and S_i the i^{th} image slice.

3.2.2 Structural features

Textural features are important structural features of medical images. The gray-level co-occurrence matrix (GLCM) (Haralick et al., 1973) is widely used to measure textures by considering the spatial relationship among pixels. It characterizes the texture of an image by calculating how combinations of discretized intensities (gray levels) of neighboring pixels are distributed along one direction of the volume. If the grayscale is separated into m levels, the frequencies of (0, 0), …, (0, m−1), …, (m−1, m−1) pairs of pixels are measured in every direction. Probability distributions are then calculated to determine high- order statistical GLCM features. The matrices are computed as follows:

$$\begin{cases} F_{\Delta x,\Delta y}(i,j) = \sum\limits_{p=1}^{h_1}\sum\limits_{q=1}^{h_2}\delta_{ij}(p,q), \\ \delta_{ij}(p,q) = \begin{cases} 1, & \text{if } X(p,q) = i \text{ and} \\ & X(p+\Delta x, q+\Delta y) = j, \\ 0, & \text{otherwise,} \end{cases} \end{cases} \tag{5}$$

$$P_{\Delta x,\Delta y}(i,j) = \sum\nolimits_{\Delta x,\Delta y}(i,j) \Big/ \sum\nolimits_{i,j} F_{\Delta x,\Delta y}(i,j), \tag{6}$$

where $\boldsymbol{X}$ is the leveled grayscale matrix, h_1 and h_2 are the width and height of $\boldsymbol{X}$ respectively, $\boldsymbol{F}$ is the frequency matrix, and $\boldsymbol{P}$ is the subsequent probability matrix. (Δx, Δy) denotes the direction vector. For

example, (Δx, Δy)=(1, 0) represents the 0° direction from the positive direction of the x axis.

Five textural features, energy, contrast, correlation, entropy, and homogeneity (inverse difference moment, IDM), are extracted from the image volumes and analyzed using the texture of four directions (0°, 45°, 90°, 135°), resulting in 20 textural features in total. These features are calculated as follows:

$$x_4 = x_{\text{contrast}} = \sum_{i,j=0}^{m-1} P_{i,j}(i-j)^2, \tag{7}$$

$$x_5 = x_{\text{IDM}} = \sum_{i,j=0}^{m-1} \frac{P_{i,j}}{1+(i-j)^2}, \tag{8}$$

$$x_6 = x_{\text{energy}} = \sqrt{\sum_{i,j=0}^{m-1} P_{i,j}^2}, \tag{9}$$

$$x_7 = x_{\text{entropy}} = \sum_{i,j=0}^{m-1} P_{i,j}\left(-\ln P_{i,j}\right), \tag{10}$$

$$x_8 = x_{\text{correlation}} = \sum_{i,j=0}^{m-1} P_{i,j}\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{\sigma_i^2\sigma_j^2}}, \tag{11}$$

where μ denotes the mean, and σ_i and σ_j the standard variations of row i and column j of $\boldsymbol{P}$, respectively. The averages of these features in each slice of the volume are calculated to form the actual feature vectors.

3.3 Proposed learning-based prediction model

With certain features extracted, the prediction problem can be formulated as

$$f(x_1, x_2, \cdots, x_M, \text{PSNR}_{\text{target}}) \to \text{QP}_{\text{opt}}, \tag{12}$$

where M denotes the total number of features. Let $\boldsymbol{x}$=(x_1, x_2, …, x_M, $\text{PSNR}_{\text{target}}$) and y=QP_{opt}. Then the problem can be rewritten as

$$f(\boldsymbol{x}) \to y. \tag{13}$$

Because QP_{opt} is located in the real domain (an integer between 1 and 30), the approximation of f can be regarded as a regression problem using the SVR method. For nonlinear regression, the input data vectors are mapped into a higher dimensional space through a nonlinear mapping $g(\boldsymbol{x})$, and linear regression is applied in the mapped space. That is, the regression function can be written as

$$y \approx f(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}} g(\boldsymbol{x}) + b, \tag{14}$$

where $\boldsymbol{w}$ and b are the linear regression parameters in the mapped space.

