No-Reference Image Quality Assessment with Reinforcement Recursive List-Wise Ranking

Jie Gu,[1,2] **Gaofeng Meng,**[1] **Cheng Da,**[1,2] **Shiming Xiang,**[1,2] **Chunhong Pan**[1]
[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
{jie.gu, gfmeng, cheng.da, smxiang, chpan}@nlpr.ia.ac.cn

Abstract

Opinion-unaware no-reference image quality assessment (NR-IQA) methods have received many interests recently because they do not require images with subjective scores for training. Unfortunately, it is a challenging task, and thus far no opinion-unaware methods have shown consistently better performance than the opinion-aware ones. In this paper, we propose an effective opinion-unaware NR-IQA method based on reinforcement recursive list-wise ranking. We formulate the NR-IQA as a recursive list-wise ranking problem which aims to optimize the whole quality ordering directly. During training, the recursive ranking process can be modeled as a Markov decision process (MDP). The ranking list of images can be constructed by taking a sequence of actions, and each of them refers to selecting an image for a specific position of the ranking list. Reinforcement learning is adopted to train the model parameters, in which no ground-truth quality scores or ranking lists are necessary for learning. Experimental results demonstrate the superior performance of our approach compared with existing opinion-unaware NR-IQA methods. Furthermore, our approach can compete with the most effective opinion-aware methods. It improves the state-of-the-art by over 2% on the CSIQ benchmark and outperforms most compared opinion-aware models on TID2013.

Introduction

Nowadays digital images are everywhere. Unfortunately, the images are often distorted at various stages of their life cycle and the distortions may lead to a reduction of the experience of human viewers. Objective image quality assessment (IQA) refers to the technique of automatically predicting the perceptual quality of distorted images. It plays an important role in many applications, like image restoration, compression, transmission, super-resolution and enhancement.

IQA has gained much attention over the past decade, with a lot of methods proposed (Bovik 2013). Existing IQA methods can be roughly classified into three categories according to the availability of the distortion-free reference (original) image. They are the full-reference (FR, where the reference image is fully available, *e.g.*, (Zhang, Shen, and Li 2014; Wang et al. 2004)), reduced-reference (RR, where only partial information about the reference is available, *e.g.*, (Ma et al. 2011)) and no-reference/blind (NR/B) IQA models. NR-IQA does not require information from the reference image. It has a wide range of applications as usually no reference is available in practice. Recently, many efforts have been made to develop general-purpose (non-distortion-specific) NR-IQA methods, which require no access to not only the reference but also the distortion type.

A large proportion of current general-purpose NR-IQA methods are regression-based (Gastaldo, Zunino, and Redi 2013; Kim et al. 2017). They typically first extract quality-aware representations, *e.g.*, natural scene statistics (NSS) features (Mittal, Moorthy, and Bovik 2012) or learning-based feature representations (Ye et al. 2012; Kang et al. 2014; Tang, Joshi, and Kapoor 2014; Kim and Lee 2017). Then a regression model, implemented by an SVR or a neural network, maps the extracted features (representations) to the quality score. This kind of methods are often considered as "opinion-aware" (OA) since the ground-truth data for supervised learning is mostly the mean opinion score (MOS). One of the main drawbacks of these methods is the lack of a large MOS-aware IQA database for training. Accordingly, their performance and generalization ability are questionable on real-world images.

Some other recent studies explore to formulate the NR-IQA as a quality-based ranking problem. They typically do not regress the image to a specific quality score but learn to rank a pair of images according to their visual quality. There are several advantages with such a formulation. One is that the preference label (representing the relative quality of two images) is generally more reliable and valid than the quality score (Keelan 2002). Thus building a pair-wise ranking model for quality assessment is fully reasonable and may probably benefit the performance (Gao et al. 2015). Another is that learning from pair-wise rankings can be used as a data augmentation technique. Collecting MOS is slow, cumbersome and expensive, which leads to the absence of large IQA databases. In contrast, the preference labels associated with image pairs can be intuitively generated by identifying distortion levels (Liu, van de Weijer, and Bagdanov 2017) or applying FR-IQA models (Ma et al. 2017b). However, there are also problems with the pair-wise approaches (Liu 2009), *e.g.*, they transforms the ranking into classification on image pairs rather than modeling it directly.

In this paper, we develop a new NR-IQA method based

on reinforcement recursive ranking. The recursive list-wise ranking formulation and reinforcement learning make our approach radically different from previous regression- and pair-wise comparison based NR-IQA methods. Specifically, we use image lists as instances in learning and separate the ranking as a sequence of nested sub-problems. During training, the recursive ranking process is formulated as a Markov decision process (MDP), and the model parameters can be trained by policy-based reinforcement learning. Our method does not require human opinion scores for training, and thus is "opinion-unaware" (OU). Moreover, it allows for end-to-end training with the standard back-propagation.

The major contributions of this paper are as follows.

