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

Beihang University

IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING,2016

2019/3/31

Motivation

- Blur and uneven illumination are two main distortions of dermoscopy images
- Poor image quality can influence the analysis(proven?)
- IQA on dermoscopy images receives little attention

Main Contribution

- Multiple distortion datasets of even illumination
- Application-driven image quality assessment model

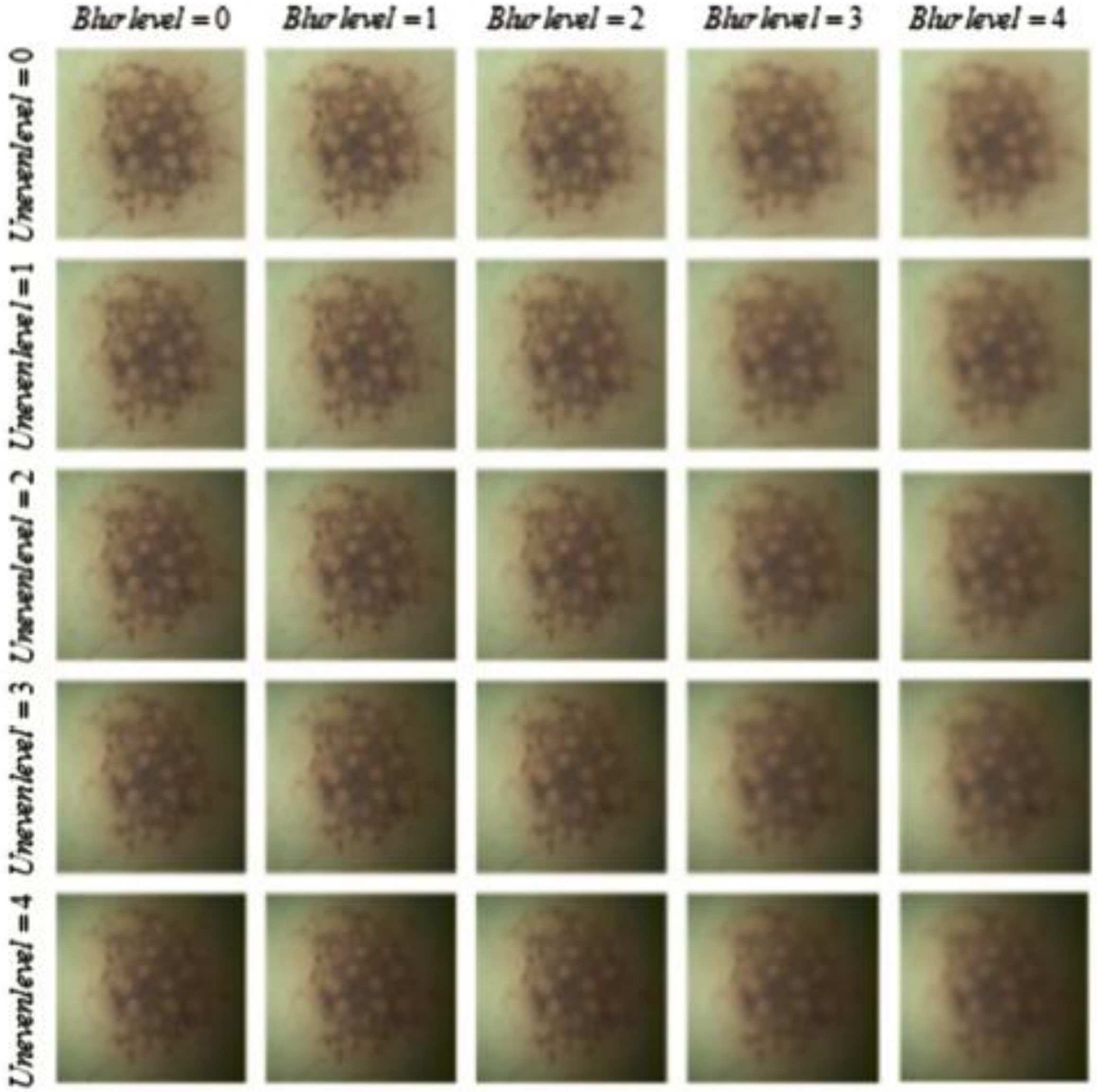
Related Work

- Models for specific image distortion types
  - Blur(P. Marziliano et al., “A no-reference perceptual blur metric,”ICIP2002)
  - JPEG(A. C. Bovik and S. Liu, “DCT-domain blind measurement of blocking artifacts in DCT-coded images,” Int. Conf. Acoust 2001.)
  - JPEG2000(H. R. Sheikh et al., “No-reference quality assessment using natural scene statistics: JPEG2000”,TIP2005)
  - Noise(X. Kong et al., “A new image quality metric for image auto-denoising,” in ICCV, 2013,)
- Models for general-purpose
  - A. K. Moorthy and A. C. Bovik, “A two-step framework for constructing blind image quality indices,” *IEEE Signal Process. Lett.,* 2010.
  - A. Mittal *et al., “Making a ‘completely blind’ image quality analyzer,” IEEE Signal Process. Lett., 2013.*

Database

- Generation
  - Reference dermoscopy images
  - Filter the reference images of four blur images
  - Add four uneven illumination masks
  - 18*25 = 450

Database



Blur level = 0
Blur level = 1
Blur level = 2
Blur level = 3
Blur level = 4
Unevenlevel = 0
Unevenlevel = 1
Unevenlevel = 2
Unevenlevel = 3
Unevenlevel = 4

Database

- Ground truth
  - Index of Influence on segmentation : XOR(border)

