

Quality Assessment for Comparing Image Enhancement Algorithms

Zhengying Chen, Tingting Jiang and Yonghong Tian*

School of Electronics Engineering and Computer Science, Peking University
{cathychenchn, ttjiang, yhtian}@pku.edu.cn

Abstract

As the image enhancement algorithms developed in recent years, how to compare the performances of different image enhancement algorithms becomes a novel task. In this paper, we propose a framework to do quality assessment for comparing image enhancement algorithms. Not like traditional image quality assessment approaches, we focus on the relative quality ranking between enhanced images rather than giving an absolute quality score for a single enhanced image. We construct a dataset which contains source images in bad visibility and their enhanced images processed by different enhancement algorithms, and then do subjective assessment in a pair-wise way to get the relative ranking of these enhanced images. A rank function is trained to fit the subjective assessment results, and can be used to predict ranks of new enhanced images which indicate the relative quality of enhancement algorithms. The experimental results show that our proposed approach statistically outperforms state-of-the-art general-purpose NR-IQA algorithms.

Keywords-enhancement algorithm; quality assessment;

I. Introduction

Acquiring clear images in bad visibility scenes plays an important role in many fields, which pushes forward the development of enhancement algorithms these years. Since the single image haze removal algorithm [1] using dark channel prior was proposed by He *et al.* in 2009, many new enhancement algorithms were designed for other tasks, such as low light scene enhancement [2], [3], underwater enhancement [4]–[6], etc. However, when we turn to compare the performances of these enhancement algorithms, we found that there is neither a public criterion at present nor a public dataset for reference.

*Corresponding to: Yonghong Tian, yhtian@pku.edu.cn

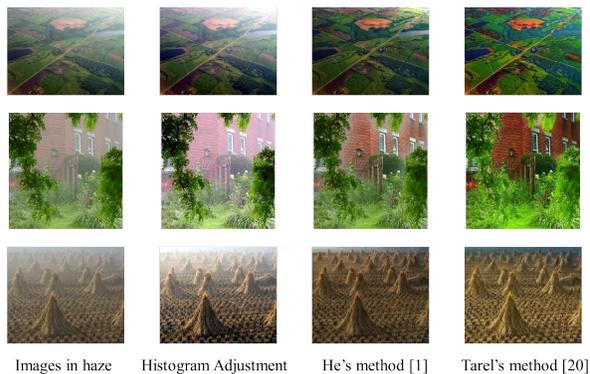


Figure 1. The first column contains source images in haze, and the other three columns contain dehazed images processed by three different algorithms. How can we compare the performances of these algorithms?

The most common existing solution is manually listing several images in bad visibility as well as their corresponding enhanced images which are processed by different algorithms, and then comparing them subjectively as shown in Figure 1. Though it complies with intuition, there are two problems. The first problem is that the amount of listed images is limited, so we can only tell which algorithm performs best in these listed cases, while performances in other cases are still unknown. The other problem is the lack of quantitative quality assessment approach for enhanced images. Moreover, subjective quality assessment is time consuming and expensive, and the problem of objective quality assessment for enhanced images itself is quite challenging.

Another possible solution is the traditional image quality assessment (IQA) technique, which mainly focuses on compressed images and cannot be applied to this problem. Usually they can be classified into three categories: full-reference (FR) IQA, reduced-reference (RR) IQA and no-reference (NR) IQA. As to enhanced images, information of reference images is rarely available, for making the definition of “clear image” is very difficult when take the

whether condition, light condition and other aspects into consideration. Most of the existing NR-IQA approaches are limited to one or several specific distortion types [7], [8] which are commonly arisen in the compressing process, such as blur, blocking artifacts or ringing. However, when come to enhanced images, unfortunately these approaches cannot be simply adopted, for the distortion types of enhanced images are quite different from those of compressed images, and they are not easy to define, either. In recent years, some general purpose NR-IQA approaches are proposed, which can make the assessment without the prior knowledge of distortion types. One of the common ideas is extracting manually designed features based on the nature scene statistics (NSS) hypothesis [9]–[11], and another trend which is based on extracting features through training [12]–[15] is getting more popular. Most of these approaches regarded the IQA task as a classification problem, i.e. images should be classified into several levels according to their qualities and that’s why SVMs were used for regression in these approaches. However, the quality of an enhanced image is strongly related with its content, so images processed from different source images usually don’t share the same classification criterion. Consequently, simply adopting SVM/SVR model may don’t work and finding a more appropriate model becomes crucial.

