Application-Driven No-Reference Quality Assessment for Dermoscopy Images With Multiple Distortions

Fengying Xie*, Yanan Lu, Alan C. Bovik, *Fellow, IEEE*, Zhiguo Jiang, and Rusong Meng

Abstract*—*Goal: **Dermoscopy images often suffer from blur and uneven illumination distortions that occur during acquisition, which can adversely influence consequent automatic image analysis results on potential lesion objects. The purpose of this paper is to deploy an algorithm that can automatically assess the quality of dermoscopy images. Such an algorithm could be used to direct image recapture or correction.** ***Methods:*** **We describe an application-driven no-reference image quality assessment (IQA) model for dermoscopy images affected by possibly multiple distortions. For this purpose, we created a multiple distortion dataset of dermoscopy images impaired by varying degrees of blur and uneven illumination. The basis of this model is two single distortion IQA metrics that are sensitive to blur and uneven illumination, respectively. The outputs of these two metrics are combined to predict the quality of multiply distorted dermoscopy images using a fuzzy neural network. Unlike traditional IQA algorithms, which use human subjective score as ground truth, here ground truth is driven by the application, and generated according to the degree of influence of the distortions on lesion analysis.** ***Results:*** **The experimental results reveal that the proposed model delivers accurate and stable quality prediction results for dermoscopy images impaired by multiple distortions.** ***Conclusion*****: The proposed model is effective for quality assessment of multiple distorted dermoscopy images.** ***Significance*****: An application-driven concept for IQA is introduced, and at the same time, a solution framework for the IQA of multiple distortions is proposed.**

***Index Terms*—Application driven, dermoscopy image, image quality assessment (IQA), multiple distortions, no reference (NR).**

I. Introduction

DERMOSCOPY is a noninvasive diagnostic technique which is useful in diagnosis of many skin diseases [1]. In recent years, dermoscopy technology has been developing toward network platforms. In 2005, Iyatomi *et al.* built an Internet-based remote diagnosis system [2], which permits ordinary people to upload dermoscopy image for analysis and diagnosis of skin diseases. In 2010, a Handyscope [3] was developed, by which dermoscopy images can be captured using a mobile phone. This development has enabled more nonclinical physicians to capture and upload dermoscopy images into remote diagnosis systems. Unfortunately, this process can easily lead to poor image quality (arising from, for example, hair, blur, and uneven illumination), which can adversely influence subsequent analysis. In [4] and [5], Rosado *et al.* pointed out that the quality of dermoscopy images captured via mobile device is problematic, and proposed a blur evaluation method for these types of images. This suggests that accurate image quality assessment (IQA) algorithms would be of great value towards assuring that the images being used for diagnosis or analysis stage are of high quality.

Blur and uneven illumination are two common distortions of dermoscopy images. It is a more complex problem to evaluate the image quality when multiple distortions (more than one distortion type) occur, since the distortions can interact and modify each other. However, recent work on dermoscopy image analysis has mainly focused on hair detection and removal, lesion segmentation, and classification [6]–[10]. IQA of dermoscopy images has thus far received little attention.

Over the past decade, numerous no-reference (NR) IQA models have been proposed for different purposes. These can be roughly divided into two groups: 1) models developed for specific image distortion types and 2) general-purpose or nondistortion-specific (sometimes called agnostic) models. IQA models developed for a single specific distortion type such as blur [11], JPEG [12], JPEG2000 [13], or noise [14] generally fail in the presence of other distortions. General-purpose IQA algorithms [15]–[18] are effective for assessing many types of single distortions. However, existing algorithms do not address two problems. First, almost all IQA algorithms have been designed to predict human judgments of perceived quality. However, in applications, high visual quality may not equate to high application value. For example, in medical image analysis, the quality of images should be expressed in terms of benefit in the sense that image analysis tasks be more efficient on images having high "task quality." Second, most of the distorted images in publicly available IQA datasets such as TID [19], IVC [20], A57 [21], LIVE II [22], and CSIQ [23] suffer from a single distortion type, although LIVE recently built a multiple distorted (MD) image quality dataset [24]. None of the images in these

Manuscript received April 3, 2015; revised August 17, 2015 and October 10, 2015; accepted October 19, 2015. Date of publication October 26, 2015; date of current version May 18, 2016. This work was supported by the National Natural Science Foundation of China under Grants 61471016, 61371134, and 61271436. The work of A. C. Bovik was supported by the U.S. National Science Foundation under Grant IIS-1116656 *Asterisk indicates corresponding author.*

*F. Xie is with the Beijing Key Laboratory of Digital Media and also with the Image Processing Center, Beihang University, Beijing 100083, China (e-mail: xfy_73@buaa.edu.cn).

Y. Lu and Z. Jiang are with the Beijing Key Laboratory of Digital Media, and also with the Image Processing Center, Beihang University (e-mail: yn_lu@buaa.edu.cn; jiangzg@buaa.edu.cn).

A. C. Bovik is with the Laboratory for Image and Video Engineering (LIVE), University of Texas at Austin (e-mail: bovik@ece.utexas.edu).

R. Meng is with the General Hospital of the Air Force, People's Liberation Army (e-mail: mrs5822@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2015.2493580

databases are directed toward any visual analysis or specific visual tasks, save the IQA task. Experiments indicate that state-of-the-art IQA models performed more poorly on the LIVE MD dataset than on single distortion datasets. The MD images are a challenge for IQA, since both the individual and joint effects of these distortions on the image as well as the effects of these distortions on each other must be considered [25]. In [26], Gu *et al.* proposed a five-step metric for assessing the quality of two types of MD images (blur followed by noise and blur followed by JPEG). Their experiments showed good performance on the LIVE MD dataset.