- We formulate the opinion-unaware NR-IQA as a recursive list-wise ranking problem, which directly learns to rank a list of images with implicit quality measures. Unlike pair-wise approaches, our method directly optimizes the whole quality ordering and can achieve better performance on the ordering-based evaluation criteria (*e.g.*, SROCC).
- The ranking process is separated as a sequence of nested sub-problems and can be modeled as an MDP, in which the model parameters can be effectively trained by reinforcement learning. Moreover, the reinforcement learning allows for a flexible step-wise training with weak supervision, in which no ground-truth ranking lists are necessary.
- Another benefit of our method is that its training does not require MOSs, which are critical for the learning of most opinion-aware NR-IQA methods. Thus it can be trained on a large database with diverse image contents to achieve better generalization ability. Experimental results demonstrate the superior performance of our method compared with current opinion-unaware methods.

To the best of our knowledge, it is the first time to exploit reinforcement learning with list-wise learning to rank for NR-IQA. The core code of our method will be released at https://github.com/m2408gj/RRLRIQA.

Related Work

We introduce previous related works in this part, including a brief review of opinion-unaware, DNN-based and ranking-based NR-IQA methods.

Opinion-unaware NR-IQA

Opinion-unaware NR-IQA models do not require subjective scores for training, and thus are of great interest given the fact that obtaining human judgments is expensive and time-consuming. Natural image quality evaluator (NIQE) (Mittal, Soundararajan, and Bovik 2013) is one of the pioneering models. It builds a multivariate Gaussian (MVG) model for fitting NSS features. The quality of a testing image is measured by the distance between the MVG model constructed from the image itself and that from a set of pristine images. Later (Zhang, Zhang, and Bovik 2015) extended NIQE to integrated local (IL-) NIQE by introducing three additional types of statistical features and performing the quality prediction in a local manner. (Xue, Zhang, and Mou 2013) proposed a quality-aware clustering (QAC) method. It assigns each patch a quality score with a FR-IQA metric. Then clustering is applied to the image patches at different quality levels, and the cluster centroids act as a codebook for quality estimation. A simple yet effective method for extending opinion-aware NR-IQA models to opinion-unaware ones is presented by (Ye, Kumar, and Doermann 2014). Instead of training on subjective scores, the authors proposed to train models on synthetic scores derived from FR-IQA metrics.

DNN-based OA NR-IQA

Recently, significant progresses have been made in NR-IQA by exploring DNNs for better quality-aware feature extraction. For example, deep belief network (DBN) has been explored in (Tang, Joshi, and Kapoor 2014), where the belief network is employed to generate better representations from pre-extracted features. (Kang et al. 2014) first applied a shallow CNN to NR-IQA by regressing raw image patches on subjective scores without hand-crafted features. (Kim and Lee 2017) developed a deep CNN model where the local targets in patch-wise training are derived by FR-IQA metrics. To achieve image-wise training, (Lu et al. 2015) presented a multi-patch aggregation method with two strategies, *i.e.*, fully-connected and statistics sorting. (Gu et al. 2018) proposed to extract features and perform quality estimation within a vector regression framework, which can be integrated with different CNNs and benefit the performance. Generative adversarial network (GAN) based NR-IQA models have been proposed by (Lin and Wang 2018) and (Ren, Chen, and Wang 2018). Typically a generative network is developed to restore the original "distortion-free" images from distorted ones, and then the quality assessment can be done by a evaluation network similar to FR-IQA metrics.

In our method, the network can be trained with no access to MOSs. The training is performed by reinforcement learning, which is very different from previous CNN-based methods. Our model has the potential to be applied in many practical situations where the quality orderings are more convenient to obtain than the opinion scores.

Ranking-based NR-IQA

This type of methods target at ranking images instead of assigning quality scores. Pair-wise learning to rank is the most widely used framework. (Gao et al. 2015) exploited preference image pairs and formulated the learning of the mapping from image features to preference label as one of classification. (Liu, van de Weijer, and Bagdanov 2017) treated learning from rankings as a data augmentation strategy. A Siamese network is first trained on synthetically generated ranking data for pair-wise comparison, and then fine-tuned with subjective scores for absolute quality score estimation. Quality discriminable image pairs, each one assigned with a perceptual uncertainty, are explored by (Ma et al. 2017b). The perceptual uncertainty serves as a weight of the corresponding image pair in the loss function. (Ma et al. 2016) applied pair-wise rank learning to retargeted NR-IQA to measure the retargeted image qualities by their orders.

Our approach differs from previous ranking-based methods in mainly two points. The first one is the formulation of the ranking with an MDP, which allows to train the model at