By all accounts, given several candidate enhancement algorithms and their enhanced images, how to compare their performances on the enhanced images and predict their performance on new cases becomes a novel task. Furthermore, when a new enhancement algorithm is proposed, how to evaluate its performance compared with known enhancement algorithms is also a new problem. In this paper, we address these two issues by proposing a new framework for comparing image enhancement algorithms. The most significant difference between our approach and the traditional IQA algorithms is that we focus on the relative quality ranking between enhanced images rather than giving an absolute quality score for a single enhanced image. To achieve this goal, we first conduct subjective quality assessment for the given enhanced images by applying a pair-wise labeling method and then construct a ground truth dataset for objective quality assessment. Based on the dataset, we learn a rank function to fit the subjective assessment results of the training enhanced images. When some new cases come, relative quality ranking between the new enhanced images can be predicted according to the learned model. The relative qualities of these candidate enhancement algorithms can be inferred from the qualities of their enhanced images. Similarly, when a new enhancement algorithm appears, the relative rankings of its enhanced images compared with the known algorithms can be predicted through the learned model. Thus, the comparison of given enhancement algorithms and the new

one can be performed according to the predicted results. To verify this proposed framework, we have conducted experiments on three most popular scenes, including haze, underwater and low light. The experimental results show that our proposed approach statistically outperforms state-of-the-art general-purpose NR-IQA algorithms.

The rest of this paper is organized as follows. Section 2 reviews previous work on quality criterion used in enhancement tasks and image quality assessment approaches that is most relevant to our work. In Section 3, a more formal definition of this task is described, and the framework of our proposed approach is introduced. The whole construction procedure of our dataset is illustrated in Section 4. Section 5 describes our proposed approach and the experimental results are shown in Section 6. Finally, Section 7 concludes with a brief discuss of our future work.

II. Related Work

A. Enhancement Algorithms and their Criteria

The quality criteria used to assess enhancement algorithms at present can mainly be classified into two approaches. The most common one is manually listing several enhanced images which are all corresponding to a same source image and then comparing their quality subjectively. Ringing artifacts, color shifting or other aspects are taken into consideration by experience. Some authors [16], [17] also show the depth map or other processing results. However, this method usually covers very limited images, so it’s easy to argue that the proposed algorithm maybe only work on the listed images.

The other solution is doing task-driven assessment. Tarel *et al.* [18] used a car detection technique on haze images and their dehazed ones, and thus showed the efficiency of their dehazing algorithm according to the detection results. Ancuti *et al.* [5] applied feature matching and image segmentation techniques to assess their underwater enhancement algorithm in a similar way. The most apparent shortcoming is that the tasks used for assessing qualities are not well solved itself, and the techniques like car detection or image segmentation may only work in limited scenes. Thus, this solution cannot guarantee the assessment accuracy nor can it be expanded easily.

There are also some approaches that are inspired by the traditional IQA techniques. Ancuti *et al.* [19] take an assessment criterion based on measuring contrast. However, these methods cannot describe all aspects of enhanced images, for image enhancement is not a simple procedure of increasing contrast. Above all, there’s still a lack of public criterion for enhancement algorithms and their enhanced images.