Our proposed model uses the Gaussian radial basis function (RBF) kernel. The kernel function is written as

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2) = g(\boldsymbol{x}_1)^{\mathrm{T}} g(\boldsymbol{x}_2) = \exp\left(-\frac{\left\|\boldsymbol{x}_1 - \boldsymbol{x}_2\right\|^2}{2\sigma^2}\right), \tag{15}$$

where σ is a scaling parameter of the RBF kernel. The best solution of f is

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}}\left(\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + C\left(v\epsilon + \frac{1}{l}\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right)\right)\right) \tag{16}$$

subject to

$$\boldsymbol{w}^{\mathrm{T}} g(\boldsymbol{x}_i) + b - y_i \le \epsilon + \xi_i,$$
$$y_i - (\boldsymbol{w}^{\mathrm{T}} g(\boldsymbol{x}_i) + b) \le \epsilon + \xi_i^*, \xi_i^* \ge 0,\ \xi_i \ge 0,\ \epsilon > 0,$$

where l is the number of training samples, C and v are the weight parameters, ϵ is the margin size of the support vectors that contribute to the generality of the algorithm, and ξ_i and ξ_i^* are slack variables for error tolerance.

To create the training data, images are chosen from the datasets and compressed with varied QPs from 1 to 30. The resultant image distortion (i.e., quality, measured by PSNR) is then recorded. The corresponding optimal QP for every target PSNR from 10 to 50 dB is then searched. The training sample ($\boldsymbol{x}$, y) is formulated as

$$\boldsymbol{x} = (\boldsymbol{x}^{(i)}, P^{(j)}), \tag{17}$$

$$y = \text{QP}_{\text{opt}}^{(i,j)}, \tag{18}$$

for each $i \in \{1, 2, \ldots, l\}$ and $P^{(j)} \in \{10, 11, \ldots, 50\}$.

The $\text{QP}_{\text{opt}}^{(i,j)}$ is the optimal QP for the i^{th} training data X_i to meet the target quality $P^{(j)}$. As mentioned at the beginning of Section 3, the optimal QP is the largest QP that meets the target PSNR. Optimal QPs for the training cases are calculated as

$$\text{QP}_{\text{opt}}^{(i,j)} = \max(q) \text{ subject to } P^{(i,q)} > P^{(j)} \tag{19}$$

for each $i\in\{1, 2, \ldots, l\}$, $q\in\{1, 2, \ldots, 30\}$, and $P^{(j)}\in\{10, 11, \ldots, 50\}$. Here $P^{(i,q)}$ denotes the corresponding PSNR of using QP=q to encode the i^{th} training data, whereas $P^{(j)}$ denotes the j^{th} target PSNR.

In the testing stage, with the extracted feature vectors and specified target distortion (quality) of a testing image X', the predicted QP_{opt} can be calculated efficiently through the trained SVR model. The complete algorithm for the SVR-based D-Q prediction process is summarized in Algorithm 1.

Algorithm 1 Prediction process of the proposed model

For $i\in\{1, 2, \ldots, l\}$
  Calculate the feature vector $\boldsymbol{x}^{(i)}$
  For $q\in\{1, 2, \ldots, 30\}$
    Pre-encode the training data X_i with QP=q and obtain the corresponding distortion $P^{(i,q)}$
  End For
  For $j\in\{10, 11, \ldots, 50\}$
    Let $P^{(j)}=j$
    Calculate $\text{QP}_{\text{opt}}^{(i,j)}$ by Eq. (19)
  End For
End For
Use $\left(\boldsymbol{x}=(\boldsymbol{x}^{(i)}, P^{(j)}), y=\text{QP}_{\text{opt}}^{(i,j)}\right)$ as the training data to construct the SVR model S
Calculate the feature vector $\boldsymbol{x}'$ for the testing image X'
Input ($\boldsymbol{x}'$, $\text{PSNR}_{\text{target}}$) into the model S to obtain the predicted optimal QP q^{pred}
Use q^{pred} to encode X' and obtain the corresponding encoding distortion $\text{PSNR}_{\text{pred}}$