Here, we develop an application-driven NR IQA algorithm for assessing the task-related quality of dermoscopy images suffering from multiple distortions (blur and uneven illumination). The algorithm is called assessment of dermoscopy images with multiple distortions, or ADMD for short. Unlike traditional IQA algorithms driven by human vision models, in the proposed algorithm, ground truth is defined in terms of the influence of distortions on the efficacy of lesion analysis algorithms applied on dermoscopy images. An assessment model learned by a fuzzy neural network is used to predict the overall quality of MD dermoscopy images. The rest of this paper is organized as follows: Section II describes the new dataset. Section III presents the proposed IQA model and algorithm in detail. Experiments and analysis are presented in Section IV, and Section V concludes the paper.

II. Distorted Dermoscopy Image Database

A. Data Generation

Following the method of distorted image generation used by other public IQA datasets [19]–[24], the distorted images were created from reference dermoscopy images by applying blur followed by uneven illumination to them. Dermoscopy images usually suffer from vignetting, with black borders at the image boundaries [27], [28]. Since reference dermoscopy images should be undistorted, the vignetting and black border regions in them were removed. Fig. 1 depicts a number of MD images. The image at leftmost in the first row is the reference image. By filtering the reference image using an approximately circular averaging filter of varying radii, four blur images were generated, as shown in the first row. Then, by adding different uneven illumination masks (shown in Fig. 2) to each of these blur images, the next four rows of images with uneven illumination were generated, shown in the second to the fifth rows. Therefore, there are five levels of each distortion. The distortion severity increases from level 1 to level 4, whereas level 0 has no introduced distortion of that type. In the overall dataset, there are 18 representative reference images, each associated with 24 corresponding distorted images. Therefore, 450 images were obtained, including the references. There are 25 different distortion level/distortion combinations, each containing 18 images.

B. Image Quality Ground Truth

In the traditional IQA research, ground truth of true perceptual quality of distorted images is obtained via subjective experiments. However, in automatic dermoscopy image analysis, the image quality can significantly affect the accuracy of the subsequent analysis in two regards: segmentation and classification. Therefore, instead of regarding human judgments as ground truth, we have created a ground-truth database that expresses image quality as a function of the degree of influence on both segmentation and classification.

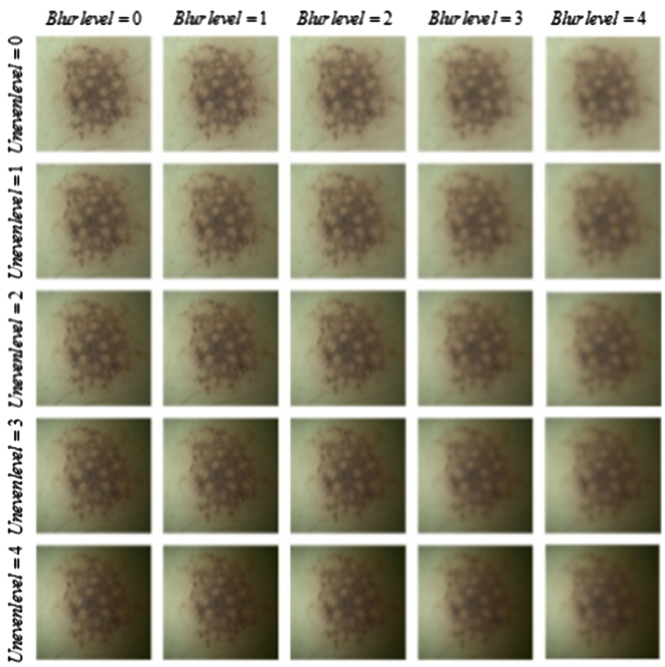

Fig. 1. Examples of MD dermoscopy images.

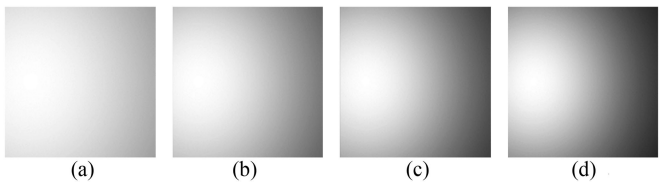

Fig. 2. Simulated uneven illumination masks. (a) Level 1. (b) Level 2. (c) Level 3. (d) Level 4.

1) Index of Influence on Segmentation: Segmentation results often change with increases of distortion level. Otsu's thresholding method [29] is simple and quick for calculation. Therefore, we used the Otsu's thresholding method to segment all of the images in the dataset. The difference in the segmentation results on the distorted image and its corresponding reference image is determined using the "XOR metric" [30]. A small XOR metric value indicates that the segmentation result of the distorted image is close to that of its reference image, which ostensibly means that the distortion is slight, whereas a large XOR metric value likely indicates a serious distortion.

Assume that the distorted images can be indexed by groups, where the images in each group have suffered the same degree of both blur and uneven illumination distortions. Then, for the ith group, the degree of influence of the composite of distortion combinations and distortion levels on segmentation is defined as the average value of the XOR metric values between the segmented versions of those images and those of the aligned

