Group Perceptual Quality Optimization for Multi-Channel Image Encoding Systems Based on Adaptive Hyper Networks

NAN CHEN1, XIAOYU MA2,3, YONG CHEN4, AND DINGGUO YU3,5

1College of Qianjiang, Hangzhou Normal University, Hangzhou 310018, China
2Institute of Intelligent Media, Communication University of Zhejiang, Hangzhou 310042, China
3Key Lab of Film and TV Media Technology of Zhejiang Province, Communication University of Zhejiang, Hangzhou 310042, China
4Hangzhou Arcvideo Technology Company Ltd., Hangzhou 311012, China
5College of Media Engineering, Communication University of Zhejiang, Hangzhou 310042, China

Corresponding author: Dingguo Yu (yudg@cuz.edu.cn)

This work was supported in part by the Key Research and Development Program of Zhejiang Province, China, under Grant 2019C01002.

ABSTRACT Images and short videos that produced by social networks surge in recent years. Image/Video encoders, such as JPEG and H.264, are indispensably involved to reduce the transmitting bandwidth. However, based on our observation, the encoding parameters and their candidates are often preset to fixed values (or fixed candidate values) in real-world scenarios, which might not be the optimal bandwidth allocation strategy. Considering that, we propose an efficient group quality optimization (GQO) framework for multi-channel image/video encoding systems in which the encoding parameters are configured in a perceptual-quality-driven manner. The GQO framework employs adaptive hyper network to predict the relationships between encoding parameters, transmitting resources, and perceptual qualities, i.e., just taking the pristine image as input, the adaptive hyper network could accurately yield a global overview of perceptual quality and transmitting resource varied along encoding parameters. A step-by-step optimization procedure is then employed to search the optimal encoding parameter for each channel so that overall perceptual quality could be maximized under limited transmitting resource. Experimental results demonstrate the proposed GQO framework could achieve higher perceptual quality whilst maintain the same bandwidth compared to traditional allocation strategies where encoding parameters are preset.

INDEX TERMS Adaptive hyper network, group quality optimization, image quality assessment.

I. INTRODUCTION

Recent years has witnessed the significant growth of social networks which produces massive amounts of images and short videos. Those digital images and videos could get contaminated in any stage of their lifecycle such as acquisition, compression, transmission. Consequently, research of Image Quality Assessment (IQA) is in great need to automatically evaluate the perceptual quality of images, based on which the Quality of Service (QoS) of video transmission systems could be monitored, guaranteed, and optimized.

Despite increasing interest has been drawn in exploiting deep convolutional neural networks (DCNN) for designing no-reference/full-reference image quality assessment (NR/FR-IQA) and notable success has achieved [1], there has largely been a disconnection between research that aimed at evaluating perceptual quality issues and research that aimed at other downstream image processing tasks [2,3]. For example, in the field of IQA, researchers always focused on how to precisely calculate the perceptual quality of images solely and somehow neglect to optimize the perceptual quality of transmitting systems in real-world based on their promising work.

The disconnection between IQA filed and other downstream image processing tasks makes it difficult to incorporate those sophisticated IQA models into other tasks which might yield exciting results with the help of accurate quality control. Specifically, given a situation that need to compress N images or N videos and transmit them from locale X to locale Y via an approximately loss-less network with limited bandwidth B, the IQA models is capable of monitoring the perceptual quality of each image/video stream via applying FR-IQA evaluation in locale X or applying
NR-IQA evaluation in locale $Y$, but that is far from satisfactory — the IQA model is capable of measuring the current quality of each image/video stream whilst fail to figure out how to re-allocate the bandwidth resources to achieve better overall perceptual quality.

Therefore, our work analyzes how to incorporate IQA models into multi-channel image encoding systems to re-allocate the bandwidth resources for each channel so that the overall perceptual quality of the output could be optimized. Such research is of importance because existing popular social networks often contain multiple images rather than single image in a webpage, video content providers also have to compress multiple video streams simultaneously for users. Moreover, encoding parameters such as $QP$ value for JPEG image encoders and bitrate/QP for video encoders could not precisely represent the perceptual quality, therefore incorporating IQA models as quality criteria and exploit a reasonable bandwidth re-allocating strategy based on such IQA quality score would benefit both the servers (saving transmitting resources) and clients (improving perceptual qualities).

It is difficult to yield bandwidth re-allocating strategy just given the current output of the encoders, that is mainly because by feeding current output into IQA model, we could only get the perceptual quality of current parameter settings and know nothing about which direction should the encoder parameters move towards by changing current encoding parameters to another.