4 Experiments and discussions

Experiments are conducted using Intel Core i5-4690 @3.55 GHz CPU (with 16 GB RAM) and NVIDIA GeForce GTX 960 GPU (with 4 GB RAM). The NVIDIA Video Codec SDK (Patait and Young, 2016) is employed to run HEVC. This interface is specially designed for commercial GPUs to accelerate the encoding/decoding processes, and is suitable for general usage. The SVR method implementation is based on the NuSVR method in the Scikit-Learn package in Python. Several open datasets of 3D medical images from The Cancer Imaging Archive (TCIA) (Clark et al., 2013) are used for testing. The details of these datasets are shown in Table 1.

Table 1 Data configuration

Dataset	Number of cases	Property	Width×Height	Snapshot
LIDC-IDRI	306	Liver CT	512×512 (axial)	
RIDER Lung CT	60	Lung CT	512×512 (axial)	
LungCT-Diagnosis	62	Lung CT	512×512 (axial)	
REM-BRANDT	109	Brain CT	256×256 (coronal, sagittal)	
TCGA-HNSC	51	Head and neck CT	512×512 (coronal, sagittal)	

The most commonly used video format, YUV 4:2:0, consists of a luma channel (Y) and two subsampled chroma channels (U/V). However, 3D medical imaging data have a single channel. One way around this is to normalize the computed tomography (CT) volumes into grayscale images and add zero-valued chroma components to form the YUV 4:2:0 sequences (Sanchez and Bartrina-Rapesta, 2014). The sequences are encoded with every QP from 1 to 30 under the fixed-QP configuration. Note that the QP ranges from 1 to 51 in HEVC settings. Too large QP, however, will cause severe loss of image quality, which is not acceptable for clinical diagnosis. The resulting PSNR of every compressed image is recorded. Seventy-five percent images are used for training, while the others are used for testing.

4.1 Model refinement

4.1.1 Hyper-parameters

The influence of three hyper-parameters in the regression model, i.e., the weight parameters C and v and the variance parameter of the Gaussian kernel $\gamma=1/(2\sigma^2)$, is examined. The values of these parameters are shown in Table 2.

Table 2 Values of the three hyper-parameters

Hyper-parameter	Values
C	$2^{-3}, 2^{-2}, \ldots, 2^{6}$
v	0.1, 0.2, …, 1.0
γ	$2^{-5}, 2^{-4}, \ldots, 2^{4}$

Each of the three parameters has 10 possible values to choose, resulting in 1000 combinations of

parameters in total. The relationship between the correction of the model and the variances of hyper-parameters in the LIDC-IDRI dataset (which is the largest dataset) is shown in Fig. 2.