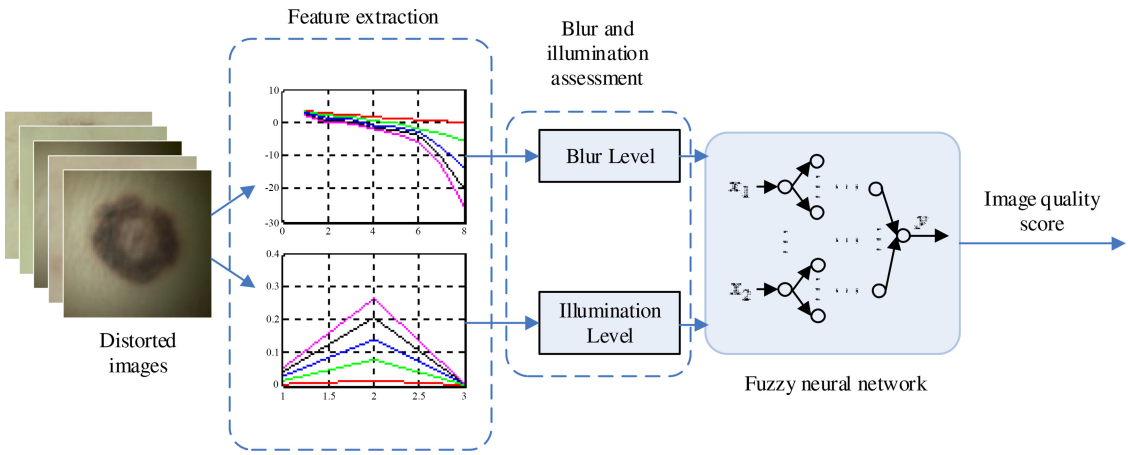

Fig. 3. Flowchart of the proposed algorithm ADMD.

reference images. The XOR metric used here is

$$gXOR_i = \frac{1}{N_i}\sum_{j=1}^{N_i} XOR_{ij}, i = 1, 2, \ldots 25 \quad (1)$$

where XOR_{ij} is the XOR value between the segmentation of the *j*th image in the *i*th group and that of its aligned reference, and N_i is the number of images in the *i*th group. Here, of course, we are summing XOR values as if they were algebraic, which is a simple way of expressing and normalizing a counting process. Here, $N_i = 18$.

2) Index of Influence on Classification: Image distortions may change the results of classification algorithms on lesions. If the lesion classification in the distorted image is different from that of its aligned reference image, then the distortion influences classification. Here, we use a classification method based on a neural network [31] to automatically classify all of the images. On the *i*th group, if n_i is the number of images having a different classification result with their aligned reference images, then the degree of influence of the corresponding distortion on lesion classification is defined as

$$r_i = \frac{n_i}{N_i}, i = 1, 2, \ldots 25 \quad (2)$$

where N_i is the same as in (1). The value of r_i falls between 0 and 1. The larger the value of r_i, the more serious the influence on lesion classification.

3) Ground-Truth Image Quality: Finally, we linearly combine $gXOR_i$ and r_i to form the overall ground-truth quality index of the MD image

$$\begin{aligned} q_i &= L(\alpha * gXOR_i + (1-\alpha) * r_i), i = 1, 2, \ldots 25 \\ L(x_i) &= \frac{1}{\max(x_i) - \min(x_i)}(x_i - \min(x_i)) \end{aligned} \quad (3)$$

where α is a weight parameter, set to 0.3 according to the dermatologist's experience in this paper; and $L(\cdot)$ remaps $q_i, i = 1, 2 \ldots 25$ to fall in the range [0,1]. Generally, the smaller the q_i value, the higher the task-affective image quality.

III. Method

Finding IQA models that can handle multiple distortions has proven difficult and to date, there exists little research providing a detailed analysis of the problem or that can guide one to a satisfactory solution [25]. Here, we propose a quality assessment framework that can handle images impaired by two coincident distortions: blur and uneven illumination. Features that are sensitive to these distortions are extracted and mapped to different distortion levels. Then, a fuzzy neural network is used to model the joint effect of blur and uneven illumination to enable prediction of the overall image quality. Fig. 3 illustrates a flowchart of the proposed method.

A. Blur Distortion Evaluation

Natural scene statistics (NSS) features are widely used in many state-of-the-art IQA algorithms [17], [32], [33]. Many NSS models are based on the empirical distributions of images decomposed by a wavelet transform into multiple subbands. The statistics of low–high (LH) and high–low (HL) subbands are similar [17], so they can be calculated together. Here, an input image is decomposed into four scales, hence, eight wavelet subbands are obtained. For each subband, a simple magnitude feature m_k is calculated as

$$m_k = \frac{1}{M_k \times N_k}\sum_{i=1}^{M_k}\sum_{j=1}^{N_k} \log_2 |C_k(i,j)| k = 1, 2, \ldots 8 \quad (4)$$

where $C_k(i,j)$ is the *k*th subband coefficient at (i,j), and M_k and N_k represent the width and length of the *k*th subband, respectively. Then, the wavelet feature vector can be written as

$$f_m = [m_1, m_2, \cdots m_8]^T. \quad (5)$$

Fig. 4 shows the magnitude features of images having different applied blur levels. Clearly, increases of the blur distortion level yields sharp decreases in the magnitude feature, which is highly indicative of the degree of blur distortion. Fig. 5 shows the magnitude features of images exhibiting different uneven illumination levels. These values lie very close to each other, because the magnitude feature is hardly affected by uneven

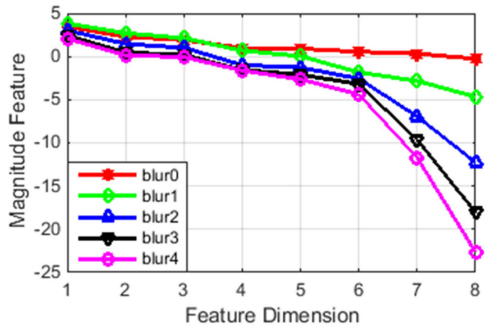

Fig. 4. Magnitude features of blur images, where red, green, blue, black, and magenta lines represent blur levels 0 to 4, respectively, and each line is the average result of ten distorted images.

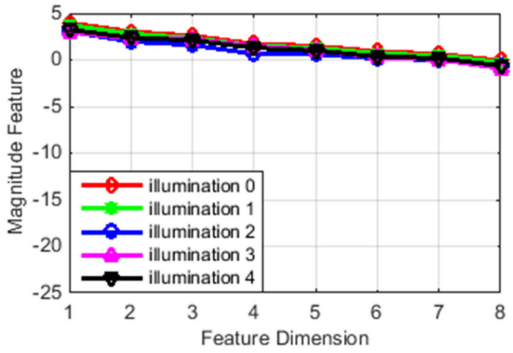

Fig. 5. Magnitude features of uneven illumination images, where different color lines represent different uneven illumination levels, and each line is the average result of ten images. The curves heavily overlap each other.