Considering the obstacle mentioned above and inspired by the ratio-distortion curve used for testing the performance of video encoders, our work tries to get a global view of relationship between bandwidth and perceptual quality for each image channel by compressing the source into different versions using various encoding parameters and calculating the IQA prediction score for each of the distorted version (shown as Fig. 1), e.g., we could use various encoding parameters (such as $QP$), denoting as $P = [p_1, p_2, \ldots, p_M]$, to compress the same source video and get corresponding $M$ versions of distorted video, video quality assessment (VQA) model (ST-MAD [4], MOVIE [5], etc.) is then employed to predict the perceptual quality for each distorted version and we could get the quality distribution (denotes as $O$) under various $P$, i.e., $O = [o_1, o_2, \ldots, o_M]$, the transmitting resource (file size or bitrate) of each version is denoted as $B = [b_1, b_2, \ldots, b_M]$. Given the quality distribution $O$ and transmitting resource distribution $B$, it is feasible to figure out the best bandwidth allocating strategy by optimizing:

$$\arg\max_{[k_i|i=1,\ldots,M]} \sum_{i=1}^{N} o_{k_i}^{(i)}.$$ (1)

$$\text{under} \sum_{i=1}^{N} b_{k_i}^{(i)} < B_L.$$ (2)

where $N$ denotes the number of image channels, $M$ denotes the number of selectable encoding parameters, $B_L$ denotes the transmitting resource limitations, $o_{k_i}^{(i)}$ and $b_{k_i}^{(i)}$ denotes the IQA prediction score and transmitting resources of the $i$-th image channel under the $j$-th encoding parameter respectively, $k_i$ denotes the index of selected encoding parameters of the $i$-th channel.

In this work, the coupled quality distribution $O$ and transmitting resource $B$ is called IQA incorporated ratio-distortion curve and abbreviated as IQA-curve. The overall quality of multi-channel image encoding system could get optimized by such IQA-curve according to Eq.1 and Eq.2. However, calculating the ratio-distortion curve for each channel is time-consuming, the image has to be compressed for $M \times N$ times and the IQA models also have to be called for $M \times N$ times.

In order to save time and calculating resources, this work tries to predict the IQA-curve by resorting to deep learning techniques, i.e., we could design a DCNN that takes the original image as input and predict the IQA-curve directly without compressing the original image and evaluate the distorted image by IQA for multiple times.

The main contribution of our work is summarized as follow:

1. We propose an effective group quality optimizing framework for multi-channel image encoding systems by re-allocating the transmitting resources of each channel via IQA-curves.
2. We design an adaptive DCNN that could predict the IQA-curves in one-shot, which significantly saves the calculating times and resources.
3. We implement our GQO framework for multi-channel JPEG image encoding systems and achieve satisfactory results, indicating the proposed work is also promising for multi-channel image and video encoding systems.

It also should be noticed that the proposed GQO framework is quite different from prevalent dynamic adaptive streaming techniques. Firstly, existing adaptive streaming techniques such as HTTP-DASH[6] are designed to fit various network environments, e.g., when the transmitting bandwidth are limited, the HTTP-DASH could switch to lower bit-rate to guarantee the fluence of the playback, whilst our proposed GQO framework are designed for the multi-channel encoding system to make full use of the limited storage (or transmitting bandwidth) and to yield optimal overall perceptual quality, i.e., the HTTP-DASH are driven by the transmitting resources.
environments but our GQO framework are driven by perceptual quality. Secondly, although HTTP-DASH could switch different bit-rates according to the network environment of the clients, the candidate bit-rates are still preset and fixed, that means if the candidate bit-rates are preset to 100kbps, 500kbps, and 2000kbps, the HTTP-DASH server could only switch between those three candidate bit-rates and cannot jump to other bit-rates, in contrary, the GQO framework could analysis the IQA-curve of each input to pursue the most suitable encoding parameters.

The rest part is organized as follow: Section II introduces several related IQA works; Section III illustrates our proposed Group Quality Optimized framework and its specific version for multi-channel JPEG image encoding system; Section IV shows the experimental results; and Section V is conclusion.

II. RELATED WORKS

With the help of Deep Neural Networks, numerous fascinating BIQA models are proposed in recent years. For example, Kang et al. [7][8] proposed a multi-task shallow CNN to learn both the distortion type and the quality score; Kim and Lee [9] applied state-of-the-art FR-IQA methods to provide proxy quality scores for each image patch as the ground truth label in the pre-training stage, and the proposed network was fine-tuned by the Subjective annotations. Similarly, Pan et al. [10] employed the U-Net to learn the local quality predicting scores previously calculated by Full-Reference IQA methods, several Dense layers were then incorporated to pool the local quality predicting scores into an overall perceptual quality score; Liang et al. [11] tried to utilize similar scene as reference to provide more prior information for the IQA model; Liu et al. [12] proposed to use RankNet to learn the quality rank information of image pairs in the training set, and then used the output of the second last layer to predict the quality score; Lin and Wang [13] tried to learn the corresponding unknown reference image from the distorted one by resorting the Generative Adversarial Networks, and to assess the perceptual quality by comparing the hallucinated reference image and the distorted image; Chiu et al. [1] proposed a new IQA framework and corresponding dataset that links the IQA issue to two practical vision tasks which are image captioning and visual question answering respectively; Su et al. [14] employed self-adaptive hyper network whose parameters could adjust according to image contents; Zhu et al. [15] leveraged meta-learning to learn a general-purpose BIQA model from training set of several specific distortion types.

