Neurocomputing 500 (2022) 307–318

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Blind image quality assessment based on progressive multi-task learning ☆

Aobo Li [a], Jinjian Wu [a,*], Shiwei Tian [b,*], Leida Li [a], Weisheng Dong [a], Guangming Shi [a]

[a] School of Artificial Intelligence, Xidian University, Xi'an 710071, China
[b] National Innovation Institute of Defense Technology, Beijing 100000, China

ARTICLE INFO

Article history:
Received 12 October 2021
Revised 16 March 2022
Accepted 14 May 2022
Available online 18 May 2022
Communicated by Zidong Wang

Keywords:
Blind image quality assessment
Multi-task learning
Deep neural network
Progressive relevance

ABSTRACT

Due to the lack of adequate training data and sufficient mining of prior knowledge related to perceived quality, most existing image quality assessment (IQA) methods show limited generalization performance. In this paper, we study the prior knowledge from the factors affecting perceived quality, and introduce a novel progressive multi-task learning based blind IQA method. Inspired by the definition of IQA: human comprehensive perception for degradation of image content, we firstly deconstruct IQA into three elements, i.e., image content, pattern of degradation, and intensity of degradation. Based on these elements, we design the corresponding auxiliary tasks for instructing the network to learn IQA. By statistical analysis on a great deal of data, we find that there is progressive relevance among the four tasks. Furthermore, we mathematically derive that introducing the progressive relevance into a multi-task learning network can effectively constrain the hypothesis space of the main task. Under the guidance of the derivation, we propose an end-to-end IQA framework based on progressive multi-task learning. Experimental results demonstrate the excellent generalization capability of the proposed method, which achieves state-of-the-art performance against these existing IQA methods.

1. Introduction

With the explosive growth of visual data, human society is entering the era of big visual data. Digital images have become an essential medium of information, and their perceptual quality has a direct impact on the user experience. Unfortunately, it is almost inevitably distorted by various distortions from acquisition, compression, processing, transmission to storage in its life cycle. Therefore, to monitor the quality of images and assure the user experience, a reliable image quality assessment (IQA) method is crucially important.

Generally, subjective IQA is considered to be the most reliable and accurate method. Nonetheless, it is not practical since it is laborious and time-consuming. Hence, objective IQA, which can perform automatically without humans, is of great value. According to the dependence of reference information, objective IQA can be categorized as full-reference IQA (FR-IQA) [7,41,23], reduced-reference IQA (RR- IQA) [37,3,11], and blind IQA (BIQA) [27,25,26]. Due to the unavailability of reference information on most occasions, BIQA method is the most widely used IQA method among them.

Most traditional BIQA methods with hand-crafted features are designed on the basis of natural scene statistics (NSS) or human visual system (HVS). However, BIQA is an extremely abstract and complicated task since it is highly correlated with human perception, and its factor cannot be analyzed simply. Therefore, it is tough to design hand-crafted features which represent the quality degradation of images efficiently and comprehensively.

With the extensive application of deep learning in various vision problems [10,30,40], several attempts have been made to apply deep learning technology to the BIQA task. Due to the complexity of BIQA and limited training data, the early methods which simply map images to quality scores by training a convolutional neural network (CNN) fail to perform satisfactorily and exist serious overfitting problems.

To relieve the overfitting problem which results from the scarcity of training data, [29] pre-trains a Siamese Network by using a ranked image database (which is generated without human labeling). However, it focuses on the lack of training data and ignores the complexity of BIQA, so that it doesn't specifically design the network by using the prior knowledge related to BIQA. [19,31]

☆ This work was supported in part by the National Natural Science Foundation of China under Grant 62022063.

* Corresponding authors.
E-mail addresses: jinjian.wu@mail.xidian.edu.cn (J. Wu), tianxwell@163.com (S. Tian).

https://doi.org/10.1016/j.neucom.2022.05.043
0925-2312/

use the idea of multi-task learning (MTL) with setting the auxiliary task of predicting distortion type. Nevertheless, they don't adequately analyze the factors affecting perceived quality and only use prior knowledge from distortion type.

In this paper, we study the prior knowledge from the factors affecting perceived quality, and introduce a novel progressive multi-task learning based BIQA method. From the definition of IQA: human comprehensive perception for degradation of image content, we can find that image quality is related to image content and its degradation, where degradation is manifested in pattern and intensity. Hence, we deconstruct IQA into three elements, i.e., image content, pattern of degradation, and intensity of degradation. According to these elements, three corresponding auxiliary tasks are set up to guide the network for IQA learning, which are content type classification task (*Task C*), distortion type classification task (*Task T*), and distortion intensity classification task (*Task I*). The image content types are artificially defined based on the analysis of the distribution of image content components (texture, edge, and smooth area).

In order to strengthen the guidance of prior knowledge, we further investigate the relationship among the tasks. By statistical analysis of a great deal of data from [44], we find that information about image content implies information about quality and distortion (type and intensity). Meanwhile, distortion intensity is meaningful only if the distortion type is given. And information about distortion contains information about image quality. As we can see, from *Task C*, *Task T*, *Task I* to IQA, tasks are set up from low level to high level, and the lower-level task is instructive to the higher-level task. There is a relationship among these tasks, and we call it progressive relevance. Motivated by the progressive relevance, a progressive multi-task learning (PMTL) strategy derived from hard parameter sharing MTL is proposed to further strengthen the lower-level task's constraint on the higher-level tasks. Furthermore, we mathematically derive that by introducing prior knowledge from auxiliary tasks and progressive relevance, the main task's hypothesis space is constrained effectively. Finally, an IQA framework based on PMTL is presented. Experimental results show the proposed method achieves state-of-the-art performance. And cross-database evaluations demonstrate the excellent generalization capability and robustness of the proposed method.

In summary, the contributions of this work are as follows:

- We deconstruct IQA into three elements and specially design three auxiliary tasks to guide the learning of IQA.
- We investigate the progressive relevance among the tasks. According to the characteristics, a progressive multi-task learning strategy is proposed.
- We prove the effectiveness of progressive multi-task learning to constrain the hypothesis space of the main task. Experimental results demonstrate state-of-the-art performance and strong generalization capability of the proposed method.

2. Related work

2.1. Image quality assessment

Early approaches are often based on hand-crafted features extracted by NSS models. The approaches assume that pristine images have inherent statistical regularity, and distortions will change these regularities. DIIVINE [34] is a two-stage method based on NSS features, which first identifies distortion types of images and applies distortion-specific methods to predict quality scores. BLIINDS-II [38] adopts the features based on discrete cosine transform (DCT) coefficients of images for IQA. BRISQUE [33] utilizes the statistical features of locally normalized luminance coefficients to acquire quality scores. CORNIA [48] uses a codebook-based method to learn local descriptors and support vector regression (SVR) is applied to obtain quality scores. Xu et al. propose a BIQA method [46] by using the statistical differences between codebook and images.

Recently, many deep learning based approaches are proposed, and achieve remarkable progress in BIQA. Kang et al. propose a shallow CNN [18] to map images to quality scores. Then they refine it to a multi-task CNN [19] to simultaneously classify distortion types and estimate quality scores. BIECON [21] uses FR-IQA methods to guide CNN to generate the local quality maps, and then pooled features are regressed onto a subjective score. Liu et al. synthesize a mass of ranked images with different distortion types and levels to train a Siamese network [29] for learning the quality rankings. MEON intrudes a cascaded multi-task framework [31], where a distortion type classification network is first trained, and then a quality regression network starting from the pre-trained early layers and the outputs of the distortion type classification network is trained subsequently. DIQA [22] let the CNN learn to predict the error map, and then to estimate a quality score. Zhang et al. propose DB-CNN [51] for quality prediction, in which two pre-trained features for synthetic distortions and authentic distortions are bilinearly pooled into a quality representation. Zeng et al. propose probabilistic quality representation (PQR) [50] to represent image quality as a probability, instead of as a scalar quantity. Zhang et al. developed a unified BIQA framework [52] that allows differentiable models to be trained on multiple IQA databases simultaneously by combining them via pairwise rankings.

We deconstruct IQA into three elements and separately design auxiliary tasks for assisting learning. And based on the exploration of the relationship among tasks, the PMTL model is designed to further enhance the utilization of prior knowledge, so as to narrow the hypothesis space of the main task and alleviate the overfitting problem.