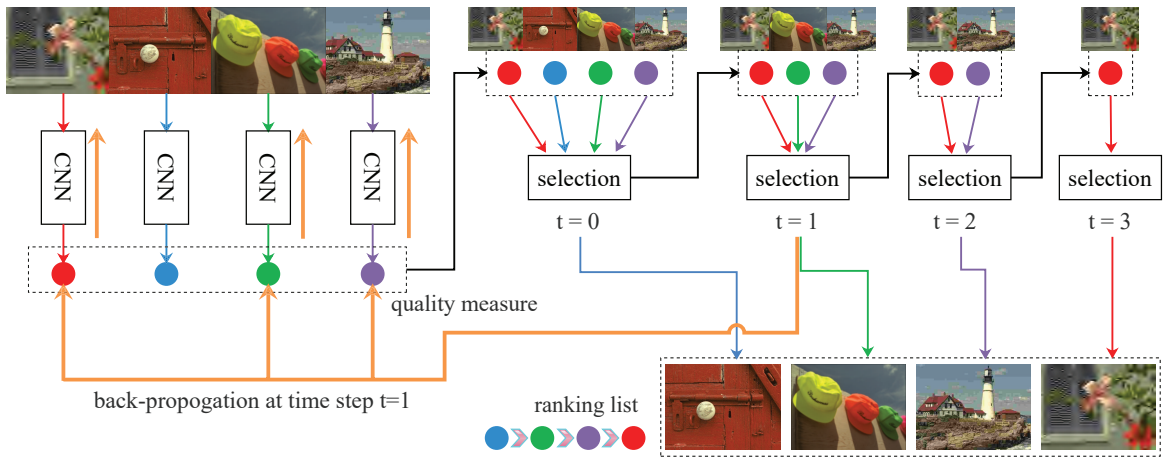

Figure 1: Illustration of our approach. The correct ranking list of the given four images can be generated by taking four actions. The first one refers to choosing the image with the best quality in these four candidate images (in this case, the blue one). The chosen image is ranked at the first position in the ranking list and then removed in the following three steps. The second action selects the best-quality image from the remaining three candidates (the green one). The chosen image is placed to the second ranking position and removed in next two steps. The process is continued until the ranking list is determined. During training, our approach allows for the full back-propagation of derivatives through the network. The dark-yellow arrows indicate the back-propagation at the second step.

each ranking position. The other one lies in that we consider the ranking positions of a list of images directly. The performance may probably benefit from making the ranking position information explicit to the learning process (Liu 2009).

NR-IQA via Reinforcement Recursive List-wise Ranking

Basic Principle

In image quality research one is interested in the relationship between the quality of images. Understanding the quality ordering of images is usually important. For example, for applications like image restoration, it is necessary to identify whether the quality of an image is improved after processing. In this study, we formulate the NR-IQA as a recursive ranking problem. Then a reinforcement list-wise approach is developed to tackle the quality-based ranking.

We explore to use image lists as learning instances mainly to optimize the quality ordering directly. The ranking is performed recursively by separating the ranking problem as a sequence of nested sub-problems. Specifically, the recursive ranking process in terms of quality estimation can be formulated as a Markov decision process (MDP) during training. The ranking list of N images can be generated by making a sequence of N decisions. An example of the ranking process is illustrated in Figure 1. The t-th decision ($t \in \{0, 1, \ldots, N-1\}$) refers to selecting an image from remaining candidates (there are $N-t$ images left) according to a policy. The chosen image is ranked at the position t in the ranking list and then removed in following steps. The process is repeated until all images are ranked.

One can see that the generated ranking list is correct if and only if the selected image always has the best quality among the candidates at every step. If so, the images will be ranked from the best-quality to the worst-quality. Accordingly, the learning is realized by rewarding or punishing each decision made by the agent depending on whether the chosen image has the best-quality. Finally, an optimal policy is achieved. Given any image list, the image with the best quality will always have a great probability to be selected according to the policy. We detail the specific MDP for ranking-based NR-IQA and the reinforcement learning in next two subsections.

MDP for Quality-based Ranking

Here we introduce how to formalize the ranking in terms of quality as an MDP. An MDP formally describes the environment for reinforcement learning and mathematically can be represented as a tuple $< \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma >$, *i.e.*, the state set, the action set, the state transition probability matrix, the reward function and the discount factor. In the following, we first give an overview of the MDP for quality-based ranking and then detail the designs.

Ranking Process: The quality ordering of a list of images can be directly inferred if we have a method that can always pick out the image with the best quality. Formally, given an image list $(I_1, I_2, \ldots, I_N)$, we aim to rank these images according to their perceptual quality from the best to the worst. There are N (*i.e.*, the size of the list) ranking positions, and each of them corresponds to a time step in MDP. At time step $t \in \{0, 1, \ldots, N-1\}$, t images with better quality have already been picked out and removed from the image list. The agent finds the best-quality image from the remaining ones. This image will be selected and placed to the ranking position t. Then it is removed from the image list in the following steps. The process is continued until all the images

are ranked, and accordingly the quality ordering of the given images is determined.

States: The state s_t of our MDP at the time step t is defined as the remaining images (candidates) that need to be further ranked in terms of the quality, denoted as $\{I_1^t, I_2^t, \ldots, I_{N_t}^t\}$. N_t is the number of the remaining images. One can observe that the presented state s_t is *Markov* because it fully defines the candidate images that the agent can select in subsequent processes.

Actions: The action set $\mathcal{A}_t$ contains all possible actions that the agent can take in the current state s_t. At time step t, we aim to pick out the image with the best quality from the remaining image list $\{I_1^t, I_2^t, \ldots, I_{N_t}^t\}$. Thus the set $\mathcal{A}_t$ can be represented as $\{1, 2, \ldots, N_t\}$, namely the list indexes of the images. An action $a_t \in \mathcal{A}_t$ being taken means that the quality of the a_t-th image is considered to be the best among the remaining ones.