As described above, recent state-of-the-art BIQA methods focus on predicting the distorted image solely, and take less consideration on how to make their proposed models incorporated into other downstream vision tasks. Amongst the IQA works above, only [1] tries to link IQA issues with other image vision tasks. Our work exploits how to leverage IQA models to optimize the perceptual quality of multi-channel transmitting systems, which is of much importance because such work could not only optimize the transmitting system both for servers and clients but also represents a beneficial attempt for linking the gap between IQA and other image vision tasks.

III. PROPOSED METHODS

In this section, we detail our group quality optimizing (GQO) framework based on learning IQA-curves. The proposed GQO framework is capable of optimizing the overall perceptual quality for multi-channel image transmitting systems.

We implement our GQO framework for multi-channel JPEG image encoding systems to verify its superiority, therefore the encoding parameters is the commonly used $Q$ value for JPEG compression.

It should be noticed that the proposed GQO framework is also supposed to be applied in video multi-channel systems with limited adjustment of the DCNN, but the implementation of this work focus on JPEG image encoding system for the consistency of narrative and convenience of implementation.

A. PROBLEM FORMULATION

For a multi-channel JPEG image encoding system, we try to optimize the overall perceptual quality given limited transmitting resources, i.e.,

$$
\max \sum_{i=1}^{N} o_i. \quad (3)
$$

$$
s.t. \sum_{i=1}^{N} b_i < B_L. \quad (4)
$$

where $N$ denotes the numbers of channels in the encoding systems meaning that $N$ images are compressed at one time, $B_L$ denotes the limitation of the transmitting resources (or storage), $o_i$ and $b_i$ denotes the perceptual quality and transmitting resources of the $i$-th channel.

As described earlier in Section I, just calculating the IQA predicting score for current multi-channel output is not enough to optimize the overall perceptual quality. That is mainly because we still cannot know how the overall perceptual quality would change if the transmitting resources is re-allocated.

Therefore, we leverage the IQA-incorporated ratio-distortion curve (abbr., IQA-curve) to take a global view of the relationship between bandwidth and perceptual quality. The IQA-curve is calculated as shown in Fig. 1. Specifically, supposing the $Q$ value candidates range from 2 to 100 with step 2 (i.e., encoding parameter $P = \{p_1, p_2, \ldots, p_M\}$, and $M = 50, p_1 = 2, p_2 = 4, \ldots$), then the IQA-incorporated curve of $i$-th channel is comprised of quality distribution $O^i = [o_1^i, o_2^i, \ldots, o_M^i]$ and transmitting resource distribution $B^i = [b_1^i, b_2^i, \ldots, b_M^i]$, where $o_i^j$ is obtained by feeding the $f$-th distorted version (compressed under $j$-th Q value) and the corresponding $i$-th original image into FR-IQA model, and
Based on the IQA-curves of each channel, the optimized overall performance could be derived by searching an index vector \( K = [k_1, k_2, \ldots, k_N] \) (denotes the selections of QP value for each channel) which obeys that

\[
K^* = \arg\max_{k_i \in [1, M]} \sum_{i=1}^{N} o_{k_i}^j, \quad (5)
\]

\[
s.t. \sum_{i=1}^{N} b_{k_i}^j < B_L. \quad (6)
\]

In order to accelerate the time-consuming calculating procedure of IQA-curve for each channel, a self-adaptive hyper network is designed inspired by [14], which could precisely predict the IQA-curve in one-shot, and get rid of \( M \times N \) times of JPEG compression and IQA prediction.

The overall procedure of our proposed group quality optimized framework for multi-channel JPEG image encoding system is shown as Fig.2, which is mainly composed of two steps, i.e., a) using self-adaptive hyper network to predict the IQA-curve for each channel, and b) optimizing the overall perceptual quality according to the predicted IQA-curves and at last the encoding system would operate based on the optimized encoding parameters.