2.2. Multi-task learning

MTL is a subfield of machine learning where multiple tasks are jointly learned by sharing the models' parameters [4]. By exploiting the information from both task-generic and task-specific, MTL offers advantages like relieving overfitting and improving data efficiency. And its effectiveness has been demonstrated in many fields like facial landmark detection [53], semantic segmentation [5], human action recognition [12], etc. According to how the parameters among tasks are shared, MTL methods can be classified as hard parameter sharing methods [53,5] and soft parameter sharing methods [6,32].

Hard parameter sharing methods are generally applied by sharing the hidden layers while keeping some task-specific output layers. TCDCN [53] simultaneously learns facial attribute inference and head pose estimation, and gets promoted on both tasks. Dai et al. propose a multi-task framework called MNCs [5]. This framework has three stages, each of which addresses one sub-task and the output of each sub-task is as the input of the next sub-task.

For soft parameter sharing methods, there is a separate network for each task. [6] uses l2 distance between the parameters of the network for regularization to encourage them to be similar. Misra et al. propose a Cross-Stitch network [32], where the cross-stitch units are used to let the model choose which tasks to leverage information from by sharing representations as linear combinations.

3. Proposed method

In this section, a novel BIQA framework based on progressive multi-task learning is proposed. The proposed framework consists

of four tasks: 1) content type classification (*Task C*); 2) distortion type classification (*Task T*); 3) distortion intensity classification (*Task I*); 4) image quality score regression (*Task Q*). For the given training data, we jointly train all four tasks. Unlike the paradigm of hard parameter sharing MTL framework, the proposed framework is designed to make full use of the prior knowledge from the progressive relevance and explicitly strengthen the lower-level task's constraint on the higher-level tasks. According to the characteristics of the tasks, each task's features are extracted hierarchically and the lower-level task's features are shared with the higher-level task progressively.

In the following, we will describe the definition of each task, introduce the proposed model's architecture, and derive the effectiveness of PMTL on constraining the main task's hypothesis space.

3.1. Tasks setting

IQA is human comprehensive perception for degradation of image content. From this definition, IQA can be deconstructed into three elements: image content, pattern of degradation, and intensity of degradation. As shown in Fig. 1, all of the three factors have an obvious impact on image perception quality. For example, the image with a higher proportion of texture is more seriously distorted by additive white Gaussian noise, while the image with a higher proportion of smooth region is more seriously distorted by Gaussian blur. Meanwhile, the higher the intensity is, the more seriously it is distorted. Therefore, according to these elements, we set up the corresponding auxiliary tasks (*Task C*, *Task T*, and *Task I*) to guide the network for *Task Q* learning. All tasks are described in detail as follows.

3.1.1. Content type classification

Image content in the pixel level falls into three categories: texture, edge, and smooth region. We artificially define the types of image content in line with the distribution of image content component proportion. Concretely, we use the structure-texture decomposition method [45] to classify the image content, and the proportion of each component is calculated as v_t, v_e and v_s. Then the index f is defined as follows to measure the imbalance of image content component distribution.

$$f = sign(v_s - v_t)(|v_s - v_t| - v_e) \tag{1}$$

where $sign(\cdot)$ is a function that extracts the sign of a real number.

Finally, according to Eq. (2), we divide the image content types c into rich textured image, biased textured image, balanced image, biased smooth image, and rich smooth image [1]. Some examples of content classification are shown in Fig. 1.

$$c = \begin{cases} 0 & f < H_1 \\ 1 & H_1 \leqslant f < H_2 \\ 2 & H_2 \leqslant f < H_3 \\ 3 & H_3 \leqslant f < H_4 \\ 4 & H_4 \leqslant f \end{cases} \tag{2}$$

where 0, 1, 2, 3 and 4 mean rich textured image, biased textured image, balanced image, biased smooth image, and rich smooth image, respectively. H_1, H_2, H_3 and H_4 are thresholds for content types classification. The values of them are respectively set to −0.3, 0, 0.2, and 0.5, which are obtained through experiments.

Task C aims to let the network learn to classify image content types. We use the cross-entropy error as the loss function for *Task C*,

$$L_C = -y_C log(\hat{y}_C) - (1 - y_C) log(1 - \hat{y}_C) \tag{3}$$

where y_C is the ground truth of *Task C*, and $\hat{y}_C$ is the corresponding predicted classification probability.

3.1.2. Distortion type classification

For *Task T*, the goal is to distinguish which distortion type the given image suffered. We use the cross-entropy error as the loss function for *Task T*,

$$L_T = -y_T log(\hat{y}_T) - (1 - y_T) log(1 - \hat{y}_T) \tag{4}$$

where y_T is the ground truth of *Task T*, and $\hat{y}_T$ is the corresponding classification probability predicted.

3.1.3. Distortion intensity classification

Task I is aimed to predict the degree of distortion. Same as before, the cross-entropy error is used as the loss function,

$$L_I = -y_I log(\hat{y}_I) - (1 - y_I) log(1 - \hat{y}_I) \tag{5}$$

where y_I is the ground truth of *Task I*, and $\hat{y}_I$ is the corresponding classification probability.

3.1.4. Image quality score regression

Task Q is to estimate quality scores for the given images. We use the squared-error as the loss function for *Task Q*,

$$L_Q = || y_Q - \hat{y}_Q ||_2^2 \tag{6}$$

where $y_Q, \hat{y}_Q$ are the ground truth of *Task Q* and the corresponding predicted quality score, respectively.

3.1.5. Global loss function

We define the global loss function of the entire model as:

$$L = \lambda_C L_C + \lambda_T L_T + \lambda_I L_I + \lambda_Q L_Q \tag{7}$$

where $\lambda_C, \lambda_T, \lambda_I, \lambda_Q$ are the weights of the corresponding loss functions.

3.2. Progressive relevance

We find that there is progressive relevance among the tasks, besides the relevance between the auxiliary tasks and the main task. Namely, from *Task C*, *Task T*, *Task I* to *Task Q*, tasks are set up from low level to high level, and the prior knowledge of the lower-level task can guide the higher-level tasks' learning.

We randomly select 2000 reference images and 30,000 corresponding distorted images with three typical distortion types (Gaussian blur, additive white Gaussian noise, and JPEG2000) at five levels of distortion intensity from the BD dataset [44] for analysis. As shown in Fig. 2(a), the distribution histograms of the content component imbalance index f are drawn for all distortion types. As a control, the reference images are resampled five times. Then we draw the distribution histograms for five levels of distortion intensity respectively when distortion types are Gaussian blur, additive white Gaussian noise, and JPEG2000 (shown in Fig. 2.

As shown in Fig. 3(a), we can see that the distributions of images with different distortion types show different deviations from the distribution of reference images. And Fig.3 (b-d) shows that the increase of distortion intensity tends to aggravate such deviation. Hence, the image content type reflects the image distortion information including the image distortion type and intensity. Besides, the distortion intensity is of significance only if the distortion type is given. Image content type, distortion type, and distortion intensity all contain image quality information. Therefore, we think that there is progressive relevance among the tasks.