Fig. 6. Examples of illumination component extraction. (a) Uneven illumination image. (b) Extracted illumination component.

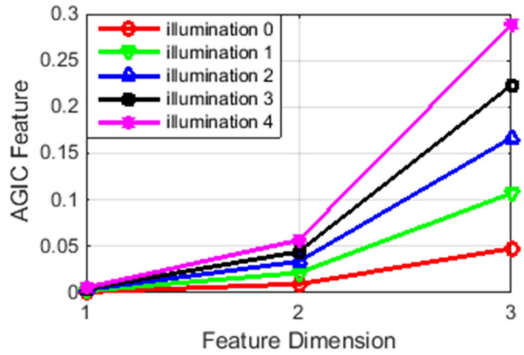

Fig. 7. AGIC features of uneven illumination images, where red, green, blue, black, and magenta lines represent uneven illumination levels ranging from level 0 to level 4 respectively. Each line is the average result of ten images.

illumination. This suggests that the magnitude feature may be used to correctly estimate the blur degree even when there is illumination distortion in the image.

The feature vector in (5) is mapped to blur distortion level using a support vector regressor (SVR). An SVR is able to handle high-dimensional data [34], and has been proved to be highly effective for solving other IQA problems [33], [35], [36]. For more details about an SVR, see [33]–[36].

B. Uneven Illumination Evaluation

As compared with blur, research reports on assessing the quality of images afflicted by uneven illumination are quite rare. In order to evaluate the level of this type of distortion, we decompose the distorted image into an illumination component and a reflectance component using a model from variational Retinex theory, and extract the illumination component via basis function fitting. Then, the average gradient magnitude of the illumination component (AGIC) is used to evaluate the level of uneven illumination distortion [37]. Fig. 6 is an instance of the extracted illumination component. AGIC can be calculated as

$$\mathrm{AGIC} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} g(i,j)$$

$$g(i,j) = \frac{\max |h(i,j) - h_k(i,j)|}{h(i,j)}, k = 1, 2, \ldots 8 \qquad (6)$$

where $i \in \{1, 2, \cdots M\}$, $j \in \{1, 2, \cdots N\}$, and $h(i,j)$ and $h_k(i,j)$ represent the average gray value of patch (i,j) and its eight-connected neighbors on the illumination component, respectively. The function $g(i,j)$ is a simple approximation of the gradient magnitude. For more details regarding the uneven illumination extraction and evaluation method, please see [37].

In order to better characterize the uneven illumination, a multiscale version of AGIC was also developed and used. AGIC was calculated over different patch sizes including 1×1 (unpatched), 10×10, and 50×50 to form a feature vector

$$f_{\mathrm{AGIC}} = [\mathrm{AGIC}_1, \mathrm{AGIC}_{10}, \mathrm{AGIC}_{50}]^T. \qquad (7)$$

Fig. 7 plots the AGIC features of images afflicted by different uneven illumination levels. Clearly, increases of the illumination distortion level yield sharp increases in the AGIC feature, hence, AGIC features are indicative of the degree of illumination distortion. Fig. 8 plots AGIC features of images having different blur levels. They are tightly clustered indicating that the AGIC feature is insensitive to blur distortion. Therefore, the AGIC feature is suitable for uneven illumination evaluation even if the image suffers from blur distortion. As with the blur evaluation, an SVR is used to map AGIC feature to illumination distortion levels.

C. Final Image Quality Prediction

The overall image quality score is predicted by combining the predicted levels of blur and uneven illumination. In [26], Gu *et al.* used a linear combination of single distortion quality measures to obtain a final image quality. However, they did not

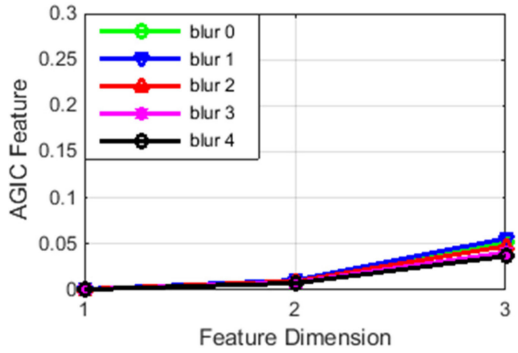

Fig. 8. AGIC features of blur images, where different color lines represent different blur levels. Each line is the average result of ten images. The curves heavily overlap each other.

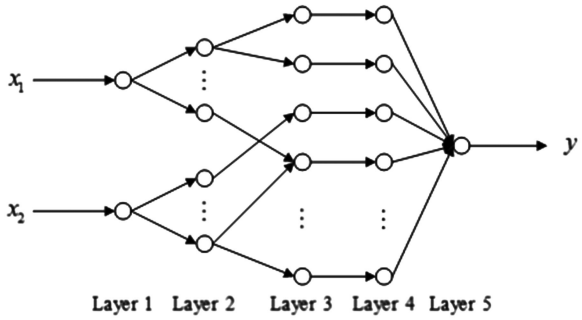

Fig. 9. Schematic diagram of the fuzzy neural network.

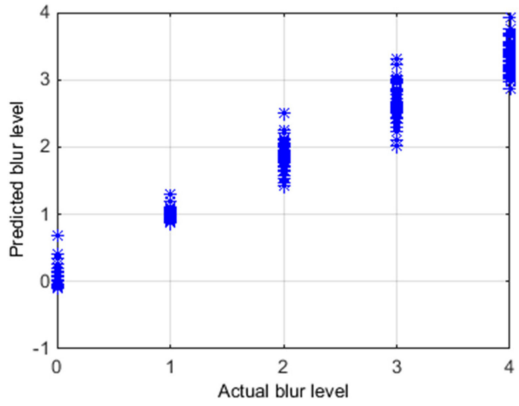

Fig. 10. Blur prediction.