Policy: A policy $\pi(a|s)$ fully defines the behaviour of an agent, and mathematically is a distribution over possible actions given states. In our framework, we use a network for encoding raw images and softmax to generate the possibilities of the a_t-th image being picked out by the agent at the time step t, *i.e.*,

$$\pi(a_t|s_t) = \frac{\exp[f(I_{a_t}^t)]}{\sum_{a \in \mathcal{A}_t} \exp[f(I_a^t)]} \ , \tag{1}$$

where $f(\cdot)$ represents the network output, which is designed to be a single scalar. The aim of the model learning is to find an optimal policy that can distinguish the perceptual quality.

Transition: Taking actions will change the states of environment. The state transition $\mathcal{P}$ defines the transition probabilities from current states to new ones. At time step t in our MDP, image $I_{a_t}^t$ is selected for the ranking position t and then removed in subsequent steps. Accordingly, the state at time step $t+1$, *i.e.*, $s_{t+1} = \{I_1^{t+1}, I_2^{t+1}, \ldots, I_{N_{t+1}}^{t+1}\}$, is

$$s_{t+1} = \mathcal{P}(s_t, a_t) = s_t \setminus \{I_{a_t}^t\}. \tag{2}$$

Reward Function: The immediate reward function $\mathcal{R}$ indicates whether or not an action is encouraged in current state with an award or a punishment. In our case, it is natural to consider the reward function as an indicator of whether the selected image has the best quality among the candidates. Specifically, a binary reward function $\mathcal{R}(s_t, a_t)$ is presented. $\mathcal{R}(s_t, a_t)$ equals to 1 if the quality of the selected image $I_{a_t}^t$ is indeed the best in $\{I_1^t, I_2^t, \ldots, I_{N_t}^t\}$, and equals to 0 otherwise.

Reinforcement Learning and Implementation

Policy Optimization Policy based reinforcement learning is an optimization problem. Given a policy π_θ with the parameters θ, the optimization is to find the best θ that maximizes a policy objective function $J(\theta)$. Policy gradient algorithms are widely used for the optimization. They search for a local maximum of $J(\theta)$ by ascending the policy gradient, *i.e.*, $\Delta\theta = \alpha \nabla_\theta J(\theta)$, where α is the step-size and $\nabla_\theta J(\theta)$ is the *policy gradient*.

According to the *policy gradient theorem*, the policy gradient can be calculated as

$$\nabla_\theta J(\theta) = \mathbb{E}_{\pi_\theta} \left[\nabla_\theta \log \pi_\theta(a|s) \ Q_{\pi_\theta}(s, a) \right] \ , \tag{3}$$

Algorithm 1: MDP for Quality-based Ranking

Input: IQA training data set $\mathfrak{D}$, size of training image list N, learning rate α, number of batches M_b, batch size M_{bz}, and discount factor γ
Output: trained network
Initialize network with res-50 pre-trained on ImageNet;
for $i = 1; i \leq M_b$ **do**
  $\Delta\theta = 0$;
  for $j = 1; j \leq M_{bz}$ **do**
    Initialize $s_0 = \{I_1^0, I_2^0, \ldots, I_{N_0}^0\}$: randomly select N images ($N_0 = N$) from $\mathfrak{D}$;
    Feed the selected images to the network that generates $\{f_\theta(I_1^0), f_\theta(I_2^0), \ldots, f_\theta(I_{N_0}^0)\}$;
    for $t = 0; t < N$ **do**
      Sample an action $a_t \in \mathcal{A}_t$ according to $\pi_\theta(a_t|s_t)$;
      Assign the reward $r_{t+1} = \mathcal{R}(s_t, a_t)$;
      Remove the selected image $I_{a_t}^t$ out of the candidate list (moving to s_{t+1});
    for $t = 0; t < N$ **do**
      $\mathrm{v}_t = \sum_{k=0}^{N_t - 1} \gamma^k \, r_{t+k+1}$;
      $\nabla_\theta \log \pi_\theta(a_t|s_t) =$
      $\frac{1}{\pi_\theta(a_t|s_t)} \sum_{a \in \mathcal{A}_t} \frac{\partial \, \pi_\theta(a_t|s_t)}{\partial f_\theta(I_a^t)} \nabla_\theta f_\theta(I_a^t)$;
    $\Delta\theta = \Delta\theta + \sum_t \nabla_\theta \log \pi_\theta(a_t|s_t) \, \mathrm{v}_t$;
  Update the parameters: $\theta = \theta + \alpha \Delta\theta$;
return trained network;

where the action-value $Q_{\pi_\theta}(s, a)$ is the expected long-term return starting from the state s, taking an action a, and then following the policy π_θ.

In this work, we adopt the Monte-Carlo Policy Gradient, *a.k.a.*, REINFORCE, to learn the parameters. There are two keys, namely to update parameters by stochastic gradient ascent, and to use the return in episodic environments, denoted as v_t, as an unbiased sample of $Q_{\pi_\theta}(s_t, a_t)$ in Eq. (3). The return v_t is the total discounted reward from time-step t in a sampled episode, *i.e.*, $\mathrm{v}_t = \sum_{k=0}^{N_t - 1} \gamma^k \, r_{t+k+1}$, where $r_{t+k+1} = \mathcal{R}(s_{t+k}, a_{t+k})$.