3.3. Network structure

In the proposed model, images of 224 × 224 × 3 are taken as input. As depicted in Fig. 3, our proposed model consists of three

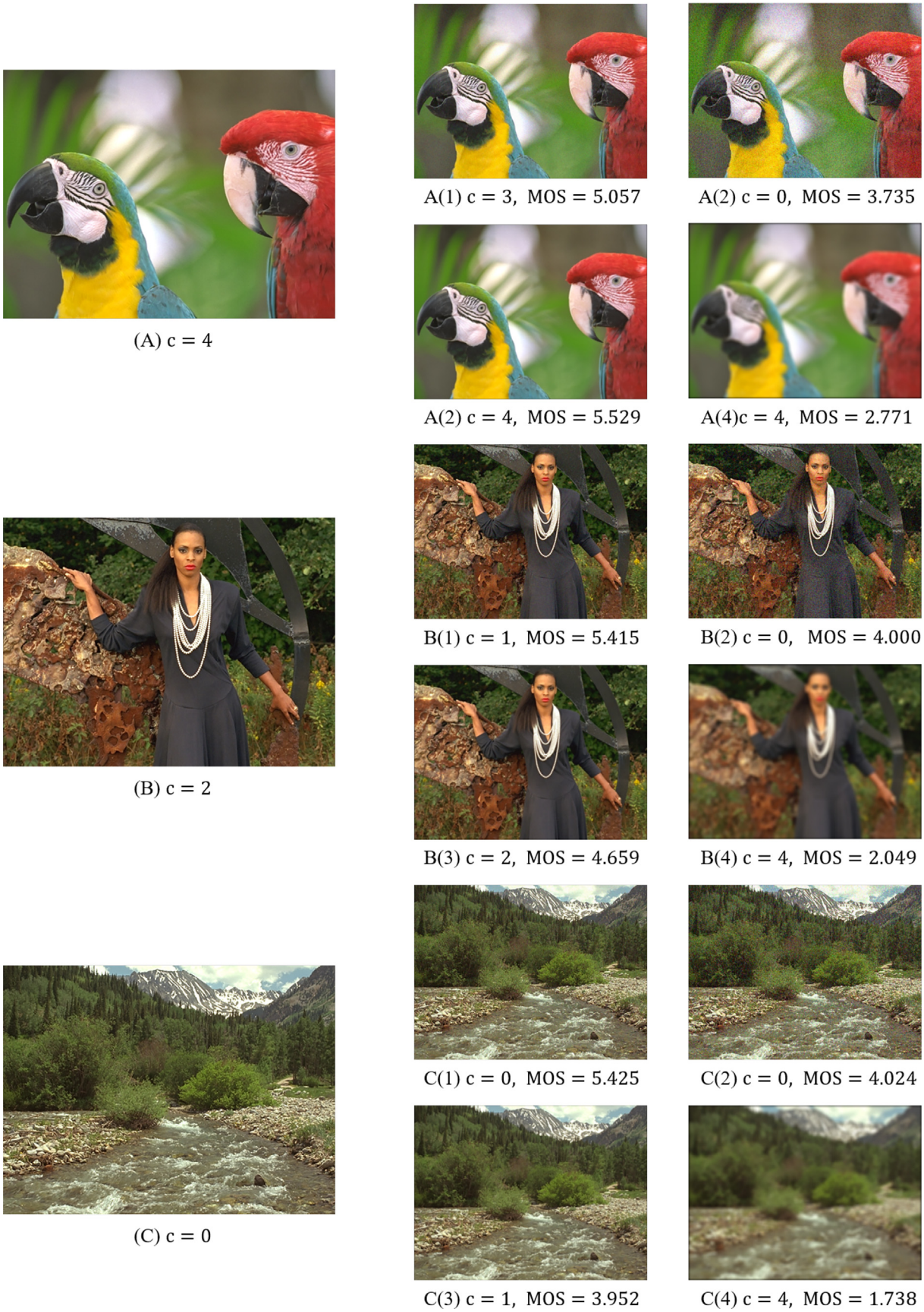

Fig. 1. Reference images with different image content and the corresponding distorted images with different distortion types and intensity: (A), (B) and (C) are respectively reference images with high smooth region, balanced content components, and high texture; A(1–4), B(1–4) and C(1–4) are the corresponding distorted images; (1), (2), (3) and (4) indicate slight additive white Gaussian noise, severe additive white Gaussian noise, slight Gaussian blur, and severe Gaussian blur, respectively. All images are from TID2013 [36]. The image content type c is computed by Eq. (2).

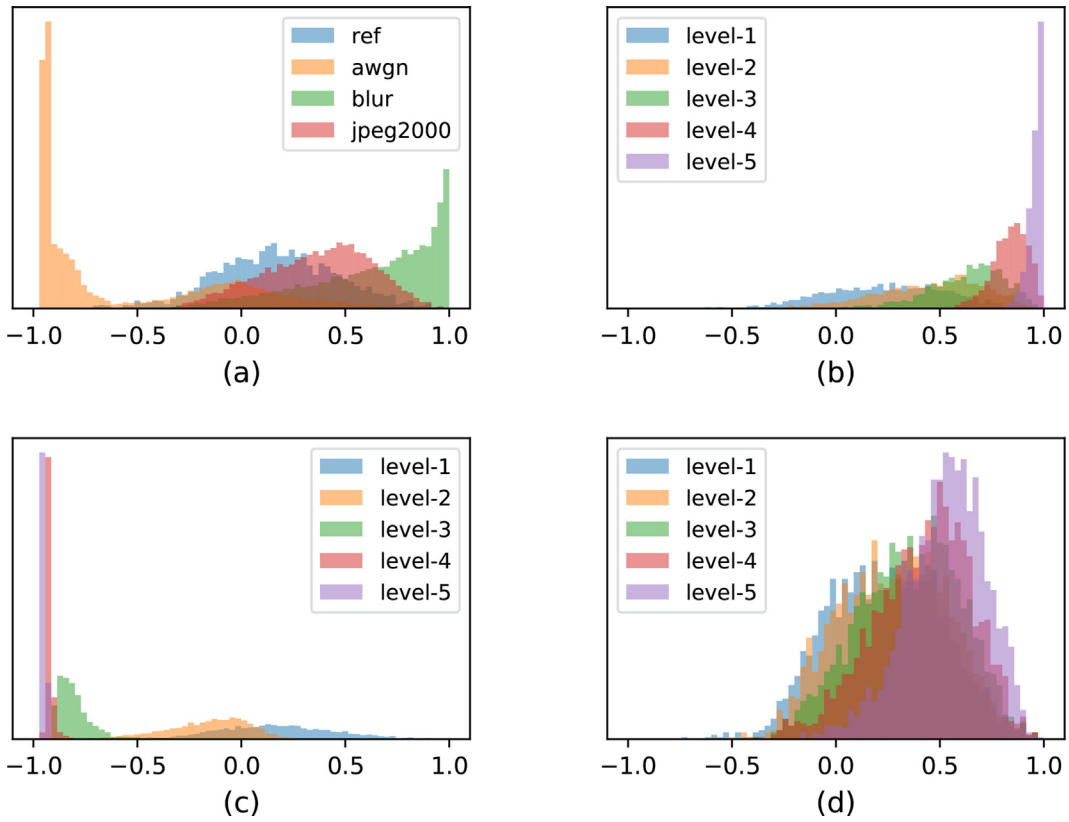

Fig. 2. Statistical analysis of selected images from BD dataset. (a) the distortion type-*f* distribution. (b) the distortion intensity-*f* distribution when distortion type is Gaussian blur. (c) the distortion intensity-*f* distribution when distortion type is additive white Gaussian noise. (d) the distortion intensity-*f* distribution when distortion type is JPEG2000.

parts: the hierarchical shared network, the progressive related network and the specific network.

3.3.1. Hierarchical shared network

The tailored ResNet34 [14], which discards the fully connected layer, is selected as our backbone. As mentioned above, from *Task C*, *Task T*, *Task I* to *Task Q*, the tasks are set up from low level to high level. And in a deep neural network, the shallow layers will learn the low-level feature then the deeper layer will learn the higher-level feature [35]. According to these characteristics, we extract hierarchical features from the backbone for the corresponding tasks. Specifically, we take the outputs from conv2_x, conv3_x, conv4_x and conv5_x in ResNet34 as features for predicting *Task C*, *Task T*, *Task I* and *Task Q*, respectively. The numbers of feature channels are respectively 64, 128, 256 and 512.

Considering the levels of the tasks, our network shares the parameters hierarchically instead of sharing the entire feature extraction network parameters directly. This allows for a more reasonable sharing of network parameters among the tasks and makes the network more interpretable.

3.3.2. Progressive related network

In the progressive related network, the features obtained by the hierarchical shared network are first fed into a convolutional layer with 3×3 kernels, 1 stride (2 stride in *Task C*) and 1 padding. Then, based on the progressive relevance among the tasks, the progressive connection structure shown in Fig. 4 is introduced. The features of the lower-level task are fused with that of the higher-level task through the progressive connection. And the fused features are passed through a convolutional layer with 3×3 kernels, 2 stride and 1 padding (zero padding in *Task Q*). The obtained features will be inputs of the specific network for predicting the corresponding tasks. The number of feature channels remains unchanged for all tasks in the progressive related network.