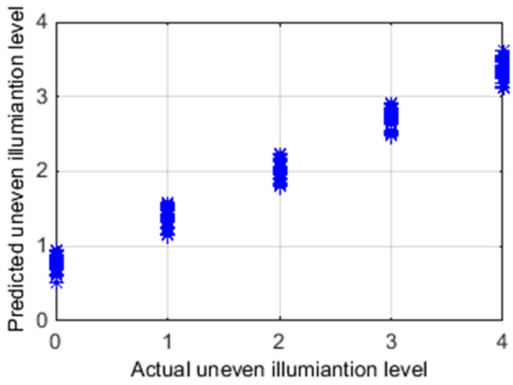

Fig. 11. Uneven illumination prediction.

account for nonlinear relationships between the degree of the single distortions and the final image quality [24], [25]. Here, we deploy a fuzzy neural network to model the complex joint effects of the two distortions on final quality score.

A fuzzy neural network [38] combines fuzzy logic with an artificial neural network to create an inference system. It is a powerful tool for handling uncertain, nonlinear, and other ill-posed problems. Fig. 9 shows the schematic diagram of the fuzzy neural network, which consists of five layers: input layer, fuzzy layer, rule layer, normalization layer, and output layer. In the proposed method, the two inputs descriptive of an image x_1 and x_2 are the blur and uneven illumination levels, respectively. The output y represents the image quality score, which is the ground truth in the training stage and the prediction in the test stage. Here, the number of nodes in the five layers is 2–10–25–25–1. In the training stage, the fuzzy neural network is trained on the image's single distortion levels and the quality ground truth to obtain a quality prediction model. In the test stage, blur and uneven illumination features are extracted from distorted images and mapped to distortion levels. Then, using the distortion levels as inputs, the quality score y is delivered.

IV. Experiments and Analysis

In order to quantitatively validate the performance of the proposed algorithm, experiments were conducted in regards to four aspects using our dataset: 1) effectiveness of the single distortion metrics; 2) effectiveness of the overall quality assessment model; 3) sensitivity of the model in relation to the training set size; and 4) performance for real distorted dermoscopy images. The performances of competing IQA algorithms were compared using two evaluation criteria: the Pearson linear correlation coefficient (LCC) and the Spearman rank-order correlation coefficient (SROCC). The LCC between the actual quality scores and the algorithm predicted scores provides an evaluation of prediction accuracy. The SROCC estimates the agreement between the rank-ordered actual quality score and the rank-ordered algorithm predictions and is a test of prediction monotonicity. The closer the value of LCC and SROCC are to 1, the better the performance of the algorithm is.

A. *Effectiveness of the Single Distortion Metric*

The accurateness of the single distortion assessment metrics substantially determines the efficacy of the proposed method. Each image group in our dataset was randomly divided into two parts, one for training and the other for testing. The images tested in this experiment included all distortion combinations. Figs. 10 and 11 show the single distortion prediction results. For each actual blur level in Fig. 10, the images include different degrees of uneven illumination ranging from level 0 to level 4, and the blur prediction results of different blur levels may be observed to be well separated. Thus, the proposed blur metric can effectively evaluate blur degree when the image also suffers from uneven

TABLE I
AVERAGE LCC AND SROCC OF THE SINGLE DISTORTION METRICS

	Blur	Uneven Illumination
LCC	0.9643	0.9838
SROCC	0.9534	0.9753

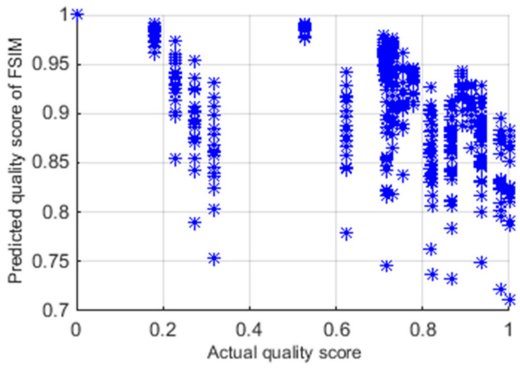

Fig. 12. Results of FSIM.

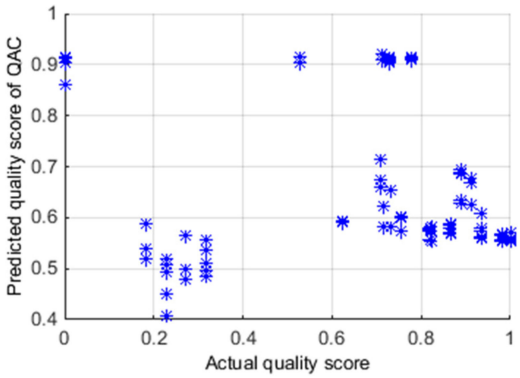

Fig. 13. Results of QAC.

illumination distortion. As with the blur metric, the proposed uneven illumination metric is insensitive to blur distortion, and can effectively evaluate uneven illumination degree even in the presence of blur distortion. Table I gives the average LCC and SROCC scores after repeating the training and testing process 1000 times. It can be seen that the proposed blur and uneven illumination metric is in high agreement with the actual single distortion level.