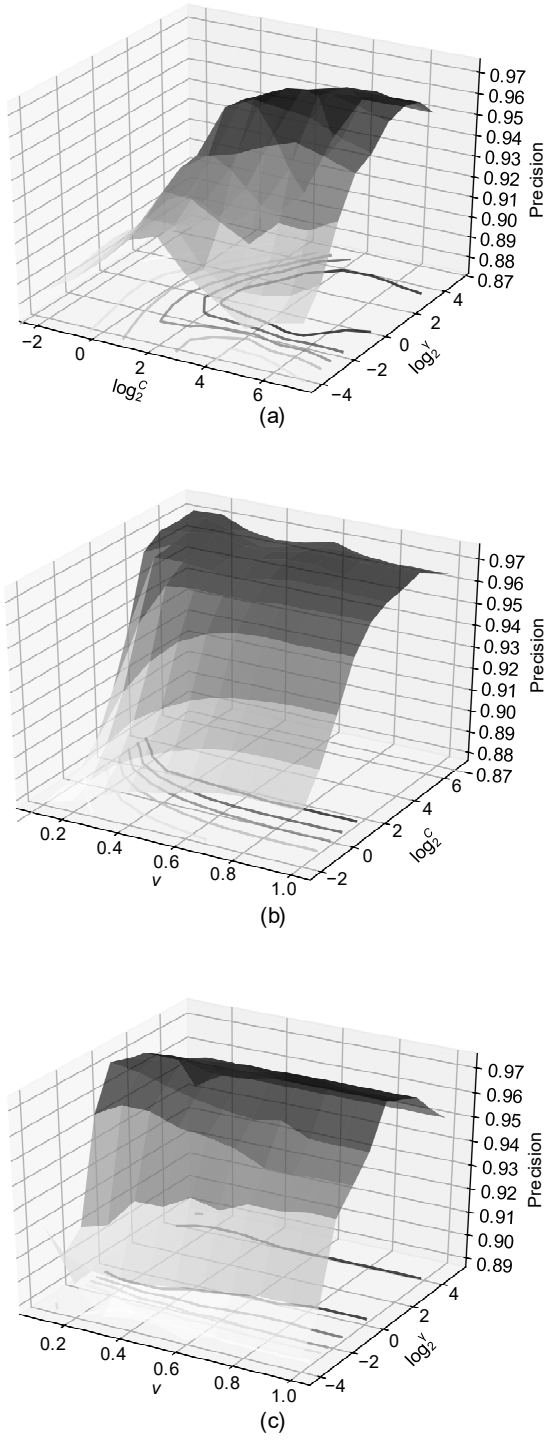

Fig. 2 Performance with different combinations of hyper-parameters in the LIDC-IDRI dataset with one of the three parameters being fixed and the others free: (a) v=0.2; (b) γ=2; (c) C=64

The correction of the model is calculated as follows:

$$\text{Correction} = 1 - \text{mean}\left(\frac{\left|\text{PSNR}_{\text{target}} - \text{PSNR}_{\text{pred}}\right|}{\text{PSNR}_{\text{target}}}\right), \quad (20)$$

where $\text{PSNR}_{\text{target}}$ denotes the target PSNR, and $\text{PSNR}_{\text{pred}}$ the PSNR reached using the predicted optimal QP.

Through grid search, the combination of (C=2^5, v=0.1, γ=1) achieves the highest average correction, hence the best generality among all five datasets. This combination is used as the standard setting of the SVR model.

4.1.2 Feature selection

Initially, 23 features are used to form the feature vector. Merging the same texture features of different directions results in eight features. Thus, 2^8 combinations are possible (the empty set means that only the target PSNR is used for prediction). The precision produced in each combination is tested in the LIDC-IDRI dataset with the hyper-parameters (C=2^5, v=0.1, γ=1). The ablation study results are listed in Table 3, whereas the leave-one-out prediction results are listed in Table 4.

Table 3 Ablation study results

Feature	Correction
No additional feature	0.898
Video-based feature only	0.947
Structural feature only	0.967
Both	0.971

Table 4 Leave-one-out prediction results of eight features

Feature eliminated	Correction	Feature eliminated	Correction
Z-slice	0.972	IDM	0.971
Zero-ratio	0.971	Energy	0.970
Mean-difference	0.970	Entropy	0.973
Contrast	0.968	Correlation	0.971

Considering the performances of all combinations, our findings are as follows. First, ablation study results show that the optimal solution is reached by combining video-based and structural features. Second, x_{entropy} is the least significant feature, while

x_{contrast} is the most significant feature in the leave-one-out test.