By using the progressive connection structure, we make full use of the progressive relevance among the tasks and explicitly strengthen the lower-level tasks' constraints on the higher-level tasks.

3.3.3. Specific network

For three classification tasks, the specific networks are composed of a convolutional layer with 3×3 kernels, 1 stride and 1 padding, a convolutional layer with 1×1 kernels, 1 stride and zero padding, and a global average pooling [28]. The number of 1×1 kernel is equal to the categories of the corresponding task, and thier dimensions are 64, 128 and 256, respectively. The quality regression network contains a fully connected layer with one dimension. Batch normalization [16] and ReLU are applied after all convolutional layers with 3×3 kernels.

3.4. Effective constraint from PMTL

The prediction errors in supervised machine learning can be decomposed into two parts: 1) the approximation error ϵ_{app} which measures how close the functions in hypothesis space H can approximate the optimal hypothesis $\hat{h}$; 2) the estimation error ϵ_{est} which measures the effect of minimizing the empirical risk $R_I(h)$ (which is the average of sample losses over the training set of I samples) instead of the expected risk $R(h)$ within H. An intuitive representation is shown in Fig. 5. Due to the lack of sufficient training data, the core issue in few-shot learning is that $R_I(h)$ may

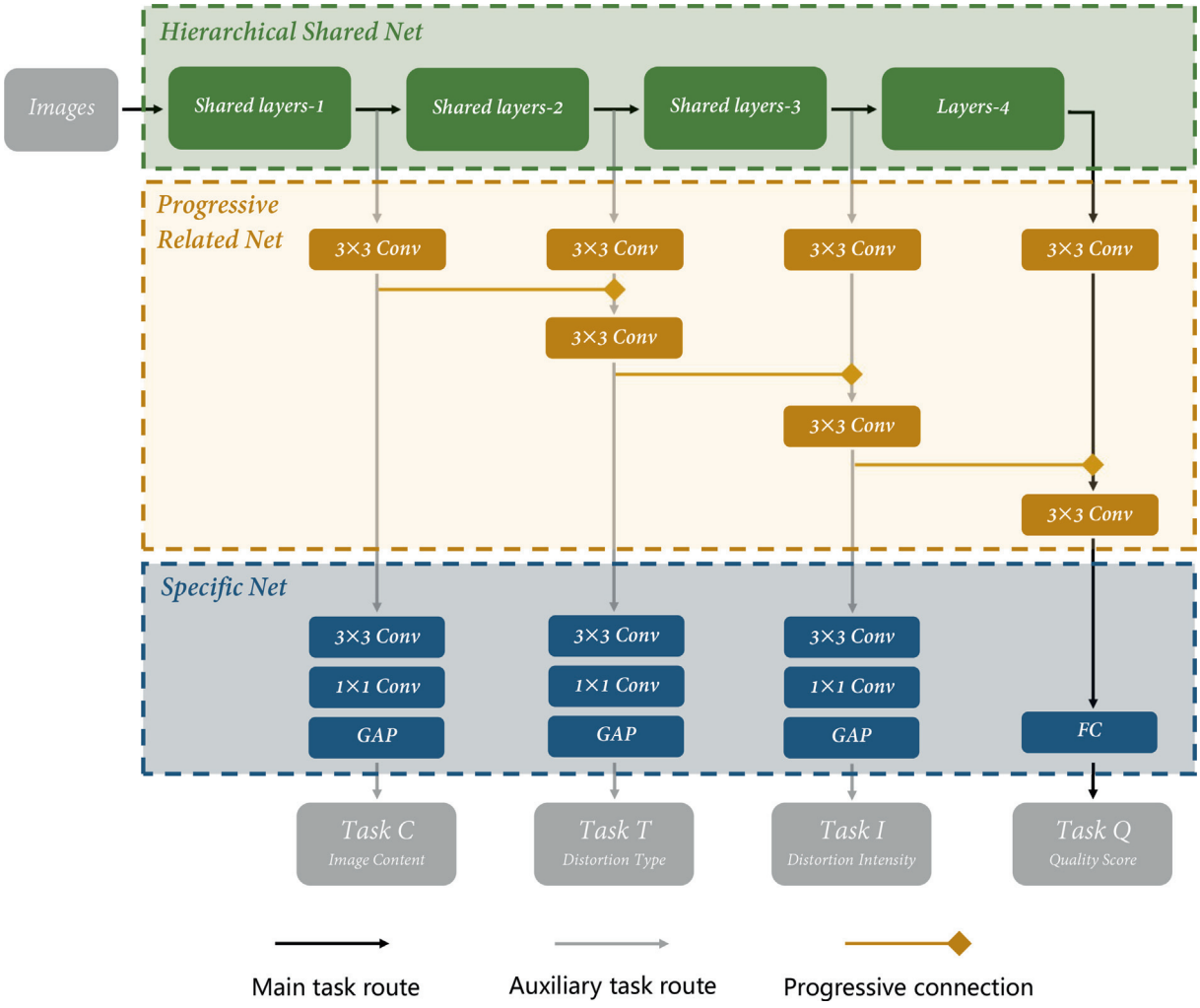

Fig. 3. Architecture of the proposed PMTL framework.

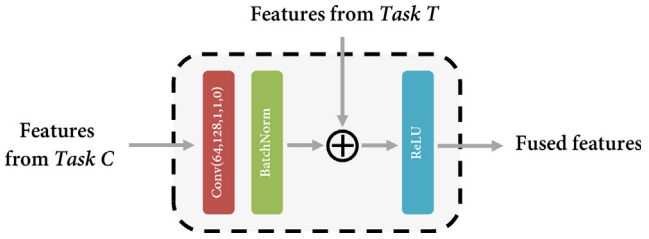

Fig. 4. The structure of the progressive connection between *Task C* and *Task T*. The presentation formats are Conv(input channels, output channels, kernel size, stride, padding).

be far from being a good approximation of $R(h)$, and the resultant empirical risk minimizer h_I overfits [43].

In this paper, we alleviate this problem by introducing prior knowledge to constrain the hypothesis space. For the sake of discussion, we simplify the PMTL model to use only an auxiliary task. The parameters of the hierarchical shared network, progressive related network and specific network in PMTL model are defined as $\theta^{sh} = [\theta_a^{sh}, \theta_m^{sh}]$, $\theta^r = [\theta_a^r, \theta_m^r]$, $\theta^{sp} = [\theta_a^{sp}, \theta_m^{sp}]$, where a indicates that the parameters are in the auxiliary task route, and m indicates that the parameters are only in the main task route. Then, we introduce MTL model and STL model as a comparison. MTL model is set to have the same skeleton as PMTL, but lacks the progressive connection structure introduced by prior knowledge from the progressive relevance. The corresponding parameters in MTL model are defined as $\theta^{sh} = [\theta_a^{sh}, \theta_m^{sh}]$, $\theta^{r} = [\theta_a^{r}, \theta_m^{r}]$, $\theta^{sp} = [\theta_a^{sp}, \theta_m^{sp}]$. For STL model,

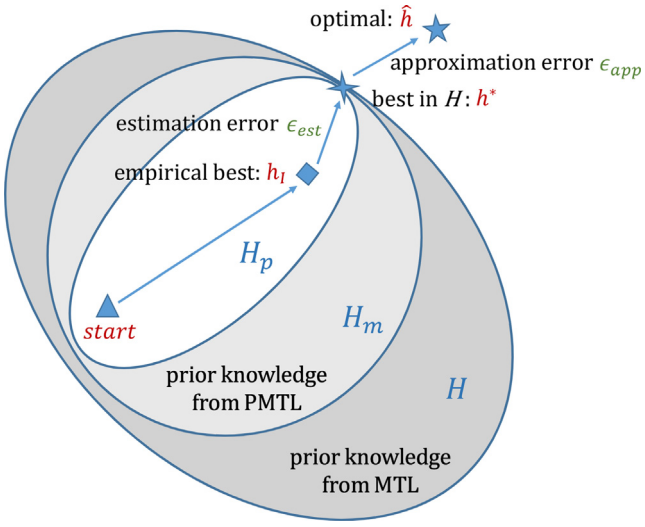

Fig. 5. The constraint on hypothesis space H obtained by introducing the prior knowledge from MTL and PMTL.

the auxiliary task branch is further removed. The corresponding parameters in STL model are defined as $\theta^{sh} = [\theta_a^{sh}, \theta_m^{sh}], \theta_m^{r\prime}, \theta_m^{sp}$.