Implementation and Training We detail the training process in Algorithm 1. The parameters of the presented model are given by the parameters of the network for encoding images in Eq. (1). The network is practically implemented by res-50 (He et al. 2016) (substituting the softmax layer with a 1-D fully connected layer), a generic classification model pre-trained on ImageNet.

At each iteration, we randomly select N images from the data set for training. The selected images are fed to the network to obtain the outputs, with which an episode can be efficiently sampled (all images are passed through the network only once). The gradients of the network parameters are calculated at each time step within the episode and accumulated. Such a process is repeated M_{bz} times (*i.e.*, mini-batch size), and then the parameters are adjusted.

Table 1: Performance comparison with opinion-unaware NR-IQA methods. The four distortions types for generating training data are included, *i.e.*, JP2K, JPEG, WN and BLUR. Then LIVE, CSIQ and TID2013 contain 808 (including references), 600 and 500 testing images, respectively. We use the res-50 after fine-tuning on the same training data as the baseline.

Dataset	LIVE		CSIQ		TID2013	
IQA methods	SROCC	LCC	SROCC	LCC	SROCC	LCC
SSIM (Wang et al. 2004)	0.963	0.950	0.902	0.896	0.867	0.871
VSI (Zhang, Shen, and Li 2014)	0.970	0.921	0.960	0.959	0.949	0.955
QAC (Xue, Zhang, and Mou 2013)	0.869	0.855	0.842	0.874	0.806	0.805
NIQE (Mittal, Soundararajan, and Bovik 2013)	0.920	0.912	0.871	0.875	0.796	0.807
ILNIQE (Zhang, Zhang, and Bovik 2015)	0.918	0.913	0.880	0.905	0.842	0.858
dipIQ (Ma et al. 2017b)	0.946	**0.954**	**0.917**	0.931	0.872	0.888
baseline	0.838	0.734	0.764	0.793	0.754	0.797
RRLRIQA	**0.949**	0.952	0.915	**0.934**	**0.885**	**0.909**

Note that we implement the learning in an efficient way. A standard and intuitive implementation of our method is to pass the candidate images through the network at each time step when sampling an episode. This would require to pass the network for a total of $(N_0 + 1)N_0/2$ times. We reduce the cost to N_0 passes by storing the networks. The episode can be directly built based on the stored network outputs.

Our model allows for the full back-propagation of derivatives through the network (end-to-end training with the standard back-propagation). At time step t, the gradient of parameters is $\nabla_\theta \log \pi_\theta(a_t|s_t)\, \mathrm{v}_t = \nabla_\theta \pi_\theta(a_t|s_t)\mathrm{v}_t/\pi_\theta(a_t|s_t)$. The term $\nabla_\theta \pi_\theta(a_t|s_t)$ can be calculated as

$$\nabla_\theta \pi_\theta(a_t|s_t) = \sum_{a \in \mathcal{A}_t} \frac{\partial\, \pi_\theta(a_t|s_t)}{\partial f_\theta(I_a^t)} \nabla_\theta f_\theta(I_a^t)\ . \qquad (4)$$

Deployment One can see that a good policy requires that the network output $f(\cdot)$ should be indicative of image quality. An image with better perceptual quality should have a higher probability to be selected, *a.k.a.*, larger network output, than the one with poorer quality. Therefore, the quality index during deployment can be implemented by the learned network. It can practically generate implicit quality measures for input images.

Experiments

Experimental Protocol

Databases: Four public IQA databases are used in our experiments, namely LIVE (Sheikh, Sabir, and Bovik 2006), CSIQ (Larson and Chandler 2010), TID2013 (Ponomarenko et al. 2015), and Waterloo Exploration (Ma et al. 2017a). The characteristics of these databases, including the number of source images (references), the number of degraded images and the number of distortion types, are summarized in Table 2. Note that MOSs are not available in Waterloo Exploration, and thus the database can hardly be used for training OA NR-IQA models.

Evaluation Criteria: Following many previous works, we adopt two criteria for the performance evaluation: the Spearman rank order correlation coefficient (SROCC) and the linear correlation coefficient (LCC). SROCC is a measure of the monotonic relationship between the ground-truth scores and predicted ones. LCC measures the linear correlation between the ground-truth and model prediction. Note that the use of LCC requires a nonlinear function, *i.e.*, $\hat{f} = \beta_1(1/2 - 1/(1 + \exp(\beta_2(f - \beta_3)))) + \beta_4 x + \beta_5$, to map raw model predictions to targeting quality scale.

Table 2: Characteristics of benchmark databases for evaluating NR-IQA methods.

IQA Database	Source Images	Distorted Images	Distortion Types	Score Types
LIVE	29	779	5	DMOS
CSIQ	30	886	6	DMOS
TID2013	25	3000	24	MOS
Waterloo	4744	94880	4	-

Comparison with OU Methods

The performance of our approach, compared with two FR-IQA and four representative OU NR-IQA methods, is shown in Table 1. The detailed setting of the inputs in Algorithm 1, as well as the strategy of the training data construction, are given in the following. We denote our method as RRLRIQA (reinforcement recursive list-wise ranking).