![FIGURE 2. Overall procedure of group quality optimized framework for multi-channel JPEG image encoding systems.](image)

**B. HYPER NETWORK FOR PREDICTING IQA-CURVE**

Inspired by [14], a self-adaptive hyper network is designed to accurately predict the IQA-curve, and its pipeline is illustrated as Fig.3. Multi-scale features extracted by ResNet-50 are employed to regress the IQA-curves. Rather than directly fed the multi-scale features into traditional regression networks whose parameters are fixed for all images, our work employs a content understanding hyper network to generate customized regression parameters for each test image. The reasons we employ the content understanding hyper network are summarized as follow: Firstly, such model has been applied for blind predict IQA score in [14] and achieved satisfactory performance, the predicting of IQA score and IQA-curve are related; Secondly, different image contents yield discriminative patterns of IQA-curves, e.g., the first-order derivative of IQA-curve obtained from different source images exhibit quite unsimilar patterns; Thirdly, based on our experience, using fixed parameters for all test images to predict a multi-dimension curve often leads to overfitting problem, which could be avoided by generating customized parameters from the hyper networks.

As described in Fig.3, the overall pipeline of our IQA-curve prediction model is similar with [14], but it is comprised of two parallel dataflows, the upper one is employed for predicting the transmitting resource distribution whilst the lower one is employed to for predicting the perceptual quality distribution. The overall structures of the two data streams are similar, so the illustration is focused on the lower one that employed for predicting the perceptual quality distribution.

Specifically, given an input image denoted as \( x \), its semantic features are extracted by ResNet-50. The semantic features extracted from the last convolutional layers, denoting as \( S(x) \), are employed as the input of the hyper network whilst the multi-scale semantic features extracted from each conv-blocks, denoting as \( S_{mx}^O(x) \), are employed as the input of target network. The parameters of the target network, denoting as \( \theta^O \), is generated from the hyper network:

\[
\theta^O = H^O(S(x), \gamma^O). \quad (7)
\]

where \( H^O \) denotes the mapping function of hyper network and \( \gamma^O \) represents hyper network parameters. Denoting the target network as \( \phi^O \), the prediction of the quality distribution \( O^p \) is then can be described as:

\[
O^p = \phi^O(S_{mx}^O(x), H^O(S(x), \gamma^O)). \quad (8)
\]

Based on similar structures, the distribution of transmitting resources \( B^p \) could also be predicted, the predicted IQA-curve could be drawn given \( B^p \) and \( O^p \).

We make several modifications to ensure the self-adaptive model are fitted to our problem. The modifications are summarized as follow:

1. Rather than using \( L1 \) norm to measure the difference between the ground truth and prediction, our work adds the KL-divergence as another cost item to ensure the overall distributions between the prediction and ground-truth are similar, e.g., when predicting the quality distribution \( O^p \), the loss is defined as:

\[
l = \frac{1}{N} \sum_{i=1}^{N} (\lambda_1 \left\| O^p_{(i)} - O^{gr}_{(i)} \right\|_1 + \lambda_2 KL(O^p_{(i)} - O^{gr}_{(i)})). \quad (9)
\]

2. The model proposed in [14] limit the input size as \( 224 \times 224 \), resulting into massive loss of local features when original input size is HD \( (1920 \times 1080) \), considering that, the input size is modified to \( 640 \times 384 \), the input HD image is firstly resize into \( 640 \times 360 \) then padded its top and bottom to meet \( 640 \times 384 \).

3. Based on our observation, when predicting the IQA-curves, shallower layers conveys more detailed feature infor-
FIGURE 3. An illustration of our proposed heterogenous reference BIQA framework.

C. IQA CRITERIA SELECTION

There is still a vital thing need to be figure out, i.e., which IQA criteria should be selected to depict the quality distribution. Since original image is available in the compression system, full-reference (FR) IQA is therefore preferred.

There are several prevalent FR-IQA metrics, e.g., SSIM [16], VIF [17], GMSD [18], FSIM [19], VSI [20], and MAD [21], almost all of them could achieve satisfactory accuracy in predicting single distortion type (e.g., in TID2013 image database, the Pearson Linear Correlation Coefficients of nearly all above FR-IQA metrics with subjective score could obtain 0.9 towards JPEG compressed images solely). However, we would like to search for a more effective FR-IQA method that could predict the subjective score as accurate as possible. Based on [22] and [23], the nonlinear combination of multiple FR-IQA metrics would achieve better performance by resort of machine learning techniques, at the same time, [24] points that such nonlinear combination might lead to a generalization ability decline of the combined FR-IQA model.

In order to improve the accuracy of FR-IQA via nonlinear combination whilst maintain the generalization ability, this work propose a pre-training engaged multi-method-fusion FR-IQA model (PTMMF-FRIQA) and firstly pre-train it by prior quality knowledge to pursue better generalization ability. Specifically, 1000 pristine images collected from internet are compressed by JPEG using 3 different Q values, which are 10, 35, 80 and representing ‘Poor’, ‘Acceptable’, ‘Satisfactory’ respectively. A classification network is then pre-trained by supervising it to successfully classify inputs (combination of FR-IQA scores) into correct quality labels (‘Poor’, ‘Acceptable’, ‘Satisfactory’), which is shown as Fig.4. After that, the last layer is substitute by a fully connect layer whose output size is 1 and activated by sigmoid that more suitable for regression task. At last, training samples with subjective score are employed to fine-tune the regression network to accurately predict the perceptual quality. Ablation Experiments demonstrates such pre-train and fine-tune pipeline would yield satisfactory multi-method fusion FR-IQA metrics.
D. GROUP QUALITY OPTIMIZATION PROCEDURE