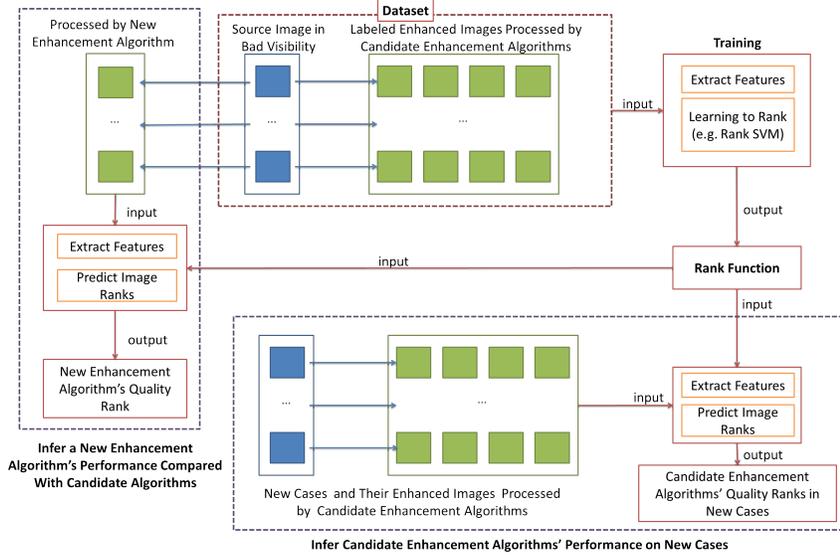


Figure 2. The framework of our proposed approach.

B. NR/RR Image Quality Assessment

Generally speaking, current no-reference (NR) IQA techniques mainly follow two trends, and are almost general-purpose. One is based on Natural Scene Statistics (NSS), which relies on the hypothesis that natural scenes possess certain statistical properties and the presence of distortion will affect these properties. Researchers usually explore statistics in the wavelet domain [10] or DCT domain [9], [11], and then design appropriate features that are highly correlated with image quality. This method acquires deep domain knowledge and rich experience, and it is extensively studied these years.

The latter trend is based on machine learning techniques, including unsupervised feature learning [13], clustering [15], codebook [14], etc. These training based approaches rely on large number of candidate features, but what features are best for the assessment problem may be not clear. At present, more and more researchers choose to follow the latter trend which may provide a better expansibility.

There are also some approaches based on information theories, and do the assessment in terms of entropies. Soundararajan *et al.* [20] proposed a reduced-reference IQA indice which is based on the entropic differencing. Gabara *et al.* [21] found that no-reference IQA could be done by measuring the anisotropy of images.

Most of these approaches mentioned above are designed for compressed images, and the features may be not able to describe the enhanced images well. What's more, these approaches formulate the IQA problem as a classification task. But when come to compare enhancement algorithms' quality, the regression model should focus more on relative

rankings between different enhancement algorithms.

Nicolas Hautiere *et al.* [22] proposed a blind contrast enhancement assessment approach in 2008, which is mainly based on the atmospheric luminance model and the concept of visibility level which is usually used in lighting engineering. Though it is claimed as a blind assessment approach, the original image which is before the enhancement is used to calculate the indicator of visibility enhancement. The indicator is used to evaluate the rate of new edges after enhancement and the percentage of saturated pixels after enhancement. Unfortunately, this approach only provides three indices of evaluation rather than a generalized assessment result, which makes automatic objective quality assessment difficult.

III. Problem and Proposed Approach

A. Problem Definition

A more formal definition of the tasks discussed in this paper will be given in this section. First of all, some mathematical symbols used in the rest of this paper are introduced below.

- A_j : the j^{th} candidate enhancement algorithm
- n_A : the number of candidate algorithms
- S_i : the i^{th} source image in bad visibility
- n_S : the number of source images
- $I_{i,j}$: the enhanced image of S_i processed by A_j
- $R_{i,j}$: the rank of $I_{i,j}$ among those enhanced images which are corresponding to S_i , *i.e.* $\langle R_{i,1}, R_{i,2}, \dots, R_{i,n_A} \rangle$ is a permutation of $1, 2, \dots, n_A$, and if $I_{i,a}$ has a higher quality than $I_{i,b}$, then $R_{i,a} < R_{i,b}$
- S_i^* : the i^{th} new source image in bad visibility