In order to make a more objective comparison, we let $\theta_m^{r\prime}$ in STL and MTL equal to θ^r, so that the parameters of all three frameworks that are involved in the prediction of the main task are the same, i.e. their unconstrained hypothesis space of the main task is the same. Hence, the parameters of the hierarchical shared network, progressive related network and specific network in STL model and MTL model are defined as $\{\theta^{sh}, \theta^r, \theta_m^{sp}\}$ and $\{\theta^{sh}, \theta^{r\prime} = [\theta_a^{r\prime}, \theta^r], \theta^{sp}\}$.

For the STL model, the optimization problem can be expressed as:

$$argmin_{\theta^{sh},\theta^r,\theta_m^{sp}} L_m(\{\theta^{sh}, \theta^r, \theta_m^{sp}\}, X) \tag{8}$$

where X is training data.

For the MTL model, the optimization problem is a multi-objective optimization problem which can be expressed as:

$$argmin_{\theta^{sh},\theta^{r\prime},\theta^{sp}} \lambda_a L_a(\{\theta_a^{sh}, \theta_a^{r\prime}, \theta_a^{sp}\}, X) + \lambda_m L_m(\{\theta^{sh}, \theta^r, \theta_m^{sp}\}, X) \tag{9}$$

where λ_a, λ_m are the weights of the auxiliary task and main task, respectively.

By using ϵ-constraint method [49], Eq. (9) can be converted to a constrained single-objective optimization problem:

$$\begin{aligned} &argmin_{\theta^{sh},\theta^r,\theta_m^{sp}} L_m(\{\theta^{sh}, \theta^r, \theta_m^{sp}\}, X) \\ &s.t. L_a(\{\theta_a^{sh}, \theta_a^{r'}, \theta_a^{sp}\}, X) \leqslant \alpha \end{aligned} \tag{10}$$

where α is the expected loss threshold of the auxiliary task.

Obviously, compared with the STL model, the MTL model constrains θ^{sh} by introducing prior knowledge from the auxiliary task. As shown in Fig. 5, the dark gray area is not considered for optimization since it is known to be unlikely to contain the optimal h^* in H according to this prior knowledge [43]. Hence, H is constrained to H_m.

For the PMTL model, the optimization problem can be expressed as:

$$argmin_{\theta^{sh},\theta^r,\theta^{sp}} \lambda_a L_a(\{\theta_a^{sh}, \theta_a^r, \theta_a^{sp}\}, X) + \lambda_m L_m(\{\theta^{sh}, \theta^r, \theta_m^{sp}\}, X) \tag{11}$$

Similarly, Eq. (11) can be converted by using ϵ-constraint method:

$$\begin{aligned} &argmin_{\theta^{sh},\theta^r,\theta_m^{sp}} L_m(\{\theta^{sh}, \theta^r, \theta_m^{sp}\}, X) \\ &s.t. L_a(\{\theta_a^{sh}, \theta_a^r, \theta_a^{sp}\}, X) \leqslant \alpha \end{aligned} \tag{12}$$

Since θ_a^r is the part of θ^r, with the same requirements for α, θ^r is constrained by introducing prior knowledge from the progressive relevance, i.e. the light gray area in Fig. 5 is not considered for optimization, and H_m is further constrained to H_p. For the smaller hypothesis space H_p, the existing training data is sufficient to obtain a relatively reliable h_l.

4. Experimental results

In this section, the experimental setups are first described, which includes IQA datasets, protocols, and implementation details of PMTL. Then, the performances of PMTL are compared with state-of-the-art BIQA methods on individual databases, individual distortion and cross databases. Finally, ablation experiments are conducted to demonstrate the contributions of the core factors in PMTL.

4.1. Datasets and protocols

We used three standard databases including LIVE [39], CSIQ [2], TID2013 [36] for evaluation. Besides singly distorted synthetic image databases, two multiply distorted synthetic image databases LIVE-MD [17], MDID [42] and two authentic image databases LIVEC [8], KonIQ-10 k [15] are employed to further verify the generalization ability of the proposed method. All databases are summarized in Table 1.

To validate prediction monotonicity and accuracy, two widely used criteria, Spearman's rank order correlation coefficient (SRCC) and Pearson's linear correlation coefficient (PLCC), are adopted in our experiments.

1) Spearman's rank order correlation coefficient (SRCC)

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)},$$

where d_i is the rank difference between the ground truth and the predicted quality score of the i-th image, N is the number of testing images.

2) Pearson's linear correlation coefficient (PLCC)

$$PLCC = \frac{\sum_i (q_i - q_a)(\hat{q}_i - \hat{q}_a)}{\sqrt{\sum_i (q_i - q_a)^2}\sqrt{\sum_i (\hat{q}_i - \hat{q}_a)^2}} \tag{14}$$

where q_i and $\hat{q}_i$ denote the ground-truth subjective score and predicted quality score of the i-th image, q_a and $\hat{q}_a$ denote the average of each.

Both criteria are in the range [0, 1], and a higher value indicates better performance.

Table 1
Summary of databases for testing.

Database	LIVE	CSIQ	TID2013	LIVE-MD	MDID	LIVEC	KonIQ-10k
Number of referance images	29	30	25	15	20	N/A	N/A
Number of distorted images	779	866	3000	405	1600	1162	10073
Number of distortion types	5	6	24	3	5	N/A	N/A
Number of distortion levels	5	4–5	5	4	4	N/A	N/A
Number of distortions in an image	1	1	1	2	1–4	N/A	N/A
Annotation	DMOS	DMOS	MOS	DMOS	MOS	MOS	MOS
Range	[0, 100]	[0, 1]	[0, 9]	[0, 100]	[0, 8]	[0, 100]	[1, 5]
Scenario	Synthetic	Synthetic	Synthetic	Synthetic	Synthetic	Authentic	Authentic

4.2. Implementation details

The weights of loss functions $\lambda_C, \lambda_T, \lambda_I$ and λ_Q in Eq. 7 are respectively set at 0.01, 0.01, 0.01 and 1 to normalize the task weights to the same scale [13,20].

In the training process, each image is uniformly sampled 224×224 patches with stride 128 as the input of the model. The ground truths of *Task C* are calculated separately by each patch, and the other ground truths of the corresponding image are assigned to these patches. Adam [24] optimization algorithm is employed to optimize the model. The batch size is set to 64 and the learning rate is set to 10^{-4}. Besides, randomly horizontal flipping is used for data augmentation.

During testing, we extract patches in the same way for the given image, and the final quality score is predicted by simply averaging the predicted scores of the corresponding patches.

4.3. Performance on individual databases

In this subsection, the performance on individual databases of our approach is compared with the ten BIQA approaches including five hand-crafted approaches (DIIVINE [34], BLIINDS-II [38], BRISQUE [33], HOSA [46], CORNIA [48]), and five deep learning-based IQA approaches (RankIQA [29], DIQA [22], DB-CNN [51], PQR [50], CaHDC [44]) on the three benchmark databases. Each benchmark database is randomly split into 80% training data and 20% testing data by reference images for no overlapping in context. Replication experiments of this random train-test splitting are run 20 times, and the median SRCC and PLCC are finally reported. For the compared hand-crafted approaches, the performances are obtained by the corresponding authors' source codes. And for the deep learning-based approaches, the results are from the corresponding papers. The performance comparison is shown in Table 2. The symbol "–" means that these results are not provided in the corresponding papers. The top two PLCC and SRCC for these approaches are highlighted.

We can observe that our method achieves a remarkable improvement against all hand-crafted methods (SRCC increases more than 2.2% on LIVE, 16.6% on CSIQ, and 19.0% on TID2013). When compared with these deep learning-based approaches, our approaches also achieves state-of-the-art performance on the three standard databases. For LIVE, the performance of the proposed method exceeds most compared deep learning-based methods. Only DIQA [22] and RankIQA [29] are slightly better than our approach. Both on CSIQ and TID2013, our approach gets the best results and SRCC increases more than 0.4% on CSIQ and 1.6% on TID2013. In summary, our approach performs consistently well on all three benchmark datasets.