The group quality optimization procedure could be conducted based on the prediction of IQA-curve. Above implementations sampled 50 encoding parameters (Q value) ranging from 2 to 100 and denoted as \( P = \{ p_1, p_2, \ldots, p_M \} \) where \( M = 50, p_1 = 2, p_2 = 4 \) and so on. The predicted IQA-curve therefore contains two 50-dimensional vectors which representing the quality distribution transmitting resource distribution respectively.

Supposing \( N \) images are engaged in the GQO, the predicted quality distribution and transmitting resource distribution of the \( i \)-th image is denoted as \( O^i = [o^i_1, o^i_2, \ldots, o^i_M] \) and \( B^i = [b^i_1, b^i_2, \ldots, b^i_M] \) respectively. Given a bandwidth limitation \( B_L \), the GQO procedure is described as follow:

1. The selection of encoding parameters for \( N \) channels is denoted as \( K = [k_1, k_2, \ldots, k_N] \) and initialized with \( K^{(0)} = 1^M \).

2. For the \((t+1)\)-th step, calculate the gradient matrix \( G \in \mathcal{R}^{M \times N} \) by \( G_{i,j} = [o^i_k - o^i_{k+j}] = [b^i_k - b^i_{k+j}] - 0 < j < (M - k^i_j) \) and other elements of \( G \) is set to -1.

3. The selection of encoding parameters is then updated by \( K^{(t+1)} = K^{(t)} + j^* \times \delta(x^*) \), where \( i^* \) and \( j^* \) are selected by \(< i^*, j^*> = \arg\max_{i,j} G(i,j) \), and \( \delta(x) \) denotes a \( N \)-dimensional vector whose elements are all zeros except the \( x \)-th element is 1.

4. If \( \sum_{j=1}^M b^i_{k^i_j+1} > B_L \), the update stops and \( K^{(t)} \) is returned as the optimal parameter settings, i.e., the \( i \)-th channel uses the \( k^i_{j^*} \)-th candidate encoding parameter to conduct image encoding; otherwise the update procedure would jump to step (2) and continue.

IV. EXPERIMENTS

The involved training and testing images, performance of each component and the overall framework are illustrated in this section.

A. TRAINING SET CONSTRUCTION

Both the proposed adaptive hyper network in Section III-B and the multi-method fusion FR-IQA model in Section III-C need large amounts of training samples. We therefore collected up to 1000 high-resolution images from digital cameras and online image resources. The collected images are cropped and rescaled into \( 1920 \times 1080 \) while their original aspect ratio is maintained. Snapshot of part of the involved training samples are shown as Fig.5, which demonstrates that the established training set contains various content types ranging from ceremonies, sceneries, people, sculptures, animals, buildings, and so on.

B. PERFORMANCE EVALUATION FOR PTMMF-FRIQA

As described in Section III-C, the PTMMF-FRIQA is employed to accurately annotate the perceptual quality of JPEG distorted images. A classification is firstly pre-trained by the established training set containing up to 1000 images, and after substituting the last layer, the changed network is fine-tuned by the IQA quality database.
The inputs of the networks are 8 traditional FR-IQA scores including PSNR, SSIM, VIF, GMSD, FSIM, FSIMC, MAD, VSI. The baseline of the classification network is shallow BP networks with size 8-56-28-6-3, as shown in Fig. 4. Images in the training set is firstly compressed by JPEG with Q value 10, 35, 80 representing ‘Poor Quality’, ‘Acceptable Quality’, ‘Satisfactory Quality’ respectively. The classification network is then pre-trained to predict the quality labels of the 3000 distorted images.

After the pre-training procedure, the last layer is substituted by a 6-1 Dense layer activated by sigmoid, this new network is then fine-tuned by training samples in IQA quality database. We employ 5-fold cross validation to demonstrate the effectiveness of the proposed multi-method fusion FR-IQA model, i.e., for each IQA database, 80% of the JPEG distorted images are employed as training set and the others as testing set, such procedure repeats 5 times to guarantee all images are involved in testing set. The predicting score in testing sets are shown in Fig. 6 and Table 2.

In order to further validate the multi-method fusion FR-IQA model, cross validation is conducted in which JPEG distorted images in LIVE is employed as training samples whilst JPEG distorted images in TID2013 is employed as testing samples. The testing results is shown as Fig.7.