B. Effectiveness of the Overall Quality Assessment

The overall effectiveness of the proposed ADMD model was evaluated on the database of MD dermoscopy images. A linear combination of blur and uneven illumination measures was compared with the ADMD method. Several other IQA algorithms, including the full reference method FSIM [39], and two nondistortion-specific NR methods NIQE [16] and QAC [18] were also used to compare with our method. Figs. 12–16 show the quality prediction results of the five methods on our dataset,

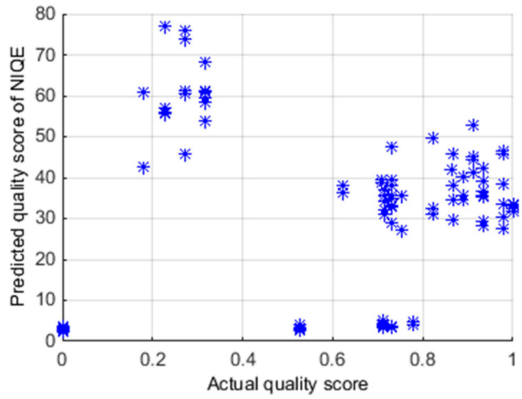

Fig. 14. Results of NIQE.

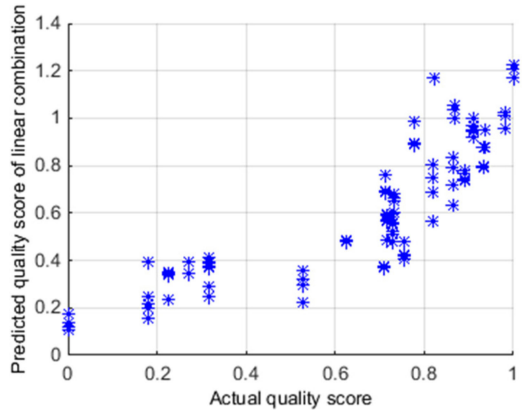

Fig. 15. Results of linear combination.

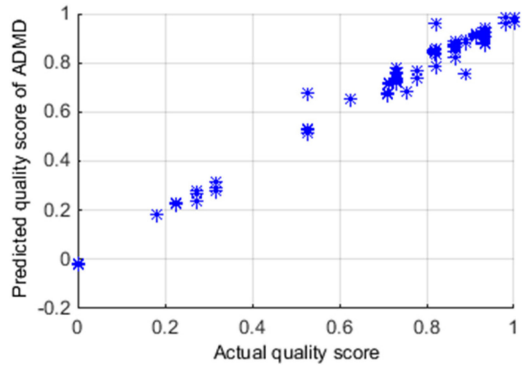

Fig. 16. Results of ADMD.

while Table II shows the LCC and SROCC scores. Obviously, FSIM, QAC, and NIQE all fail. The FSIM metric is designed to simulate human judgments of image quality, while here the ground truth of dermoscopy image quality is taken to be the degree of influence of the distortion on lesion analysis. QAC is based on FSIM, so they both fail. The failure of FSIM and QAC highlights the idea that the goals of application-driven IQA are different from those of the traditional IQA with regards to both target and solution. NIQE is a method based on NSS features. While NSS features are used to evaluate blur distortion in

TABLE II
AVERAGE LCC AND SROCC OF THE COMPETING IQA METHODS

	FSIM	QAC	NIQE	Linear Combination	Proposed ADMD
LCC	0.4774	0.1074	0.1520	0.8310	0.9740
SROCC	0.5623	0.1144	0.1083	0.8899	0.9544

TABLE III
LCC FOR DIFFERENT TRAINING SET SIZE

Ratio of Training Samples	60%	70%	80%	90%
Linear combination	0.8322	0.8304	0.8293	0.8299
Proposed ADMD	0.9731	0.9742	0.9752	0.9738

TABLE IV
SROCC FOR DIFFERENT TRAINING SET SIZE

Ratio of Training Samples	60%	70%	80%	90%
Linear combination	0.8901	0.8892	0.8883	0.8814
Proposed ADMD	0.9531	0.9547	0.9549	0.9524

ADMD. NIQE is also unsuccessful in this experiment, owing to its inability to gauge uneven illumination. The linear combination of blur and uneven illumination metrics achieves a better result than FSIM, QAC, and NIQE, but it does not compete well with ADMD, which delivers the most accurate prediction results among the five methods.

C. Sensitivity in Relation to Training Set Size

Many IQA algorithms heavily rely on the size of the training set. Their performance often becomes worse when the number of training samples decreases. In this experiment, the percentage of samples used for training ranged from 60% to 90%, while the percentage of test samples correspondingly ranged from 40% to 10%. Repeating the training and test procedure 1000 times yielded average LCC and SROCC scores for the linear combination method and for ADMD as shown in Tables III and IV. The results of FSIM, QAC, and NIQE are not given here, since they are not based on train-test processes; moreover, all of them failed in the test application in Section IV-B. From Tables III and IV, it can be seen that the proposed ADMD algorithm achieved better LCC and SROCC scores under all divisions of training samples. With LCC and SROCC greater than 0.95, the performance of ADMD is quite stable. As a check, we also conducted a complete cross-validation study ranging from twofold to tenfold, and found substantially the same results over all divisions as we report here using random selection. For brevity, we only report the latter results.

D. Performance for Real Distorted Dermoscopy Images

We collected 162 images suffering from real distortions, including blur and uneven illumination, obtained under a variety of capture conditions. In this section, the ground-truth quality is determined and the effectiveness of the proposed ADMD is verified for real distorted dermoscopy images.

TABLE V
SINGLE DISTORTION METRICS FOR REAL DISTORTED DERMOSCOPY IMAGES

	Blur	Uneven Illumination
LCC	0.9254	0.9118
SROCC	0.9193	0.9060

TABLE VI
AVERAGE LCC AND SROCC FOR REAL DISTORTED DERMOSCOPY IMAGES

	QAC	NIQE	Linear Combination	Proposed ADMD
LCC	0.1996	0.1859	0.7639	0.8415
SROCC	0.2566	0.2167	0.8389	0.8592

1) Ground-Truth Quality for Real Distorted Images: Unlike simulated images, real distorted dermoscopy image do not have associated reference images. Therefore, it is not possible to calculate gXOR and *r* indexes as in (1) and (2), hence, ground truth (3) is unavailable. As an alternative, a synthetic ground-truth comparison was accomplished using simulated distorted images, as follows.