4.2 Performance comparisons

The accuracy of the proposed D-Q curve estimation is examined. Results are provided for the proposed model and the five other D-Q models, i.e., the model based on the Cauchy density function (Kamaci et al., 2005) (denoted by model I), the model using SATD to predict QP (Ma SW et al., 2012) (denoted by model II), the model using SVR to approximate the map among PSNR, SATD, and QP (Pan and Chen, 2016) (denoted by model III), and the two models using convolutional neural networks, Dinh et al. (2018)'s model (denoted by model IV) and Santamaria et al. (2018)'s model (denoted by model V).

Model I contains two sequence-dependent parameters. To determine these parameters appropriately, each sequence is encoded twice using two extreme QP values (i.e., 1 and 30) to obtain distinct D-Q data points. Linear regression is applied to acquire the parameters. Model II has a form similar to that of model I, but the SATDs of the sequences are calculated at the same time to help predict an optimal QP. Once the model parameters have been determined, both models can predict the complete D-Q curve for this sequence, and identify the optimal QP. Model III and our model are constructed without any encoding process, because the two models use features extracted directly from the raw data. Models IV and V split the data into the blocks, feed them into the neural networks, and reach the final result by averaging all result vectors of these blocks.

The target PSNR for testing ranges from 10 to 50, and the real optimal QP for the target PSNR is searched from the training dataset. The mean errors of the methods at every target PSNR in the LIDC-IDRI dataset are shown in Fig. 3. It can be concluded that our method has better performance than others.

The correction of the models is calculated using Eq. (20). Table 5 reports their prediction accuracy among all datasets.

It is not surprising that model III outperforms its predecessor, model II, with respect to average. Also, note that the performance of model I that is based on the Cauchy distribution is highly content-dependent;

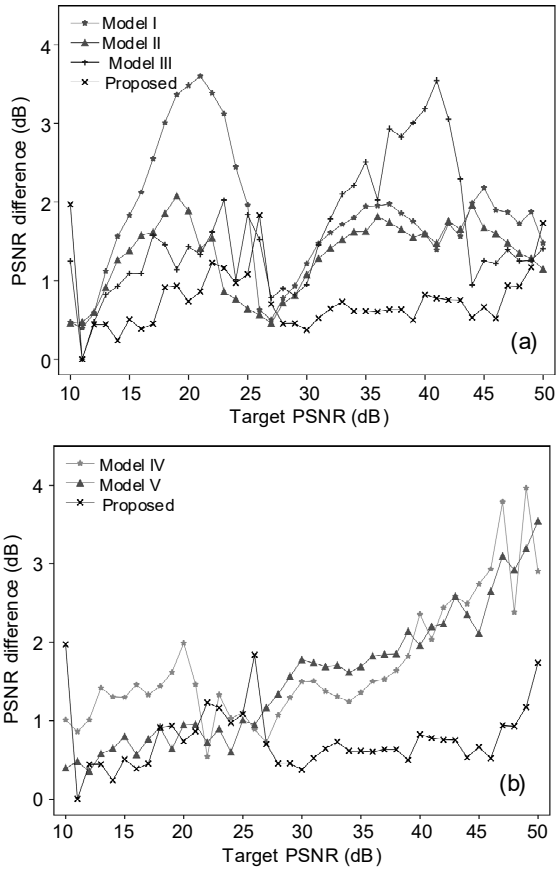

Fig. 3 The mean error of PSNR vs. the target PSNR of six models: (a) models I, II, and III vs. the proposed model; (b) models IV and V vs. the proposed model
Our model obtains the smallest error