Algorithm	Scene	Our Adjustment
Tan's method [16]	Haze	Not applied the optimization of MRFs
He's method [1]	Haze	Used the guided filter [23] instead of soft-matting
Tarel's method [18]	Haze	/
Ancuti's method in 2010 [19]	Haze	Used the LUV chromaticity definition
white balance [24]	Underwater	/
Li's method [4]	Underwater	Used the guided filter [23] instead of soft-matting
Wen's method [6]	Underwater	/
Ancuti's method in 2012 [5]	Underwater	/
classic histogram equalization	Low Light	/
classic Gama correlation	Low Light	/
Cai's method [2]	Low Light	/
Dong's method [3]	Low Light	Used the guided filter [23] instead of soft-matting
classic histogram adjustment	Haze, Underwater, Low Light	/

Table I. Applied enhancement algorithms.

n_S^* : the number of new source images

$I_{i,j}^*$: the new enhanced image of S_i^* processed by A_j

$R_{i,j}^*$: the rank of $I_{i,j}^*$ among those new enhanced images which are corresponding to S_i^*

A^* : a novel enhancement algorithm that needs assessment

I_i^* : the enhanced image of S_i processed by A^*

R_i^* : the rank of I_i^* among those enhanced images which are all corresponding to S_i

Given n_A candidate enhancement algorithms $\{A_j\}$, n_S source images $\{S_i\}$ in bad visibility and their enhanced images $\{I_{i,j}\}$ processed by candidate enhancement algorithms, the quality assessment task for comparing image enhancement algorithms can be defined as two parts below.

The first part is to compare given candidate enhancement algorithms' performances on new cases, *i.e.* given n_S^* new source images $\{S_i^*\}$ in bad visibility and their enhanced images $\{I_{i,j}^*\}$. The goal is to predict the quality ranks $\{R_{i,j}^*\}$ of these new enhanced images so as to compare the given candidate enhancement algorithms' performances on new cases.

The second part is to evaluate a new enhancement algorithm's quality compared with given candidate enhancement algorithms, *i.e.* given a new enhancement algorithm A^* and new enhanced images I_i^* , the goal is to predict the quality ranks $\{R_i^*\}$ of these new enhanced images with respect to known enhancement algorithms so as to compare the performances of given candidate enhancement algorithms and that of the new one.

B. Proposed Approach

The framework of our proposed approach is illustrated in Figure 2, including a dataset, a training step and an inference step. Our framework is quite simple but effective, and can be capable to assess enhancement algorithms in different scenes.

First of all, we design a subjective quality assessment experiment for given enhanced images in a pair-wise way, in order to obtain the quality ranks of these enhanced images. In this way, the dataset which consists of the

source images, their enhanced images and the subjective quality ranks is constructed. Then in the training step, a rank function can be learned on this dataset by adopting some learning to rank tools like rank SVM [25]. The rank function is expected to fit the subjective assessment results well, and give clues of quality rankings between given enhanced images.

When some new source images come, quality ranks of their enhanced images are predicted by the pre-trained rank function. Then the quality rank of a given candidate enhancement algorithm A_j can be estimated as the average quality rank of its corresponding enhanced images. When a new enhancement algorithm comes, quality ranks of its enhanced images can be predicted by the pre-trained rank function, too. The quality rank of this new algorithm can be evaluated in a similar way, and the variance of the ranks also indicates the robustness of this algorithm.

IV. Dataset

A. Data Collection

We picked three popular topics in image enhancement fields for instance, including image dehazing, underwater image enhancement and low light scene enhancement. For each enhancement task, we collected 100 source images which are all in bad visibility from related literature and image search engines. These 300 source images are resized so that the maximum of height and width is 400 pixels.¹

In order to generate enhanced images, we applied several enhancement algorithms for each source image. For each scene, we chose 5 enhancement algorithms, and Table I gives a detailed view. Some of these algorithms may be too time consuming, so we only applied some part of the algorithm and the enhanced images processed in this way may seem not as good as the ones processed by the original version. The enhanced images are only used for providing enough samples for quality assessment. Hence

¹This dataset can be downloaded at <http://mlg.idm.pku.edu.cn/resources/dataset.html>

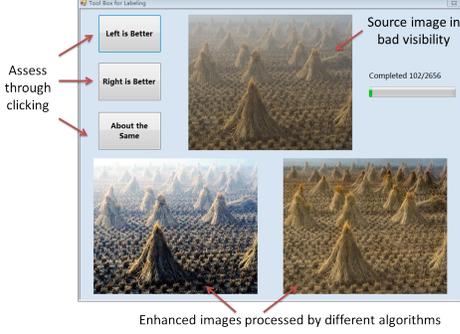


Figure 3. The interface of labeling.

it does not matter whether we implemented the original version of the enhancement algorithm or a simpler version.