It should be noticed that above results demonstrate the employed lightweight PTMMF-FRIQA could accurately predict the perceptual quality of JPEG distorted images and overwhelm others. That is mainly because PTMMF is trained specially for JPEG distortion whilst other metrics are towards a more general purpose. However, considering the proposed group quality optimization framework is for server-side, in which the distortion types are fixed aforehand (i.e., JPEG in this implementation), The proposed PTMMF is therefore the most suitable IQA metric for our GQO framework.

### C. PERFORMANCE EVALUATION FOR ADAPTIVE HYPER NETWORK

The adaptive hyper network takes pristine images as input and yield its IQA-curve by predicting the quality distribution and transmitting resource distribution respectively.

Therefore, we construct the training sample as follow: each image in the training set is compressed by JPEG compression with Q values ranging from 2 to 100 (the interval is 2). The perceptual quality of each distorted image is predicted by the multi-method fusion FR-IQA model, and its file size is also recorded, let $o_i^j$ and $b_i^j$ denote the perceptual quality and file size of the $j$-th distorted version of the $i$-th image, the ground-truth quality distribution and transmitting resource distribution of the $i$-th image is then denoted as $O_i = [o_i^1, o_i^2, \ldots, o_i^{50}]$ and $B_i = [b_i^1, b_i^2, \ldots, b_i^{50}]$ respectively.

The adaptive hyper network is trained on one Tesla V100 GPU card with 32G memory installed on Dell Quarterly Server Tracker, the mini-batch size is set to 16 and other settings are similar with [14].

The 5-fold cross validation is also employed to verify the effectiveness of the IQA-curve prediction accuracy of the adaptive hyper network. The experimental results are shown in Fig.8. It should be noticed the file size and perceptual quality are normalized.

As shown in Fig.8, the adaptive hyper network could accurately predict the IQA-curve given solely one pristine image without cumbersome JPEG compression and PTMFF calculation. In addition, time costs of the proposed adaptive hyper network are also tested, in average, the time that predicting IQA-curve via adaptive hyper network for per image takes only 0.037 seconds (tested on Intel i7-8700@3.2GHz with GTX2080 by pytorch), whilst generating IQA-curve via multiple times of JPEG compression and PTMFF calculation costs up to 76.325 seconds ($M = 50$, tested on similar host PC by MATLAB, with GPU acceleration on). Therefore, predicting the IQA-curve via DCNN significantly saves the time-consuming.


**D. PERFORMANCE ANALYSIS OF GROUP QUALITY OPTIMIZATION FRAMEWORK**

Experimental results in section IV-B and IV-C indicate each component of our GQO framework is sensible and effective. The overall performance of the GQO framework is evaluated in this section.

The proposed Group Quality Optimization framework (shown as Fig.2) aims to allocates the transmitting resources for each encoding channel in a perceptual-quality-driven manner, i.e., by predicting the IQA-Curve of each channel, the GQO framework could intelligently yield customized encoding parameters for each channel so that the overall quality of the output achieves optima under limited bandwidth.

In opposite, existing multi-channel encoding framework neglects the diversity of IQA-curve between various image contents and tends to compress the images via fixed (or pre-set) encoding parameters ($Q$ value or output file-size).

In order to demonstrate the superiority of the proposed GQO framework, two prevalent allocating strategies for multi-channel image encoding systems are involved as the comparison, i.e., the ‘Q-fixed’ mode and the ‘B-fixed’ mode. In the ‘Q-fixed’ mode, all inputs of the multi-channel encoding system are compressed via the same $Q$ values, and in the ‘B-fixed’ mode, all inputs are compressed into nearly the same bandwidth (i.e., output file-size).

The evaluation procedure is shown as Fig.9 and described as follow. The encoding channels $N$ are set to 40, i.e., 40 images are encoded simultaneously via the multi-channel encoding system. 100 images with various contents are downloaded from public-available website as the candidate inputs of the multi-channel encoding system. The evaluation procedure repeats for 3 epochs. For each epoch, 40 images are randomly sampled from the 100 downloaded images. As for each evaluation epoch, the 40 involved images are firstly compressed via ‘Q-fixed’ mode. The multi-channel encoding of ‘Q-fixed’ mode repeats 8 periods with ascending $Q$ values, i.e., in the 1st period, the $Q$ value of all channels are set to 10, and the 2nd period, the $Q$ value of all channels are set to 20, etc. The overall output file-size of the ‘Q-fixed’ mode in each period is recorded, which is then employed as the
bandwidth limitation of the GQO framework and the overall output file-size of the ‘B-fixed’ mode.

The ‘Quality Loss’ is introduced to measure the performance of different modes of multi-channel image encoders. Specifically, considering the proposed PTMMF-FRIQA value (ranging from 0 to 1) could accurately assess the perceptual quality of images compressed with JPEG and higher PTMMF-FRIQA value means better perceptual quality, therefore, the output of the i-th channel coupled with its original input is fed into the PTMMF-FRIQA model yields $o_i$, and the value $\sum_{i=1}^{K}(1-o_i)$ is then employed to measure the quality loss, i.e., higher value means the output of the multi-channel image encoder system suffers more quality degradation. The quality loss of the ‘Q-fixed’ mode, ‘B-fixed’ mode, and GQO mode is denoted as $L_Q^i$, $L_B^i$, and $L_{GQO}^i$, respectively.