Training Set Construction: As previously done in (Ma et al. 2017b), our training dataset contains 700 source images (collected from the Waterloo database). We generate four types of distortions with five levels. They are JP2K and JPEG compression, white Gaussian noise (WN) and Gaussian blur (BLUR). Accordingly, there are a total of 14700 images (700 references and 14000 distorted ones). Each image is assigned with an objective score by performing VSI, a state-of-the-art and efficient FR-IQA measure. During training, we randomly select 10 images (*i.e.*, $N = 10$) and crop a patch from each one (with size of 224×224) for building an episode. The training is performed with no access to MOSs. The reward, indicating whether the chosen image has the best quality, is determined by the assigned objective scores.

Table 3: Performance comparison with opinion-aware NR-IQA methods. We partition the database into a training and a testing subsets. The models are trained on the partitioned training set and then tested on the other one. The procedure is repeated and the median SROCC and LCC values are reported. Note that all types of distortions are included in this experiment. The best two methods are highlighted in boldface.

Dataset	LIVE		CSIQ		TID2013	
IQA methods	SROCC	LCC	SROCC	LCC	SROCC	LCC
SSIM (Wang et al. 2004)	0.948	0.945	0.876	0.861	0.742	0.790
VSI (Zhang, Shen, and Li 2014)	0.952	0.948	0.942	0.928	0.897	0.900
BRISQUE (Mittal, Moorthy, and Bovik 2012)	0.939	0.942	0.756	0.797	0.572	0.651
CORNIA (Ye et al. 2012)	0.942	0.943	0.714	0.781	0.549	0.613
FRIQUEE (Ghadiyaram and Bovik 2016)	0.948	0.962	0.839	0.863	0.669	0.704
BIECON (Kim and Lee 2017)	0.958	0.960	0.815	0.823	0.717	0.762
RankIQA (Liu, van de Weijer, and Bagdanov 2017)	**0.981**	**0.982**	-	-	0.780	0.799
H-IQA (Lin and Wang 2018)	**0.982**	**0.982**	**0.885**	**0.910**	**0.879**	**0.880**
QAC (Xue, Zhang, and Mou 2013)	0.874	0.868	0.486	0.654	0.390	0.495
NIQE (Mittal, Soundararajan, and Bovik 2013)	0.908	0.908	0.627	0.725	0.317	0.426
ILNIQE (Zhang, Zhang, and Bovik 2015)	0.902	0.906	0.822	0.865	0.521	0.648
baseline	0.950	0.954	0.876	0.905	0.712	0.756
RRLRIQA	0.956	0.962	**0.907**	**0.916**	**0.806**	**0.833**

Training Setting: We adopt the adaptive moment estimation optimizer (ADAM) (Kingma and Ba 2015) to train the model. The learning rate α and the batch-size M_{bz} are set as 10^{-5} and 10, respectively. The training process is repeated 9000 times, *i.e.*, $M_b = 9000$. Our MDP is discounted. In experiments, the discount factor γ is set as 0. We will discuss the performance effects of γ later.

Table 1 lists the comparisons (tested on LIVE, CSIQ and TID2013), where the method with the best performance is highlighted in boldface. All the results are reported by ourselves with realized codes or models. To demonstrate the effectiveness of the presented reinforcement list-wise method, we use the res-50 after fine-tuning on the same training data (*i.e.*, the same 14700 training images) as the baseline. The ground-truth for fine-tuning is the objective scores derived by VSI. During testing, we randomly sample 10 patches with size of 224×224 for each image. The quality estimation is the average over the network outputs of the selected patches (both for the baseline and RRLRIQA).

One can see that RRLRIQA achieves state-of-the-art performance on these three databases. It delivers higher results than other competitors on TID2013, and is comparable to dipIQ on LIVE and CSIQ. Moreover, although RRLRIQA and baseline are both based on res-50, we observe consistent performance gains of our RRLRIQA on the three databases over the baseline. This may be because the objective scores are capable of indicating whether an image is of higher quality than others, but are not accurate enough and nosier for training OA NR-IQA methods.

Comparison with OA Methods

The second part of the experiments is to compare RRLRIQA with OA NR-IQA methods by training and testing on MOS-aware databases. The detailed settings are as follows.

Training and Testing Settings: The experiment is conducted on the entire LIVE, CSIQ, and TID2013 databases. We divide each database into two subsets: 80% for training and 20% for testing. Specifically, the distorted images are grouped according to their references. On LIVE, CSIQ, and TID2013, we randomly select 23, 24 and 20 groups for training respectively, while retaining the other groups for testing. This is to ensure that no image contents used in testing have been seen by the models during training. The training-testing partition is repeated 20 times and the median is reported to reduce the influence of random selection.

MOSs are used for training OA methods for comparison, including BRISQUE, CORNIA, FRIQUEE, BIECON, RankIQA and H-IQA. The results of first four methods are given by (Kim et al. 2017), and those of the other two are reported by original authors. Three OU methods are included, namely QAC, NIQE and ILNIQE. Their results, taken from (Zhang, Zhang, and Bovik 2015), are reported on the testing subsets. We use the res-50 fine-tuned on the training subsets as the baseline. MOSs are employed as the ground-truth for fine-tuning. Our RRLRIQA is also trained on the partitioned training subsets. The training configurations are the same as in the last subsection. The reward is still given by the objective scores for a fair comparison with OU methods.