B. Labeling and Data Cleaning

We developed a labeling tool on .NET framework with C#, which includes a annotator information collecting interface and a simple tutorial. We recruited 30 college students to do the labeling work, involving 14 males and 16 females. Most annotators have a basic knowledge of image processing, and 3 of them are doing related researches currently. Only 2 annotators said that they had never touched the field of image processing.

To illustrate the usage of the labeling tool as well as the common aspects the subjective assessment usually concerns, our labeling tool gives a simple tutorial at the beginning of the labeling work. The annotators are required to give subjective assessment on pairs of enhancement images. Formally speaking, for a source image S_i in bad visibility, we have already got its n enhanced images $\{I_{i,j}\}$. Our labeling tool will show all pairs of these enhanced images $\langle I_{i,a}, I_{i,b} \rangle$ ($1 \leq a < b \leq n_A$) one by one, and will give the source image S_i for annotators to refer to as well. The annotators gave ‘a is better than b’, ‘a is worse than b’ or ‘a and b are about the same’ assessment through clicking buttons, and the results were recorded automatically. Figure 3 shows the labeling interface. After labeling all these pairs corresponding to S_i , next source image and its enhanced images would be assessed in the same way until all images were got labeled.

After obtained 30 annotators’ subjective assessment results, we did data cleaning with these original labeling records. We implemented an outlier detection method proposed by Xu *et al.* [26] that relies on a Huber-LASSO model to eliminate unreliable pair-wise records, and then summarized the 30 annotators’ results into ranks $\{R_{i,j}\}$. $\langle R_{i,1}, R_{i,2}, \dots, R_{i,n_A} \rangle$ is a permutation of 1 to n_A , and if $I_{i,a}$ has a higher quality than $I_{i,b}$, then $R_{i,a} < R_{i,b}$. Figure 4 shows some samples of images’ ranks after labeling.

V. Training Model and Inference

A. Training

The first part of the training step is extracting appropriate features from each enhanced image. The features should be able to capture the characteristics of enhanced images. Here we denote the extracted features as $f_{i,j}$, where $f_{i,j}$ is a p-dimension feature vector. The second part of the training step is to find a rank function which fits the labeling results best. The rank function $F(x) = xw^T$ can be learned by learning to rank tools like rank SVM. x is a p-dimension feature vector, and w is a p-dimension weight vector.

For each pair of enhanced images $I_{i,a}$ and $I_{i,b}$, the desired rank function should satisfy the constraints below. With the introduction of non-negative slack variables, this problem can be solved with an approximate solution which is similar to SVM classification.

$$f_{i,a}w^T - f_{i,b}w^T > 0, \forall R_{i,a} > R_{i,b} \quad (1)$$

$$f_{i,a}w^T - f_{i,b}w^T < 0, \forall R_{i,a} < R_{i,b} \quad (2)$$

$$f_{i,a}w^T - f_{i,b}w^T = 0, \forall R_{i,a} = R_{i,b} \quad (3)$$

B. Inference

When turn to assess the candidate enhancement algorithms’ performances on new cases, the new enhanced images $\{I_{i,j}^*\}$ can be obtained by applying candidate algorithms $\{A_j\}$. With the pre-trained rank function $F(x)$, quality ranks $\{R_{i,j}^*\}$ can be calculated after extracting the same p-dimension features of these new enhanced images. The quality ranks of these candidate enhancement algorithms on new cases can be inferred from the average rank of their corresponding new enhanced images:

$$Quality(A_j) = \frac{1}{n_S^*} \sum_{i=1}^{n_S^*} R_{i,j}^* \quad (4)$$

What’s more, the robustness of these enhancement algorithms can be inferred from the variance of $\{R_{i,j}^*\}$. A small variance indicates a more robust algorithm, and vice versa.