Based on the ‘Quality Loss’, we could conduct a more intuitive measurement to illustrate the superiority of the GQO framework, which is called the ‘Quality Gain’. The Quality Gain of the GQO framework compared to ‘Q-fixed’ mode is calculated by $(L_B^i - L_{GQO}^i)/L_{GQO}^i$. Higher value of quality gain means that the proposed GQO framework could preserve more perceptual quality compared to ‘Q-fixed’ or ‘B-fixed’ mode.

Experimental results of the 1st, 2nd, and 3rd epochs are shown in Fig.10, Fig.11, and Fig.12 respectively, demonstrating the proposed GQO framework could significantly reduce the overall quality loss by a range from 1.5 to 49.2% compared to traditional Q-fixed and B-fixed mode, regardless of the image contents.

As shown from Fig.10 to Fig.12, although various image contents could achieve quality gain via GQO framework, but the effects are quite different, e.g., results in Fig.10-(b) is much obvious than that in Fig.12-(b). Further analysis is conducted to investigate what factors leads to such diversity of effects. Intuitively, if inputs of the multi-channel encoding system are all the same, their IQA-curves are also the same, and GQO framework could gain noting. Therefore, we measure the difference of IQA-curves amongst all input channels (denoted as $D$) as follow:

Let $O_i^{(i)}$ and $B_i^{(i)}$ denote the quality distribution and transmitting distribution of the i-th channel, bilinear interpolating

![FIGURE 10. Performance comparison between proposed GQO framework and fixed-Q/fixed-B mode (1st period), (a) involved images of 1st experiment; (b) corresponding experimental results of (a).](image1)

![FIGURE 11. Performance comparison between proposed GQO framework and fixed-Q/fixed-B mode (2nd period), (a) involved images of 1st experiment; (b) corresponding experimental results of (a).](image2)
is firstly employed to obtain the mapping function between $O^{(i)}$ and $b^{(i)}$, denoted as $f^{(i)}$, and $O^{(j)} = f^{(j)}(b^{(j)})$, $i \in [1, M]$. The fixed interval sequence $X = [0.01, 0.02, \ldots, 1]$ is then fed to $f^{(i)}$ and yield $Y^{(i)}, Y^{(j)} = \{y_1^{(i)}, y_2^{(i)}, \ldots, y_{100}^{(i)}\}$. The difference of IQA-curves (denoted as $D$) is then obtained by calculating the standard deviation of $Y$ amongst axis $i$ and then calculating the mean of standard deviation, i.e., supposing $N$ images are involved in the multi-channel image encoding, their difference of IQA-curves ($D$) is calculated by $D = \text{mean}(\text{std}(y_1^{(i)}), \text{std}(y_2^{(i)}), \ldots, \text{std}(y_{100}^{(i)}))$.

The relationship between $D$ and the quality gain in above three experiments are shown as Fig.13., demonstrating that large difference between IQA curves might leads to more obvious quality gain in our GQO framework.

V. CONCLUSION

This work presents an efficient group quality optimization framework for multi-channel Image encoding systems. Specifically, the GQO for multi-channel JPEG image encoding system is implemented and estimated. Without any modification of the native image encoder, the proposed GQO framework could effectively analyze the IQA-curve and improve the overall perceptual quality of output given limited transmitting resources. The time-cost is imperceptible via acceleration by adaptive hyper network. We therefore think the proposed work is capable of deployed into real-world multi-channel encoding systems and benefit a lot. The implementation of GQO for multi-channel H.264/HEVC video encoding system would be our future work.

REFERENCES

[1] T.-Y. Chiu, Y. Zhao, and D. Gurari, “Assessing image quality issues for real-world problems,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2020, pp. 3646–3656.

[2] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, “PCANet: A simple deep learning baseline for image classification,” IEEE Trans. Image Process., vol. 24, no. 12, pp. 5017–5032, Dec. 2015.

[3] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2015, pp. 1116–1124.

[4] P. V. Vu, C. T. Vu, and D. M. Chandler, “A spatiotemporal most-apparent-distortion model for video quality assessment,” in Proc. 18th IEEE Int. Conf. Image Process., Sep. 2011, pp. 2505–2508.

[5] K. Seshadrinathan and A. C. Bovik, “Motion-based perceptual quality assessment of video,” Proc. SPIE, vol. 7240, Feb. 2009, Art. no. 72400X.