4.4. Performance on individual distortion

To investigate the performance on individual distortion, we train our framework using images with all kinds of distortion types and test it on a specific distortion type for three standard databases separately. The performance is compared with seven methods (BLIINDS-II [38], BRISQUE [33], HOSA [46], RankIQA [29], DIQA [22], MEON [31], DB-CNN [51]). The results are shown in Table 3, and the top results are highlighted. The number of times (N.o.T) of achieving the best performance for each method is listed in the last row. And the top two results are highlighted.

As can be seen, our approach gets the best results on 27 of 35 distortion types. It shows the high correlation of our approach to human perception for images with most kinds of distortion types. Meanwhile, the performance of our approach is worse than other approaches on local block-wise distortion in TID2013. In the TID2013 database, the greater the intensity of local block-wise distortion, the weaker the distortion. It is the opposite of the other types so as to make the prediction of distortion intensity mislead the prediction of quality. Moreover, in the case of low distortion intensity, most of the sampled patches are distortion-free, which may also have a negative impact on our prediction. In the remaining 7 types, the performance of our approach is slightly worse or competitive compared with the best one.

4.5. Cross-database evaluations

To measure the generalization ability of the proposed method, we test it in a more challenging cross-database setting. Specially, we conduct all experiments by training on a single entire dataset and testing on the other datasets respectively. Firstly, we run cross-database tests among the three singly distorted synthetic image databases. The results compared with eight methods (BRISQUE [33], M3 [47], FRIQUEE [9], CORNIA [48], HOSA [46], DB-CNN [51], PQR [50], CaHDC [44]) are listed in Table 4. All results of the compared methods come from existing papers.

We can see that the performance of our approach achieves a significant improvement against the compared approaches on all three cross-database evaluations. Although it performs a little worse than PQR when trained on CSIQ and tested on LIVE, it shows a high correlation with human perception (SRCC and PLCC are both larger than 0.9). It provides strong evidence that by setting the three auxiliary tasks and designing the PMTL model, the hypothesis space for IQA is effectively constrained and the generalization ability is improved significantly.

Then, we explore the generalization ability of our approach in more complex scenarios. We run the cross-database experiment by training on LIVE and testing on LIVE-MD, MDID, LIVEC and KonIQ-10 k. The results compared with six methods (BRISQUE [33], DIIVINE [34], FRIQUEE [9], HOSA [46], DB-CNN [51], PQR [50] are shown in Table 5. For the compared hand-crafted approaches, the performances are obtained by the corresponding authors' source codes. And for the deep learning-based approaches, the results are from the corresponding papers.

The proposed method also shows the fantastic generalization ability even though it is trained on a singly distorted synthetic image database and tested on authentic (or multiply distorted synthetic) image databases. We believe that more sufficient mining of prior knowledge related to perceived quality enables our method to learn IQA more essentially, so that it can adapt to more complex scenarios.

Table 2
Performance comparison on individual databases.

DB	Crit.	Proposed	DIIVINE	BLIINDS-II	BRISQUE	HOSA	CORNIA	RankIQA	DIQA	DB-CNN	PQR	CaHDC
LIVE	SRCC	0.970	0.925	0.919	0.939	0.948	0.942	**0.981**	**0.975**	0.968	0.965	0.965
	PLCC	0.972	0.923	0.920	0.942	0.949	0.943	**0.982**	**0.977**	0.971	0.971	0.964
CSIQ	SRCC	**0.950**	0.784	0.570	0.750	0.781	0.714	–	0.884	**0.946**	0.873	0.903
	PLCC	**0.958**	0.836	0.534	0.829	0.842	0.781	–	0.915	**0.959**	0.901	0.914
TID2013	SRCC	**0.878**	0.654	0.536	0.573	0.688	0.549	0.780	0.825	0.816	0.740	**0.862**
	PLCC	**0.896**	0.549	0.628	0.651	0.764	0.613	–	0.850	0.865	0.798	**0.878**

Table 3
Performance comparison (SRCC) on individual distortions.

	Type	BLIINDS-II	BRISQUE	HOSA	RankIQA	DIQA	MEON	DB-CNN	Proposed
LIVE	FF	0.874	0.828	0.954	0.954	0.912	–	0.930	**0.969**
	GB	0.915	0.964	0.954	**0.988**	0.962	–	0.935	0.960
	WN	0.947	0.982	0.975	**0.991**	0.988	–	0.980	0.899
	JPEG	0.950	0.965	0.954	0.978	0.976	–	0.972	**0.984**
	JP2K	0.930	0.929	0.935	0.970	0.961	–	0.955	**0.982**
CSIQ	CG	0.336	0.396	0.716	–	0.718	–	0.870	**0.931**
	GB	0.880	0.808	0.841	–	0.870	0.918	0.947	**0.954**
	WN	0.702	0.682	0.604	–	0.835	**0.951**	0.948	0.907
	PN	0.812	0.743	0.500	–	0.893	–	0.940	**0.946**
	JPEG	0.846	0.846	0.733	–	0.931	0.948	0.940	**0.965**
	JP2K	0.850	0.817	0.818	–	0.927	0.898	0.953	**0.977**
TID2013	AGN	0.714	0.711	0.853	0.891	0.915	0.813	0.790	**0.933**
	ANC	0.728	0.432	0.625	0.799	0.755	0.722	0.700	**0.857**
	SCN	0.825	0.746	0.782	0.911	0.878	0.926	0.826	**0.951**
	MN	0.358	0.252	0.368	0.644	**0.734**	0.728	0.646	0.631
	HFN	0.852	0.842	0.905	0.873	0.939	0.911	0.879	**0.956**
	IN	0.664	0.765	0.775	0.869	0.843	0.901	0.708	**0.955**
	QN	0.780	0.662	0.810	0.910	0.858	0.888	0.825	**0.913**
	GB	0.852	0.871	0.892	0.835	0.920	0.887	0.859	**0.946**
	DEN	0.754	0.612	0.870	0.894	0.788	0.797	0.865	**0.958**
	JPEG	0.808	0.764	0.893	0.902	0.892	0.850	0.894	**0.926**
	JP2K	0.862	0.745	**0.932**	0.923	0.912	0.891	0.916	0.928
	JGTE	0.251	0.301	0.747	0.579	0.861	0.746	0.772	**0.882**
	J2TE	0.755	0.748	0.701	0.431	0.812	0.716	0.773	**0.903**
	NEPN	0.081	0.269	0.199	0.463	0.659	0.116	0.270	**0.800**
	Block	0.371	0.207	0.327	0.693	0.407	**0.500**	0.444	0.074
	MS	0.159	0.219	0.233	0.321	0.299	0.177	−0.009	**0.525**
	CTC	−0.082	−0.001	0.294	0.657	0.687	0.252	0.548	**0.838**
	CCS	0.109	0.003	0.119	0.622	−0.151	0.684	0.631	**0.782**
	MGN	0.699	0.717	0.782	0.845	**0.904**	0.849	0.711	0.869
	CN	0.222	0.196	0.532	0.609	0.655	0.406	0.752	**0.867**
	LCNI	0.451	0.609	0.835	0.891	0.930	0.772	0.860	**0.947**
	ICQD	0.815	0.831	0.855	0.788	**0.936**	0.857	0.833	0.898
	CHA	0.568	0.615	0.801	0.727	0.756	0.779	0.732	**0.836**
	SSR	0.856	0.807	0.905	0.768	0.909	0.855	0.902	**0.949**
N.o.T		0	0	1	2	3	2	0	**27**

Table 4
Performance comparison (SRCC) on cross-database evaluations.