Similarly, when given a new enhancement algorithm A^* , we can obtain its enhanced images $\{I_i^*\}$ according to the source images $\{S_i\}$ in the dataset. After extracting the same p-dimension features from these enhanced images, the quality ranks $\{R_i^*\}$ of these enhanced images can be inferred by the pre-trained rank function $F(x)$. The quality rank of this new enhancement algorithm with respect to known algorithms can be represented as the average rank of its enhanced images, too.

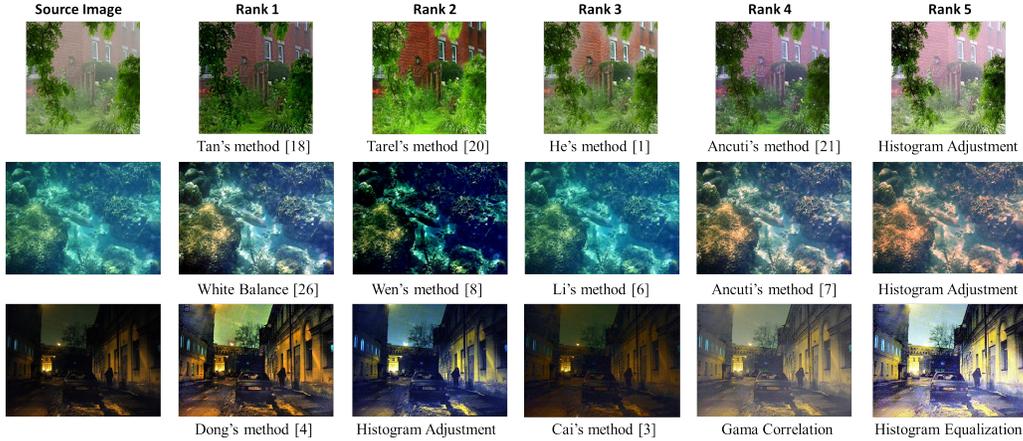


Figure 4. Some samples of images after labeling. The first column contains some sample of source images in three scenes. The other columns are enhanced images processed by different algorithms, and are listed according to the subjective quality assessment results. The quality of image decreases from left to right.

VI. Experiment

A. Protocol

Since there is no image quality assessment dataset for enhanced images, we evaluated our approach on the dataset described in Section 4. We use a concatenation of the GIST descriptor [27] and the color motion [28] as our image features (*i.e.* a combination feature of 521 dimensions), and rank SVM [25] was adopted to do the regression. The distortion types that enhanced images may contain are not known yet, so we choose four general-purpose NR-IQA approaches [9]–[11], [29] as baselines.

B. Evaluate the Performance on New Cases

First of all, we try to evaluate candidate enhancement algorithms' performances on new cases. We randomly separated the dataset into 2 disjoint sets with a ratio of 8:2, *i.e.* 80% of the dataset is used as the training set, and the rest 20% is used for testing. And then we calculated the SROCCs (Spearman Rank Order Correlation Coefficient) of the predict ranks in the test set. The baseline NR-IQA algorithms trained their models on the same dataset, and the LibSVM [30] was used to do the regression. We repeated the experiment for 1000 times and take the average SROCC as the final result. Table II shows the details. Figure 5, Figure 6 and Figure 7 show the boxplot of the 1000 iterations' distribution in 3 scenes respectively.

As we can see, our approach gets the highest SROCC in all three scenes and looks more robust than other baseline NR-IQA algorithms according to the boxplots.

C. Evaluate the Performance of a New Algorithm

We also tested our framework to evaluate the performance of a new algorithm. In order to do the experiment

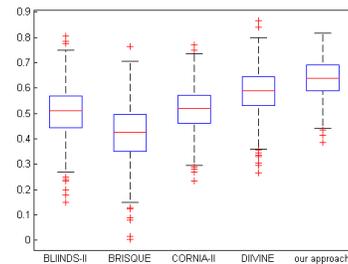


Figure 5. The boxplot of 1000 iterations in haze scene.