Given a real distorted dermoscopy image, we found a simulated distorted image from the simulated dataset, suffering from similar degrees of blur and uneven illumination. The blur and uneven illumination distortion levels of the simulated image were considered as an approximation to the distortion levels of the real image. Following Section II-B, when two images suffer from similar blur and uneven illumination levels, they are regarded as having similar quality ground truth. Therefore, the ground-truth quality of the real distorted image is set to be the same as that of the corresponding simulated image.

2) Correlation Analysis: The blur and uneven illumination levels of each real distorted image are predicted using the methods described in Sections III-A and III-B. The prediction accuracy was evaluated using two correlation measures: LCC and SROCC. The results are tabulated in Table V. It may be seen that the single distortion indices proposed here are able to predict the degree of blur and uneven illumination degree effectively. Table VI compares four assessment methods with respect to overall quality assessment. The proposed ADMD model delivers the most accurate prediction results among these methods. Comparing Table VI with Table II, the quality prediction accuracy of ADMD on real distorted images is lower than on simulated dermoscopy images, which is to be expected [40]. However, the values LCC $= 0.8415$ and SROCC $= 0.8592$ obtained by ADMD on the real distorted images are quite satisfactory.

3) Removal of Low-Quality Filtration: The performance of ADMD on real distorted images can also be verified by assessing the statistical accuracy of subsequent lesion analysis against the quality predictions. The quality scores y on real distorted images lie between 0 and 1, where a higher quality score indicates worse image quality. When the quality y of an image

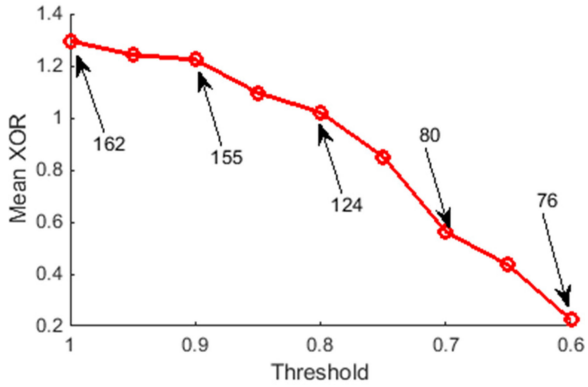

Fig. 17. Segmentation accuracy using different quality thresholds. The numbers with arrows indicate the number of images remaining following application of the threshold T.

was greater than a threshold T, the image was removed from consideration. Therefore, only images of high quality $(y < T)$ were saved and used for the subsequent lesion analysis. We used tenfold cross validation to obtain quality predictions on the real images. Fig. 17 plots segmentation accuracy against the quality threshold T. The segmentation accuracy is expressed as the mean XOR value of the remaining images. A smaller XOR indicates a higher segmentation accuracy. The relationship in Fig. 17 is largely monotonic. The mean XOR value decreases as the quality threshold is reduced. This suggests that our application-driven IQA algorithm is quite useful for ensuring the accuracy of lesion segmentation and is effective on real distorted images. The classification accuracy is not given here since very few malignant lesion samples were present in the data.

V. Conclusion

An IQA process would be a highly desirable part of the process of acquiring dermoscopy images to ensure that the images being analyzed are not distorted. Toward this end, we introduced a novel application-driven IQA model called ADMD for analyzing the quality of MD dermoscopy images. In this algorithm, two single distortions: blur and uneven illumination are separately evaluated. Then, taking the two single distortion levels as inputs, the overall image quality is predicted by a fuzzy neural network. There are four main contributions made. The first is a new application-driven IQA concept is proposed; the second is a new dataset for MD dermoscopy images with blur and uneven illumination that we created. The third contribution is single distortion metrics for blur and uneven illumination that are only sensitive to their corresponding distortion, while insensitive to the other. The fourth and last contribution is an overall dermoscopy image quality assessment framework based on a fuzzy neural network. The overall framework effectively captures the highly nonlinear mutual effects of multiple distortions and it provides an IQA solution framework that may prove useful for assessing multiple distortions not only on dermoscopy images but also on other types of images. A series of experiments showed that the proposed algorithm is effective, stable, and insensitive to training set size.