Train	Test	Proposed	BRISQUE	M3	FRIQUEE	CORNIA	HOSA	DB-CNN	PQR	CaHDC
LIVE	CSIQ	**0.880**	0.562	0.621	0.722	0.649	0.594	**0.758**	0.717	–
	TID2013	**0.626**	0.358	0.344	0.461	0.360	0.361	0.524	**0.551**	–
CSIQ	LIVE	**0.921**	0.847	0.797	0.879	0.853	0.773	0.877	**0.930**	–
	TID2013	**0.662**	0.454	0.328	0.463	0.312	0.329	0.540	**0.546**	–
TID2013	LIVE	**0.931**	0.790	0.873	0.755	0.846	0.846	0.758	0.891	**0.930**
	CSIQ	**0.895**	0.590	0.605	0.635	0.672	0.612	**0.807**	0.632	0.736

Table 5
Performance comparison (SRCC) on cross-database evaluations.

DB	Proposed	BRISQUE	DIIVINE	FRIQUEE	HOSA	DB-CNN	PQR
LIVEC	**0.568**	0.378	0.361	0.469	0.460	**0.567**	0.547
KonIQ-10 k	**0.613**	0.411	0.356	**0.549**	0.435	–	–
LIVE-MD	**0.825**	0.500	**0.692**	0.600	0.647	–	–
MDID	**0.820**	0.402	0.698	0.668	**0.707**	–	–

4.6. Ablation experiment

As mentioned above, we deconstruct IQA into three elements, i.e., image content, pattern of degradation, and intensity of degradation. According to these elements, three corresponding auxiliary tasks are set up to guide network for IQA learning. To investigate the impact of the auxiliary tasks on performance, we conducted several ablation experiments. All of them are conducted by training on TID2013 and testing on LIVE and CSIQ. All experimental settings are the same. The results are shown in Table 6 when different numbers of auxiliary tasks are combined. The symbol "✓" indicates that the corresponding auxiliary task is employed.

It can be seen that the performance of the model increases gradually along with the number of auxiliary tasks. And performance achieves the best performance when all auxiliary tasks are employed. We believe this is because when fewer auxiliary tasks are used, the factors of IQA are not fully considered and the implicit prior knowledge is not adequately exploited. The proposed method

Table 6
Performance comparison when different numbers of auxiliary tasks are combined.

Auxiliary tasks			LIVE		CSIQ	
Task C	Task T	Task I	PLCC	SRCC	PLCC	SRCC
✓			0.833	0.852	0.826	0.777
	✓		0.823	0.885	0.846	0.825
		✓	0.627	0.718	0.859	0.828
✓	✓		0.887	0.897	0.847	0.825
✓		✓	0.904	0.909	0.876	0.851
	✓	✓	0.902	0.920	0.904	0.884
✓	✓	✓	**0.928**	**0.931**	**0.909**	**0.895**

Table 7
Performance comparison with and without the progressive connection.

	LIVE		CSIQ	
	PLCC	SRCC	PLCC	SRCC
PMTL -H -P	0.827	0.853	0.838	0.872
PMTL -P	0.902	0.927	0.870	0.894
PMTL	**0.928**	**0.931**	**0.909**	**0.895**

adequately considers these factors and makes full use of the prior knowledge so that it can learn the effective and robust image quality representation.

In this work, due to the level of tasks and the progressive relevance among the tasks, we design the hierarchical sharing structure and progressive connection. To demonstrate the contribution of them, we train a base MTL model (all tasks are sharing the entire tailored ResNet34 for feature extracting) as our backbone model and analyze the effect of each individual component. It is trained on TID2013 and tested on LIVE, CSIQ. The performance comparison is listed in Table 7. "-H" and "-P" mean removing the hierarchical sharing structure and removing the progressive connection, respectively.

By modifying the hierarchical sharing structure to the backbone model, SRCC and PLCC increased 7.4% and 7.5% on LIVE, and 3.2% and 2.2% on CSIQ. By introducing the progressive connection, SRCC and PLCC further increased 0.4%, 2.6% on LIVE, and 0.1% and 3.9% on CSIQ. The results demonstrate the validity of the hierarchical sharing structure and progressive connection structure.

5. Conclusion

In this paper, an end-to-end IQA framework based on progressive multi-task learning has been proposed. Inspired by the definition of IQA, we have deconstructed IQA into the factors affecting perceived quality and designed the corresponding auxiliary tasks to provide prior knowledge for network learning IQA. Furthermore, the progressive relevance among the tasks has been studied. We have modified the hard parameter sharing MTL model to make full of this relationship. The experiments show that the proposed method achieves state-of-the-art performance and has strong generalization ability.

CRediT authorship contribution statement

Aobo Li: Investigation, Software, Data curation, Writing – original draft. **Jinjian Wu:** Conceptualization, Methodology, Visualization. **Shiwei Tian:** Writing – review & editing. **Leida Li:** Writing – review & editing. **Weisheng Dong:** Supervision. **Guangming Shi:** Resources, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Andalibi, D.M. Chandler, Digital image watermarking via adaptive logo texturization, IEEE Trans. Image Process. 24 (2015) 5060–5073.
[2] L.D.M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (2010) 011006.
[3] H.w. Chang, Q.w. Zhang, Q.g. Wu, Y. Gan, Perceptual image quality assessment by independent feature detector, Neurocomputing 151 (2015) 1142–1152..
[4] M. Crawshaw, Multi-task learning with deep neural networks: A survey, 2020. arXiv preprint arXiv:2009.09796..
[5] J. Dai, K. He, J. Sun, Instance-aware semantic segmentation via multi-task network cascades, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3150–3158.
[6] L. Duong, T. Cohn, S. Bird, P. Cook, Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser, in: Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing (volume 2: short papers), 2015, pp. 845–850.
[7] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, Y. Zhu, Deepsim: Deep similarity for image quality assessment, Neurocomputing 257 (2017) 104–114.
[8] D. Ghadiyaram, A.C. Bovik, Massive online crowdsourced study of subjective and objective picture quality, IEEE Trans. Image Process. 25 (2015) 372–387.
[9] D. Ghadiyaram, A.C. Bovik, Perceptual quality prediction on authentically distorted images using a bag of features approach, J. Vis. 17 (2017), 32–32.
[10] R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-based convolutional networks for accurate object detection and segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2015) 142–158.
[11] S. Golestaneh, L.J. Karam, Reduced-reference quality assessment based on the entropy of dwt coefficients of locally weighted gradient magnitudes, IEEE Trans. Image Process. 25 (2016) 5293–5303.
[12] W. Guo, G. Chen, Human action recognition via multi-task learning base on spatial–temporal feature, Inf. Sci. 320 (2015) 418–428.
[13] K. Hassani, M. Haley, Unsupervised multi-task feature learning on point clouds, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 8160–8171.
[14] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[15] V. Hosu, H. Lin, T. Sziranyi, D. Saupe, Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment, IEEE Trans. Image Process. 29 (2020) 4041–4056.
[16] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, International conference on machine learning, PMLR (2015) 448–456.
[17] D. Jayaraman, A. Mittal, A.K. Moorthy, A.C. Bovik, Objective quality assessment of multiply distorted images, in: 2012 Conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR), IEEE, 2012, pp. 1693–1697.
[18] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1733–1740.
[19] L. Kang, P. Ye, Y. Li, D. Doermann, Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks, in: 2015 IEEE International Conference on Image Processing, 2015, pp. 2791–2795.
[20] A. Kendall, Y. Gal, R. Cipolla, Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7482–7491.
[21] J. Kim, S. Lee, Fully deep blind image quality predictor, IEEE J. Sel. Top. Signal Process. 11 (2016) 206–220.
[22] J. Kim, A.D. Nguyen, S. Lee, Deep cnn-based blind image quality predictor, IEEE Trans. Neural Networks Learn. Syst. 30 (2018) 11–24.