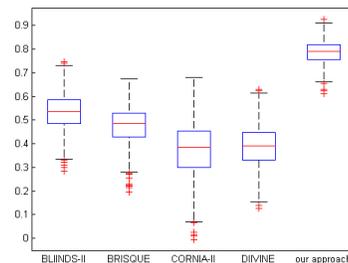


Figure 6. The boxplot of 1000 iterations in underwater scene.

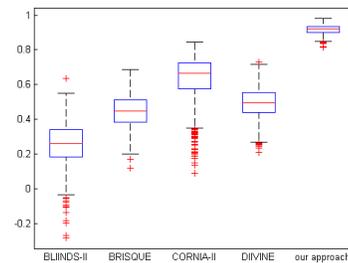


Figure 7. The boxplot of 1000 iterations in low light scene.

more efficiently, the first step we did is to determine the number of candidate algorithms which are used to be compared with the new algorithm, and what these algorithms

SROCC	Haze	Underwater	Low Light	Over All
BLINDS-II [11]	0.5055	0.5320	0.2572	0.4316
BRISQUE [9]	0.4179	0.4781	0.4461	0.4474
CORNIA-II [29]	0.5153	0.3716	0.6318	0.5062
DIIVINE [10]	0.5835	0.3894	0.4948	0.4893
our approach	0.6302	0.7858	0.9155	0.7792

Table II. Average SROCC with 1000 iterations of experiments on the dataset described in Section 4.

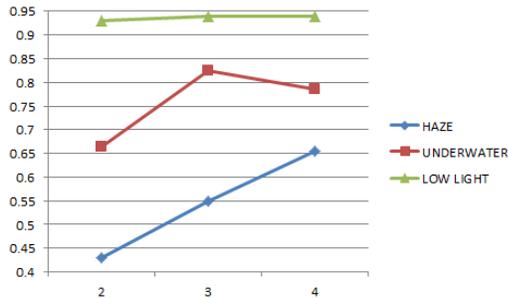


Figure 8. The SROCC curve when picking different number of candidate algorithms.

Haze	Underwater	Low Light	Over All
1.5800	0.8400	0.2600	0.8933

Table III. MSE of predicted ranks and ground-truth ranks.

are. To solve this problem, we randomly separated the dataset into three disjoint sets. 60% of the dataset was used as a training set, 20% was used as a validation set and the rest 20% was used as a testing set. For each enhancement task, we tried every subset of candidate algorithms. For each subset, we only used the enhanced images processed by this subset in the training set to obtain the rank function. Then the ranks of the enhanced images processed by the subset in the validation set was predicted through the rank function, and an average SROCC of these ranks can be computed. For each enhancement task, we chose the subset on which the highest SROCC was obtained as the most appropriate set of candidate algorithms. According to the experiment result, 4 candidate algorithms are needed in haze scene, and 3 are needed in the other two scenes. In order to figure out the trend of SROCC when the number of algorithms in the subset changes, for those subsets which contain the same number of algorithms, we recorded the highest SROCC they could achieve. Figure 8 shows the curves in three scenes respectively.

After the candidate algorithms were determined, we trained the rank function by using enhanced images processed by candidate algorithms in the training set, and then predicted the ranks of the other algorithms in the testing set. The MSE (Mean Squared Error) of predicted ranks and ground-truth ranks indicated the prediction accuracy of this framework. Table III shows the details.

D. Discussion of Results

Generally speaking, our approach is statistically better than baselines, and gets a high prediction accuracy in the low light scene. However, the experiment results in haze scene are not so good as that in other two scenes. One possible reason maybe is that the haze scene is more complicated than other two scenes, so the features we chose cannot capture the characteristics of dehazed images well. According to Figure 8, in the haze scene, the highest SROCC keeps growing when the number of candidate algorithms increases. But in the other two scenes, the highest SROCC first grows and then drops when the number of candidate algorithms increases. It indicates that maybe more than 4 candidate algorithms are needed in the haze scene to predict a new enhancement algorithm's rank, so the MSE of haze scene is much larger than that of the other two scenes.