Table 3 shows the results, from which we have several observations. First, our model outperforms the three OU competitors by a large margin, *e.g.*, about 28% SROCC improvement than ILNIQE on TID2013. Second, RRLRIQA achieves state-of-the-art performance in comparison with recent OA NR-IQA methods. It reaches the best performance on CSIQ, and the second best on TID2013 even without access to MOS for training. Finally, our model works consistently better than the baseline. RRLRIQA achieves about 3% and 9% higher SROCC than the baseline on CSIQ and TID2013, respectively. This may be because learning from

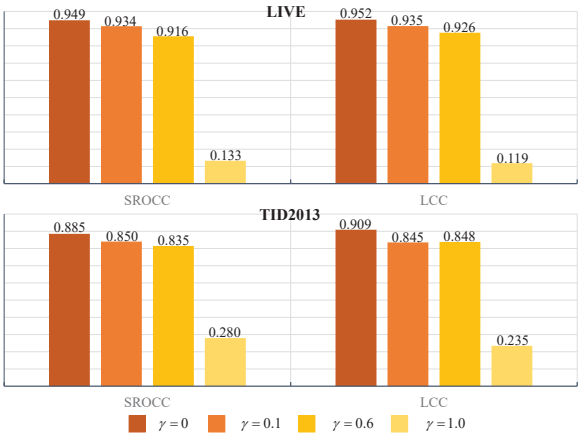

Figure 2: SROCC and LCC with respect to the discount factor γ. The models are trained on the collected training data and tested on LIVE and TID2013.

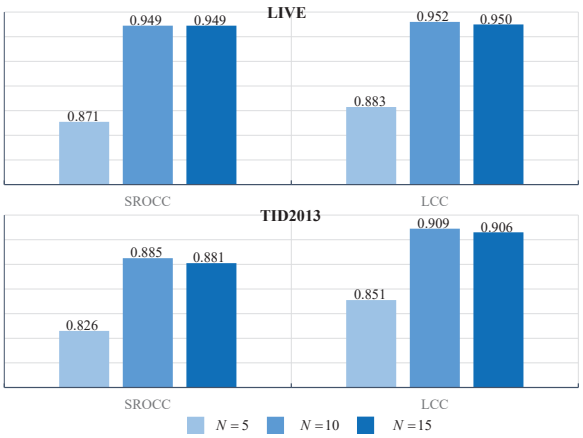

Figure 3: SROCC and LCC with respect to the size of training image list, *i.e.*, N. The models are trained on the collected training data and tested on LIVE and TID2013.

rankings can be considered as a data augmentation strategy since various image lists can be generated from a database for training our model. It may help to reduce the overfitting.

Discussion on Performance Issues

In this subsection, we discuss two performance issues about the proposed RRLRIQA, namely the effects of the discount factor γ and the size of an image list in learning.

Discount Factor The effect of the discount factor γ is investigated in this part. We compare four models with γ being set as 0, 0.1, 0.6 and 1.0 during training. These four models share the same training data (the collected 700 images and their distorted versions) and training configurations. LIVE and TID2013 are used for testing.

One can see from Figure 2 that the performance decreases if we increase the γ value. It is probably because the long-term returns are not capable of indicating whether the best-quality image is selected at each time step (especially when $\gamma = 1$), which leads to the performance degradation.

Size of Training Image List In our method, an image list with size of N is used as an instance (*i.e.*, for building an episode) in learning. We then conduct an experiment to explore the effect of this training parameter. Figure 3 shows the SROCC and LCC metrics with N being set to be 5, 10 and 15. The three models are all trained on the collected training data and tested on LIVE and TID2013.

It can be seen from Figure 3 that the performance benefits from a relatively large list size. The models with N being set as 10 and 15 significantly achieve better results than the one with $N = 5$. Moreover, we also notice that the performance of the model trained with a large list size is generally more stable. To show that, we test the model every 300 training iterations (the training is performed 9000 iterations). The standard deviations of the obtained SROCCs are $0.0119/0.0080$ ($N = 10/15$) on LIVE, and $0.0183/0.0124$ ($N = 10/15$) on TID2013. But training such a model will cost more time. In our experiments, N is set to be 10 as a trade-off between the performance and training efficiency.

Conclusion

In this paper, we develop an effective OU NR-IQA method with reinforcement recursive list-wise ranking. We formulate the NR-IQA as a recursive ranking problem and use image lists as learning instances. The recursive ranking process is separated as a sequence of nested sub-problems and can be modeled as an MDP during training. The model parameters can be effectively trained by the policy gradient algorithm. Experimental results show that our approach achieves state-of-the-art performance compared with current OA and OU NR-IQA methods.