Table 5 The correction of D-Q models

Dataset	Correction					
	Model I	Model II	Model III	Model IV	Model V	Ours
LIDC-IDRI	0.935	0.951	0.940	0.946	0.951	**0.973**
RIDER Lung CT	0.879	0.930	0.956	0.955	0.949	**0.959**
LungCT-Diagnosis	0.964	0.962	0.966	0.957	0.965	**0.980**
REMBRANDT	0.967	0.932	0.965	0.941	0.961	**0.980**
TCGA-HNSC	0.958	0.962	0.962	0.968	0.967	**0.977**
Average	0.941	0.947	0.958	0.953	0.959	**0.974**

i.e., the performance of this model is case-sensitive. This is because the real distribution of source data is not always consistent with the Cauchy distribution assumption. Model III and our model both use SVR to predict the D-Q model, but model III uses only two features, i.e., PSNR and SATD, while our proposed model uses more than 20 features and achieves higher correction. Models IV and V, which use neural networks, perform moderately in precision because 3D medical images are too large for them to handle. Meanwhile, these two block-based methods suffer from loss of structural information (especially when the target PSNR is high), and this loss becomes the main contributor to the total error. Finally, our method outperforms all the reference models for precision.

4.3 Inference time comparisons

The inference times of the six models are listed in Table 6. The total time cost is divided into encoding/decoding time, feature extraction time, and prediction time. Results are tested on the largest dataset LIDC-IDRI.

Among the models, models I and II need extra encoding/decoding processes using the lowest/largest QP for linear regression. The calculation of test data features for prediction is necessary in models I, II, III and our model. As mentioned, models IV and V need to split the data into blocks before feeding them into the neural networks, thus requiring more time than the other models for prediction.

As demonstrated by results, model III and our proposed model, which use the SVR method, have the lowest time cost. Models I and II are slow due to the pre-encoding process. As for models IV and V, when the original data are large, the number of blocks becomes large, leading to much time consumption on splitting and prediction.

5 Conclusions

A learning-based quantization parameter prediction scheme for quality control in 3D medical imaging data compression is proposed in this paper. The optimal QP can be predicted from image features using the SVR model. By assigning QP values to the sequences appropriately, the resultant PSNR meets the targeted PSNR. Experimental results show that our proposed method outperforms existing solutions in precision, achieving good results in time consumption. Thus, our approach has much potential in solving the quality control problem of medical images in real applications such as telemedicine or medical imaging database regulation. In the future we will consider the adaptive QP scheme, rather than the current fixed QP scheme, to develop a better model.

Contributors

Zhong REN and Wei CHEN designed the research. Yuxuan HOU implemented the algorithm and drafted the paper. Yubo TAO participated in the technical discussions. Zhong REN, Yubo TAO, and Wei CHEN revised and finalized the paper.

Compliance with ethics guidelines

Yuxuan HOU, Zhong REN, Yubo TAO, and Wei CHEN declare that they have no conflict of interest.

References

Clark K, Vendt B, Smith K, et al., 2013. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Dig Imag*, 26(6):1045-1057. https://doi.org/10.1007/s10278-013-9622-7

Dinh KQ, Lee J, Kim J, et al., 2018. Only-reference video quality assessment for video coding using convolutional neural network. Proc 25th IEEE Int Conf on Image Processing, p.2496-2500. https://doi.org/10.1109/ICIP.2018.8451262

Table 6 The inference time of the models (per query)

Model	Inference time (ms)			
	Encoding/Decoding	Feature extraction	Prediction	Total
I	16 526	–	6	16 532
II	16 526	2120	2	18 648
III	–	2165	2	2167
IV	–	9506*	7539	17 045
V	–	4720*	15 566	20 286
Ours	–	3339	2	3341

* For data splitting

El-Naqa I, Yang YY, Galatsanos NP, et al., 2004. A similarity learning approach to content-based image retrieval: application to digital mammography. *IEEE Trans Med Imag*, 23(10):1233-1244. https://doi.org/10.1109/TMI.2004.834601

Haralick RM, Shanmugam K, Dinstein IH, 1973. Textural features for image classification. *IEEE Trans Syst Man Cybern*, 3(6):610-621. https://doi.org/10.1109/TSMC.1973.4309314

Huynh-Thu Q, Ghanbari M, 2008. Scope of validity of PSNR in image/video quality assessment. *Electron Lett*, 44(13): 800-801. https://doi.org/10.1049/el:20080522