References

[1] H. Kittler *et al.*, "Diagnostic accuracy of dermoscopy," *Lancet. Oncol.*, vol. 3, no. 3, pp. 159–165, 2002.
[2] H. Iyatomi *et al.*, "An internet-based melanoma diagnostic system-toward the practical application," in *Proc. IEEE Symp. Comput. Intell. Bioinformat. Comput. Biol.*, 2005, pp. 1–4.
[3] Handyscope-mobile dermatoscope: Handyscope (2010). [Online]. Available: http://www.handyscope.net
[4] M. Vasconcelos and L. Rosado, "No-reference blur assessment of dermatological images acquired via mobile devices," in *Proc. Int. Conf. Image Signal Process.*, 2014, pp. 350–357.
[5] L. Rosado *et al.*, "From dermoscopy to mobile teledermatology," in *Dermoscopy Image Analysis*. Boca Raton, FL, USA: CRC Press, 2015, pp. 385–418.
[6] F. Xie *et al.*, "PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma," *Comput. Med. Imag. Graph.*, vol. 33, no. 4, pp. 275–282, 2009.
[7] F. Xie and A. C. Bovik, "Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm," *Pattern Recog.*, vol. 46, no. 3, pp. 1012–1019, 2013.
[8] Y. He and F. Xie, "Automatic skin lesion segmentation based on texture analysis and supervised learning," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 330–341.
[9] Q. Abbas *et al.*, "Pattern classification of dermoscopy images: A perceptually uniform model," *Pattern Recog.*, vol. 46, no. 1, pp. 86–97, 2013.
[10] M. Sadeghi *et al.*, "Global pattern analysis and classification of dermoscopic images using textons," in *Proc. SPIE Med. Imag.*, vol. 8314, pp. 83144X-1–83144X-6, 2012.
[11] P. Marziliano *et al.*, "A no-reference perceptual blur metric," in *Proc. Int. Conf. Image Process.*, 2002, pp. III–57–III-60.
[12] A. C. Bovik and S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2001, pp. 1725–1728.
[13] H. R. Sheikh *et al.*, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.
[14] X. Kong *et al.*, "A new image quality metric for image auto-denoising," in *Proc. Int. Conf. Comput. Vis.*, 2013, pp. 2888–2895.
[15] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.
[16] A. Mittal *et al.*, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[17] L. He *et al.*, "Sparse representation for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 1146–1153.
[18] W. Xue *et al.*, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 995–1002.
[19] N. Ponomarenko *et al.*, "TID2008-A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.
[20] Subjective quality assessment IVC database. (2005). [Online]. Available: http://www.irccyn.ec-nantes.fr/ivcdb/
[21] A57 Database. [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html
[22] H. R. Sheikh *et al.* Live image quality assessment data set release 2. (2006). [Online]. Available: http://live.ece.utexas.edu/research/ quality/subjective.htm
[23] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011006-1–011006-21, 2010.
[24] D. Jayaraman *et al.*, "Objective quality assessment of multiply distorted images," in *Proc. Conf. Rec.46th Asilomar Conf. Signals Syst. Comput.*, 2012, pp. 1693–1697.
[25] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Process.*, vol. 2013, pp. 905685-1–905685-53, 2013.
[26] K. Gu *et al.*, "FISBLIM: A five-step blind metric for quality assessment of multiply distorted images," in *Proc. IEEE Workshop Signal Process. Syst.*, 2013, pp. 241–246.
[27] H. Wang *et al.*, "Watershed segmentation of dermoscopy images using a watershed technique," *Skin Res. Technol.*, vol. 16, no. 3, pp. 378–384, 2010.
[28] M. E. Celebi *et al.*, "Lesion border detection in dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 33, no. 2, pp. 148–153, 2009.

[29] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, pp. 23–27, 1975.
[30] M. E. Celebi *et al.*, "An improved objective evaluation measure for border detection in dermoscopy images," *Skin Res. Technol.*, vol. 15, no. 4, pp. 444–450, 2009.
[31] F. Xie, "Segmentation and recognition of dermoscopy melanoma images based on computational intelligence," Ph.D. thesis, Beihang University, Beijing, China, May 2009.
[32] H. R. Sheikh *et al.*, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.
[33] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[34] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.
[35] A. Mittal *et al.*, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[36] J. Wu *et al.*, "Image quality assessment with degradation on spatial structure," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 437–440, Apr. 2014.
[37] Y. Lu *et al.*, "No reference uneven illumination assessment for dermoscopy images," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 534–538, May 2015.
[38] C. T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Trans. Comput.*, vol. 40, no. 12, pp. 1320–1336, Dec. 1991.
[39] L. Zhang *et al.*, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[40] M. A. Saad *et al.*, "Objective consumer device photo quality evaluation," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1516–1520, Oct. 2015.

Fengying Xie received the Ph.D. degree in pattern recognition and intelligent system from Beihang University, Beijing, China, in 2009.

She was a Visiting Scholar in the Laboratory for Image and Video Engineering, University of Texas at Austin, Austin, TX, USA, from 2010 to 2011. She is currently a Professor in Image Processing Center, School of Astronautics, Beihang University. Her research interests include biomedical image and remote sensing image processing, image quality assessment, and image segmentation and classification.

Yanan Lu received the B.Eng. degree from Beihang University, Beijing, China, in 2011, where she is currently working toward the Ph.D. degree majored in pattern recognition and intelligent system.

Her research interests include biomedical image processing and image quality assessment and restoration.

Alan C. Bovik (S'80–M'81–SM'89–F'96) is the Cockrell Family Regents Endowed Chair Professor at The University of Texas at Austin, Austin, TX, USA, where he is the Director of the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering and the Institute for Neuroscience. His research interests include image and video processing and visual perception. He has published more than 750 technical articles and holds four U.S. patents. His several books include the recent companion volumes *The Essential Guides to Image and Video Processing* (San Diego, CA, USA: Academic Press, 2009).

Prof. Bovik has received numerous awards from the IEEE Signal Processing Society: the Society Award (2013), the Best Paper Award (2009), the Best Magazine Paper Award (2013), the Education Award (2007), the Technical Achievement Award (2005), the Meritorious Service Award (1998), and the young author Best Paper Award (coauthor, 2013). He also received Honorary Membership in the Society for Imaging Science and Technology in 2013, the SPIE Technology Achievement Award in 2012, and was named the IS&T/SPIE Imaging Scientist of the Year in 2011. He is a Fellow of the Optical Society of America and the Society of Photo-Optical and Instrumentation Engineers. He cofounded and served as an Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002 and founded and served as the first General Chairman of the IEEE International Conference on Image Processing, Austin, in 1994.

Zhiguo Jiang received the B.Eng., M.S., and Ph.D. degrees from the Beihang University, Beijing, China, in 1987, 1990, and 2005, respectively.

He is currently a Professor in Image Processing Center, School of Astronautics, Beihang University. His research interests include medical image processing, segmentation and classification, remotely sensed image processing, and target detection, tracking, and recognition.

Rusong Meng is a Deputy Chief Physician of the General Hospital of the Air Force of PLA, Beijing, China. His research interests include morphological analysis of histiocytes, derma pathology, and the clinical application of image analysis technology.