We believe that our method can give some useful insights to the NR-IQA community, not only the reinforcement list-wise ranking model but also the way to develop OU methods. Our method can be further extended. For example, we plan to explore the ranking uncertainty (different people may have different opinions about the quality ordering of images) in our future work. Moreover, since reinforcement learning allows for a step-wise training, designing different training strategies at different time steps for a potential performance improvement is also very interesting.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grants 91646207, 61773377, and 61573352, and the Beijing Natural Science Foundation under Grants 4162064 and L172053.

References

Bovik, A. C. 2013. Automatic prediction of perceptual image and video quality. *Proceedings of the IEEE* 101(9):2008–2024.

Gao, F.; Tao, D.; Gao, X.; and Li, X. 2015. Learning to rank for blind image quality assessment. *IEEE Transactions on Neural Networks and Learning Systems* 26(10):2275–2290.

Gastaldo, P.; Zunino, R.; and Redi, J. 2013. Supporting visual quality assessment with machine learning. *EURASIP Journal on Image and Video Processing* 2013(1):1–15.

Ghadiyaram, D., and Bovik, A. C. 2016. Perceptual quality prediction on authentically distorted images using a bag of features approach. *arXiv preprint arXiv:1609.04757*.

Gu, J.; Meng, G.; Redi, J. A.; Xiang, S.; and Pan, C. 2018. Blind image quality assessment via vector regression and object oriented pooling. *IEEE Transactions on Multimedia* 20(5):1140–1153.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 770–778.

Kang, L.; Ye, P.; Li, Y.; and Doermann, D. S. 2014. Convolutional neural networks for no-reference image quality assessment. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 1733–1740.

Keelan, B. 2002. *Handbook of image quality: characterization and prediction*. CRC Press.

Kim, J., and Lee, S. 2017. Fully deep blind image quality predictor. *IEEE Journal of Selected Topics in Signal Processing* 11(1):206–220.

Kim, J.; Zeng, H.; Ghadiyaram, D.; Lee, S.; Zhang, L.; and Bovik, A. C. 2017. Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment. *IEEE Signal Processing Magazine* 34(6):130–141.

Kingma, D. P., and Ba, J. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.

Larson, E. C., and Chandler, D. M. 2010. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging* 19(1):011006–011006.

Lin, K., and Wang, G. 2018. Hallucinated-iqa: No-reference image quality assessment via adversarial learning. *CoRR* abs/1804.01681.

Liu, X.; van de Weijer, J.; and Bagdanov, A. D. 2017. Rankiqa: Learning from rankings for no-reference image quality assessment. In *IEEE International Conference on Computer Vision*, 1040–1049.

Liu, T. 2009. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval* 3(3):225–331.

Lu, X.; Lin, Z.; Shen, X.; Mech, R.; and Wang, J. Z. 2015. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. In *IEEE International Conference on Computer Vision*, 990–998.

Ma, L.; Li, S.; Zhang, F.; and Ngan, K. N. 2011. Reduced-reference image quality assessment using reorganized dct-based image representation. *IEEE Transactions on Multimedia* 13(4):824–829.

Ma, L.; Xu, L.; Zhang, Y.; Yan, Y.; and Ngan, K. N. 2016. No-reference retargeted image quality assessment based on pairwise rank learning. *IEEE Transactions on Multimedia* 18(11):2228–2237.

Ma, K.; Duanmu, Z.; Wu, Q.; Wang, Z.; Yong, H.; Li, H.; and Zhang, L. 2017a. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing* 26(2):1004–1016.

Ma, K.; Liu, W.; Liu, T.; Wang, Z.; and Tao, D. 2017b. dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs. *IEEE Transactions on Image Processing* 26(8):3951–3964.

Mittal, A.; Moorthy, A. K.; and Bovik, A. C. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing* 21(12):4695–4708.

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2013. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters* 20(3):209–212.

Ponomarenko, N. N.; Jin, L.; Ieremeiev, O.; Lukin, V. V.; Egiazarian, K. O.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, F.; and Kuo, C. J. 2015. Image database TID2013: peculiarities, results and perspectives. *Signal Processing: Image Communication* 30:57–77.

Ren, H.; Chen, D.; and Wang, Y. 2018. RAN4IQA: restorative adversarial nets for no-reference image quality assessment. In *AAAI Conference on Artificial Intelligence*.

Sheikh, H. R.; Sabir, M. F.; and Bovik, A. C. 2006. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing* 15(11):3440–3451.

Tang, H.; Joshi, N.; and Kapoor, A. 2014. Blind image quality assessment using semi-supervised rectifier networks. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2877–2884.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4):600–612.

Xue, W.; Zhang, L.; and Mou, X. 2013. Learning without human scores for blind image quality assessment. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 995–1002.

Ye, P.; Kumar, J.; Kang, L.; and Doermann, D. S. 2012. Unsupervised feature learning framework for no-reference image quality assessment. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 1098–1105.

Ye, P.; Kumar, J.; and Doermann, D. S. 2014. Beyond human opinion scores: Blind image quality assessment based on synthetic scores. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 4241–4248.

Zhang, L.; Shen, Y.; and Li, H. 2014. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing* 23(10):4270–4281.

Zhang, L.; Zhang, L.; and Bovik, A. C. 2015. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing* 24(8):2579–2591.