Kamaci N, Altunbasak Y, Mersereau RM, 2005. Frame bit allocation for the H.264/AVC video coder via Cauchy-density-based rate and distortion models. *IEEE Trans Circ Syst Video Technol*, 15(8):994-1006. https://doi.org/10.1109/TCSVT.2005.852400

Kwon DK, Shen MY, Kuo CCJ, 2007. Rate control for H.264 video with enhanced rate and distortion models. *IEEE Trans Circ Syst Video Technol*, 17(5):517-529. https://doi.org/10.1109/TCSVT.2007.894053

Lazzerini B, Marcelloni F, Vecchio M, 2010. A multi-objective evolutionary approach to image quality/compression trade-off in JPEG baseline algorithm. *Appl Soft Comput*, 10(2):548-561. https://doi.org/10.1016/j.asoc.2009.08.024

Liu F, Hernandez-Cabronero M, Sanchez V, et al., 2017. The current role of image compression standards in medical imaging. *Information*, 8(4):131. https://doi.org/10.3390/info8040131

Ma S, Gao W, Lu Y, 2005. Rate-distortion analysis for H.264/AVC video coding and its application to rate control. *IEEE Trans Circ Syst Video Technol*, 15(12):1533-1544. https://doi.org/10.1109/TCSVT.2005.857300

Ma SW, Si JJ, Wang SS, 2012. A study on the rate distortion modeling for high efficiency video coding. Proc 19th IEEE Int Conf on Image Processing, p.181-184. https://doi.org/10.1109/ICIP.2012.6466825

Miaou SG, Chen ST, 2004. Automatic quality control for wavelet-based compression of volumetric medical images using distortion-constrained adaptive vector quantization. *IEEE Trans Med Imag*, 23(11):1417-1429. https://doi.org/10.1109/TMI.2004.835312

Pan X, Chen ZZ, 2016. Multi-layer quantization control for quality-constrained H.265/HEVC. *IEEE Trans Image Process*, 26(7):3437-3448. https://doi.org/10.1109/TIP.2016.2627818

Patait A, Young E, 2016. High performance video encoding with NVIDIA GPUs. GPU Technology Conf. https://goo.gl/Bdjdgm

Pratt WK, Kane J, Andrews HC, 1969. Hadamard transform image coding. *Proc IEEE*, 57(1):58-68. https://doi.org/10.1109/PROC.1969.6869

Said A, Pearlman WA, 1996. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans Circ Syst Video Technol*, 6(3):243-250. https://doi.org/10.1109/76.499834

Sanchez V, Bartrina-Rapesta J, 2014. Lossless compression of medical images based on HEVC intra coding. IEEE Int Conf on Acoustics, Speech and Signal Processing, p.6622-6626. https://doi.org/10.1109/ICASSP.2014.6854881

Santamaria M, Izquierdo E, Blasi S, et al., 2018. Estimation of rate control parameters for video coding using CNN. IEEE Visual Communications and Image Processing, p.1-4. https://doi.org/10.1109/VCIP.2018.8698721

Schölkopf B, Smola AJ, Williamson RC, et al., 2000. New support vector algorithms. *Neur Comput*, 12(5):1207-1245. https://doi.org/10.1162/089976600300015565

Wang HL, Kwong S, 2008. Rate-distortion optimization of rate control for H.264 with adaptive initial quantization parameter determination. *IEEE Trans Circ Syst Video Technol*, 18(1):140-144. https://doi.org/10.1109/TCSVT.2007.913757

Wang SJ, Summers RM, 2012. Machine learning and radiology. *Med Image Anal*, 16(5):933-951. https://doi.org/10.1016/j.media.2012.02.005

Wu CY, Su PC, 2013. A content-adaptive distortion-quantization model for H.264/AVC and its applications. *IEEE Trans Circ Syst Video Technol*, 24(1):113-126. https://doi.org/10.1109/TCSVT.2013.2273656