[6] L. De Ciccio, V. Califano, V. Palmisano, and S. Mascolo, “ELASTIC: A client-side controller for dynamic adaptive streaming over HTTP (DASH),” in Proc. Int. Packet Video Workshop, Dec. 2013, pp. 1–8.

[7] L. Kang, P. Ye, Y. Li, and D. Doermann, “Convolutional neural networks for no-reference image quality assessment,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 1733–1740.

[8] L. Kang, P. Ye, Y. Li, and D. Doermann, “Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks,” in Proc. Int. Conf. Image Process., Sep. 2015, pp. 2791–2795.

[9] J. Kim and S. Lee, “Fully deep blind image quality predictor,” IEEE J. Sel. Topics Signal Process., vol. 11, no. 1, pp. 206–220, Feb. 2017.

[10] D. Pan, P. Shi, M. Hou, Z. Ying, S. Fu, and Y. Zhang, “Blind predicting similar quality map for image quality assessment,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6373–6382.

[11] Y. Liang, J. Wang, X. Wan, Y. Gong, and N. Zheng, “Image quality assessment using similar scene as reference,” in Proc. European Conf. Comput. Vis. Cham, Switzerland: Springer, 2016, pp. 3–18.

[12] X. Liu, J. Van De Weijer, and A. D. Bagdanov, “RankIQA: Learning from rankings for no-reference image quality assessment,” in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 1040–1049.

[13] K.-Y. Lin and G. Wang, “Hallucinated-IQA: No-reference image quality assessment via adversarial learning,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 732–741.
[14] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, “Blindly assess image quality in the wild guided by a self-adaptive hyper network,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2020, pp. 3667–3676.

[15] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, “MetaQA: Deep meta-learning for no-reference image quality assessment,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2020, pp. 14143–14152.

[16] 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.

[17] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” IEEE Trans. Image Process., vol. 15, no. 2, pp. 430–444, Feb. 2006.

[18] X. Mou, L. Zhang, X. Mou, and A. C. Bovik, “Gradient magnitude similarity deviation: A highly efficient perceptual image quality index,” IEEE Trans. Image Process., vol. 23, no. 2, pp. 684–695, Feb. 2014.

[19] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: A feature similarity index for image quality assessment,” IEEE Trans. Image Process., vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[20] L. Zhang, Y. Shen, and H. Li, “VSI: A visual saliency-induced index for perceptual image quality assessment,” IEEE Trans. Image Process., vol. 23, no. 10, pp. 4270–4281, Oct. 2014.

[21] E. C. Larson and D. M. Chandler, “Most apparent distortion: Full-reference image quality assessment and the role of strategy,” J. Electron. Imag., vol. 19, no. 1, pp. 011006-1–011006-21, Jan. 2010.

[22] T.-J. Liu, W. Lin, and C.-C.-J. Kuo, “Image quality assessment using multi-method fusion,” IEEE Trans. Image Process., vol. 22, no. 5, pp. 1798–1807, May 2013.

[23] Oszust, M. Oszust, “Decision fusion for image quality assessment using an optimization approach,” IEEE Signal Process. Lett., vol. 23, no. 1, pp. 65–69, Jan. 2016.

[24] X. Ma, X. Jiang, and D. Pan, “Performance validation and analysis for multi-method fusion based image quality metrics in a new image database,” China Commun., vol. 16, no. 8, pp. 147–161, Aug. 2019.

[25] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” IEEE Trans. Image Process., vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[26] N. Ponomarenko, O. Ieremeiev, V. Lukin, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, “Color image database TID2013: Peculiarities and preliminary results,” in Proc. European Workshop Vis. Inf. Process., 2013, pp. 106–111.

NAN CHEN received the master’s degree from the Jiangsu University of Science and Technology, in 2005. She is currently a Lecturer with Hangzhou Normal University. Her research interests include intelligent processing of video and digital image, optimization of generalization abilities for deep neural networks towards IQA tasks, content-aware video compression, and digital signal processing.

XIAOYU MA received the Ph.D. degree in computer science from the Communication University of China, Beijing, China, in 2019. He is currently a Research Assistant with the Institute of Intelligent Media Technology, Communication University of Zhejiang, China. He has published seven articles and filed one patent. His current research interests include image quality assessment, video compression, and medical image processing.

YONG CHEN is currently pursuing the M.B.A. degree with Zhejiang University. He has worked in the field of image and video encoding for more than 20 years. He is also the Vice President with Hangzhou Arcvideo Technology Company Ltd. He is also a Senior Engineer. He has filed more than 30 patents that are relevant to image and video encoding techniques.

DINGGUO YU was born in 1976. He received the Ph.D. degree in computer application technology from Tongji University, China, in 2011. He is currently a Professor and the Director of Key Laboratory of Film and TV Media Technology, Zhejiang, China. His research interests include media fusion technology, big data, and artificial intelligence for media.

* * *