VII. Conclusion

In this paper, a new quality assessment task for enhancement algorithms was proposed. To solve this task, a general-purpose framework which formulates the task as a learning to rank problem was given. A quality assessment dataset for enhancement algorithms was constructed, and was used to do experiments based on our framework. The experiment results showed that the new formulation of enhancement algorithm quality assessment makes sense and can help to evaluate a new enhancement algorithm. In the future work, we will try to find better features and test the framework in more scenes.

Acknowledgement

This work is partially supported by grants from the National Natural Science Foundation of China (61035001, 61390514, 61121002, 61103087).

References

- [1] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1956–1963. 1, 4
- [2] L. Cai and J. Qian, "Night color image enhancement using fuzzy set," in *Image and Signal Processing, 2009. CISP'09. 2nd International Congress on*. IEEE, 2009, pp. 1–4. 1, 4

- [3] X. Dong, G. Wang, Y. Pang, W. Li, J. Wen, W. Meng, and Y. Lu, "Fast efficient algorithm for enhancement of low lighting video," in *Multimedia and Expo (ICME), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1–6. 1, 4
- [4] L. Chao and M. Wang, "Removal of water scattering," in *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on*, vol. 2. IEEE, 2010, pp. V2–35. 1, 4
- [5] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 81–88. 1, 2, 4
- [6] H. Wen, Y. Tian, T. Huang, and W. Gao, "Single underwater image enhancement with a new optical model," in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 753–756. 1, 4
- [7] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: Jpeg2000," *Image Processing, IEEE Transactions on*, vol. 14, no. 11, pp. 1918–1927, 2005. 2
- [8] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *Signal Processing Letters, IEEE*, vol. 17, no. 5, pp. 513–516, 2010. 2
- [9] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *Signal Processing Letters, IEEE*, vol. 17, no. 6, pp. 583–586, 2010. 2, 3, 6, 7
- [10] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *Image Processing, IEEE Transactions on*, vol. 20, no. 12, pp. 3350–3364, 2011. 2, 3, 6, 7
- [11] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3339–3352, 2012. 2, 3, 6, 7
- [12] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 305–312. 2
- [13] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1098–1105. 2, 3
- [14] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *Image Processing, IEEE Transactions on*, vol. 21, no. 7, pp. 3129–3138, 2012. 2, 3
- [15] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013. 2, 3
- [16] R. T. Tan, "Visibility in bad weather from a single image," in *Computer Vision and Pattern Recognition, 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8. 2, 4
- [17] R. Fattal, "Single image dehazing," in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3. ACM, 2008, p. 72. 2
- [18] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 2201–2208. 2, 4
- [19] C. O. Ancuti, C. Ancuti, and P. Bekaert, "Effective single image dehazing by fusion," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 3541–3544. 2, 4
- [20] R. Soundararajan and A. C. Bovik, "Rred indices: reduced reference entropic differencing for image quality assessment," *Image Processing, IEEE Transactions on*, vol. 21, no. 2, pp. 517–526, 2012. 3
- [21] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *JOSA A*, vol. 24, no. 12, pp. B42–B51, 2007. 3
- [22] N. Hautière, J.-P. Tarel, D. Aubert, É. Dumont *et al.*, "Blind contrast enhancement assessment by gradient ratioing at visible edges," *Image Analysis & Stereology Journal*, vol. 27, no. 2, pp. 87–95, 2008. 3
- [23] K. He, J. Sun, and X. Tang, "Guided image filtering," in *European Conference on Computer Vision, 2010*. Springer, 2010, pp. 1–14. 4
- [24] M. Ebner, *Color constancy*. Wiley.com, 2007, vol. 6. 4
- [25] T. Joachims, "Training linear svms in linear time," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006, pp. 217–226. 4, 6
- [26] Q. Xu, J. Xiong, Q. Huang, and Y. Yao, "Robust evaluation for quality of experience in crowdsourcing," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 43–52. 5
- [27] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001. 6
- [28] M. A. Stricker and M. Orengo, "Similarity of color images," in *IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology*. International Society for Optics and Photonics, 1995, pp. 381–392. 6
- [29] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Real-time no-reference image quality assessment based on filter learning," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013. 6, 7
- [30] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. 6