[23] W. Kim, A.D. Nguyen, S. Lee, A.C. Bovik, Dynamic receptive field generation for full-reference image quality assessment, IEEE Trans. Image Process. 29 (2020) 4219–4231.
[24] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014. arXiv preprint arXiv:1412.6980..
[25] Q. Li, W. Lin, Y. Fang, Bsd: Blind image quality assessment based on structural degradation, Neurocomputing 236 (2017) 93–103.
[26] Q. Li, W. Lin, K. Gu, Y. Zhang, Y. Fang, Blind image quality assessment based on joint log-contrast statistics, Neurocomputing 331 (2019) 189–198.
[27] Y. Li, L.M. Po, X. Xu, L. Feng, F. Yuan, C.H. Cheung, K.W. Cheung, No-reference image quality assessment with shearlet transform and deep neural networks, Neurocomputing 154 (2015) 94–109.
[28] Lin, M., Chen, Q., Yan, S., 2013. Network in network. arXiv preprint arXiv:1312.4400..
[29] X. Liu, J. van de Weijer, A.D. Bagdanov, Rankiqa: Learning from rankings for no-reference image quality assessment, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1040–1049.
[30] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
[31] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, W. Zuo, End-to-end blind image quality assessment using deep neural networks, IEEE Trans. Image Process. (2018) 1–1..
[32] I. Misra, A. Shrivastava, A. Gupta, M. Hebert, Cross-stitch networks for multi-task learning, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3994–4003.
[33] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Processing 21 (2012) 4695–4708.
[34] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: From natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (2011) 3350–3364.
[35] M.M. Najafabadi, F. Villanustre, T.M. Khoshgoftaar, N. Seliya, R. Wald, E. Muharemagic, Deep learning applications and challenges in big data analytics, J. Big Data 2 (2015) 1–21.
[36] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F.a. Battisti, Color image database tid2013: Peculiarities and preliminary results, in: European Workshop on Visual Information Processing, 2013, pp. 106–111..
[37] A. Rehman, Z. Wang, Reduced-reference image quality assessment by structural similarity estimation, IEEE Trans. Image Process. 21 (2012) 3378–3389.
[38] M.A. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the dct domain, IEEE Trans. Image Process. 21 (2012) 3339–3352.
[39] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (2006) 3440–3451.
[40] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014. arXiv preprint arXiv:1409.1556..
[41] W. Sun, Q. Liao, J.H. Xue, F. Zhou, Spsim: A superpixel-based similarity index for full-reference image quality assessment, IEEE Trans. Image Process. 27 (2018) 4232–4244.
[42] W. Sun, F. Zhou, Q. Liao, Mdid: A multiply distorted image database for image quality assessment, Pattern Recogn. 61 (2017) 153–168.
[43] Y. Wang, Q. Yao, J.T. Kwok, L.M. Ni, Generalizing from a few examples: A survey on few-shot learning, ACM Comput. Surv. 53 (2020) 1–34.
[44] J. Wu, J. Ma, F. Liang, W. Dong, W. Lin, End-to-end blind image quality prediction with cascaded deep neural network, IEEE Trans. Image Process. (2020) 1–1..
[45] J. Wu, Y. Wu, R. Che, Y. Liu, Perceptual sensitivity based image structure-texture decomposition, in: 2020 IEEE Conference on Multimedia Information Processing and Retrieval, IEEE, 2020, pp. 336–341.
[46] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, D. Doermann, Blind image quality assessment based on high order statistics aggregation, IEEE Trans. Image Process. 25 (2016) 4444–4457.
[47] W. Xue, X. Mou, L. Zhang, A.C. Bovik, X. Feng, Blind image quality assessment using joint statistics of gradient magnitude and laplacian features, IEEE Trans. Image Process. 23 (2014) 4850–4862.
[48] P. Ye, J. Kumar, L. Kang, D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, in: 2012 IEEE conference on computer vision and pattern recognition, IEEE, 2012, pp. 1098–1105.
[49] Y.H. Yv, L. Lasdon, D.W. Da, On a bicriterion formation of the problems of integrated system identification and system optimization, IEEE Trans. Syst. Man Cybern. (1971) 296–297.
[50] H. Zeng, L. Zhang, A.C. Bovik, Blind image quality assessment with a probabilistic quality representation, in: 2018 25th IEEE International Conference on Image Processing, 2018, pp. 609–613.
[51] W. Zhang, K. Ma, J. Yan, D. Deng, Z. Wang, Blind image quality assessment using a deep bilinear convolutional neural network, IEEE Trans. Circuits Syst. Video Technol. 30 (2018) 36–47.
[52] W. Zhang, K. Ma, G. Zhai, X. Yang, Uncertainty-aware blind image quality assessment in the laboratory and wild, IEEE Trans. Image Process. 30 (2021) 3474–3486.
[53] Z. Zhang, P. Luo, C.C. Loy, X. Tang, Facial landmark detection by deep multi-task learning, European conference on computer vision, Springer (2014) 94–108.

Aobo Li received the B.S. degree from Xidian University, Xi'an, China, in 2019. He is currently pursuing the Ph.D. degree with the School of Artificial Intelligence, Xidian University, Xi'an, China. His research interests include image processing, and image/video quality assessment.

Jinjian Wu received the B.Sc. and Ph.D. degrees from Xidian University, Xi'an, China, in 2008 and 2013, respectively. From 2011 to 2013, he was a Research Assistant with Nanyang Technological University, Singapore, where he was a Post-Doctoral Research Fellow from 2013 to 2014. From 2015 to 2019, he was an Associate Professor with Xidian University, where he had been a Professor since 2019. His research interests include visual perceptual modeling, biomimetic imaging, quality evaluation, and object detection. He received the Best Student Paper Award at the ISCAS 2013. He has served as associate editor for the journal of Circuits, Systems and Signal Processing (CSSP), the Special Section Chair for the IEEE Visual Communications and Image Processing (VCIP) 2017, and the Section Chair/Organizer/TPC member for the ICME2014-2015, PCM2015-2016, ICIP2015, VCIP2018, and AAAI2019 Quality Assessment.

Shiwei Tian received the B.S. degree in electronic information engineering from Xidian University, Xian, in 2008, and the M.S and Ph.D. degrees in communication and information system from the Army Engineering University of PLA, Nanjing, in 2011 and 2015, respectively. Since 2015, he has been an Assistant Professor with the College of Communications Engineering, Army Engineering University, and since 2021, He has been working in National Innovation Institute of Defense Technology. His main research interests are satellite navigation, cooperative positioning and machine learning.

Leida Li (M'14) received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively. In 2008, he was a Research Assistant with the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan. From 2014 to 2015, he was a Visiting Research Fellow with the Rapid-Rich Object Search (ROSE) Lab, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he was a Senior Research Fellow from 2016 to 2017. He is currently a Professor with the School of Artificial Intelligence, Xidian University, China. His research interests include multimedia quality assessment, affective computing, information hiding, and image forensics. He has served as a SPC for IJCAI 2019–2021, Session Chair for ICMR 2019 and PCM 2015, and TPC for CVPR 2021, ICCV 2021, AAAI 2019–2021, ACM MM 2019–2020, ACM MM-Asia 2019, ACII 2019, and PCM 2016. He is currently an Associate Editor of the Journal of Visual Communication and Image Representation and the EURASIP Journal on Image and Video Processing.

Weisheng Dong (M'11) received the B.S. degree in electronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2004, and the Ph.D. degree in circuits and system from Xidian University, Xi'an, China, in 2010. He was a Visiting Student with Microsoft Research Asia, Bejing, China, in 2006. From 2009 to 2010, he was a Research Assistant with the Department of Computing, Hong Kong Polytechnic University, Hong Kong. In 2010, he joined Xidian University, as a Lecturer, and has been a Professor since 2016. He is now with the School of Artificial Intelligence, Xidian University. His research interests include inverse problems in image processing, sparse signal representation, and image compression. He was a recipient of the Best Paper Award at the SPIE Visual Communication and Image Processing (VCIP) in 2010. He is currently serving as an associate editor of IEEE Transactions on Image Processing and SIAM Journal of Imaging Sciences.

Guangming Shi (Fellow, IEEE) received the B.S. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi'an, China, in 1985, 1988, and 2002, respectively. He had studied at the University of Illinois and University of Hong Kong. Since 2003, he has been a Professor with the School of Electronic Engineering, Xidian University. He awarded Cheung Kong scholar Chair Professor by ministry of education in 2012. He is currently the Academic Leader on circuits and systems, Xidian University. His research interests include compressed sensing, brain cognition theory, multirate filter banks, image denoising, low-bitrate image and video coding, and implementation of algorithms for intelligent signal processing. He has authored or co-authored over 200 papers in journals and conferences. He served as the Chair for the 90th MPEG and 50th JPEG of the international standards organization (ISO), technical program chair for FSKD06, VSPC 2009, IEEE PCM 2009, SPIE VCIP 2010, IEEE ISCAS 2013.