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

Database

- Ground truth
  - Index of Influence on classification

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

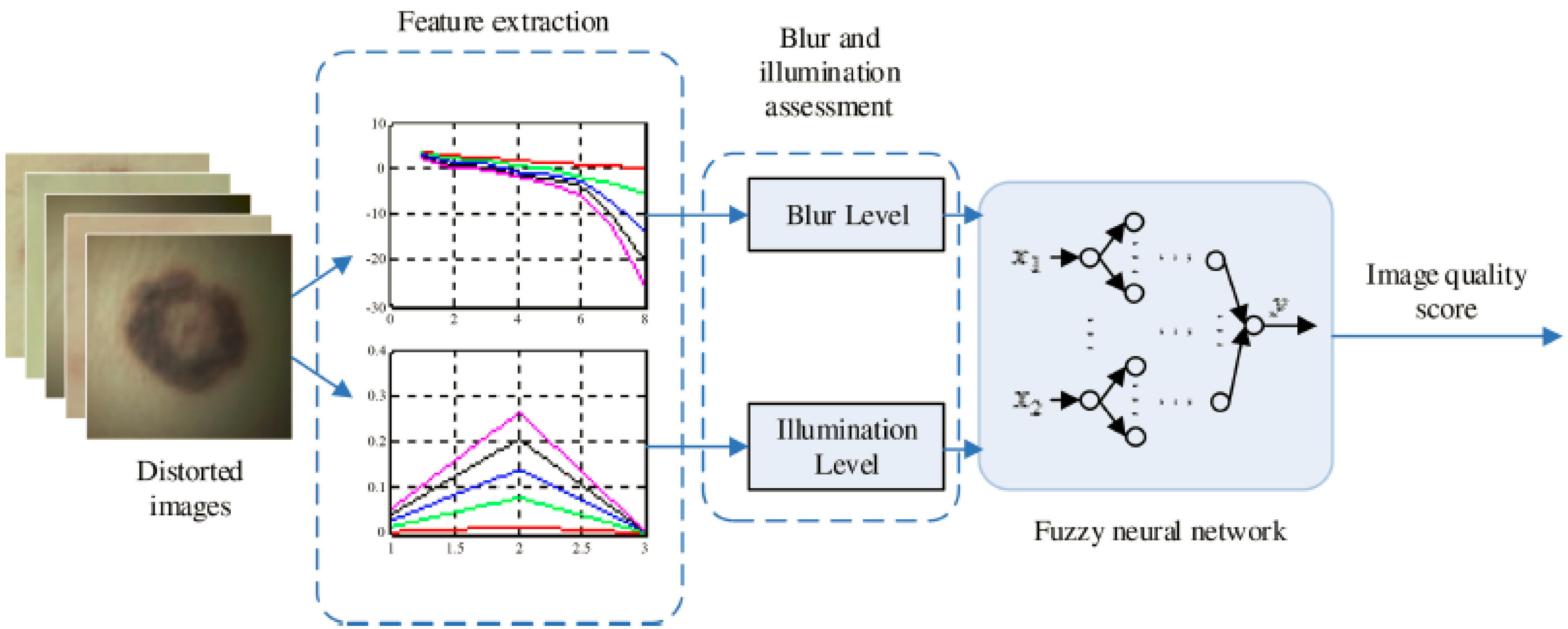
Database

- Ground truth
  - Ground truth image quality:

$$q_i = L(\alpha * g\mathrm{XOR}_i + (1-\alpha) * r_i), i = 1, 2, \ldots 25$$

$$L(x_i) = \frac{1}{\max(x_i) - \min(x_i)}(x_i - \min(x_i)) \qquad (3)$$

Method


Feature extraction
Blur and illumination assessment
Blur Level
Illumination Level
Image quality score
Distorted images
Fuzzy neural network

Method

- Blur distortion evaluation
    - Natural scene statistics (NSS) features
    - Magnitude feature can estimate the blur degree even if there is illumination distortion

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

$$f_m = [m_1, m_2, \cdots m_8]^T.$$

Method

- Blur distortion evaluation
  - Magnitude feature can estimate the blur degree even if there is illumination distortion

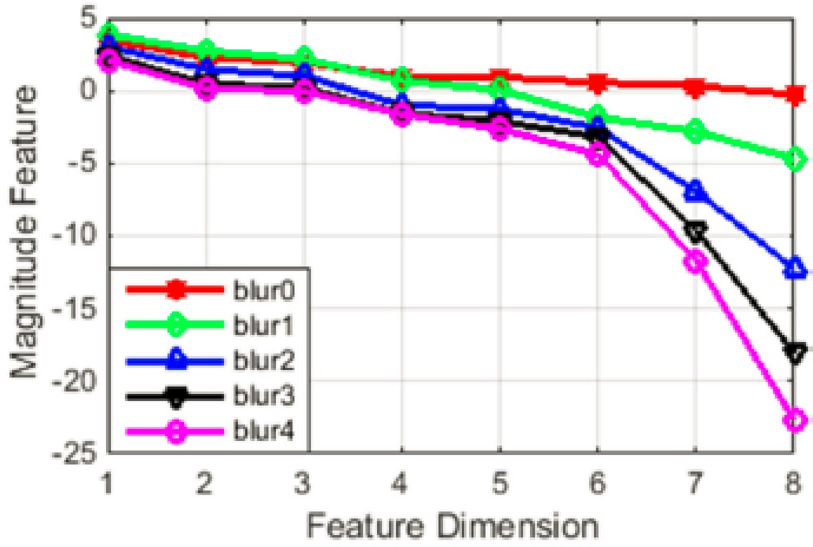


Fig. 4. Magnitude features of blur images, where red, green, blue, black, and magenta lines represent blur levels 0 to 4, respectively, and each line is the average result of ten distorted images.

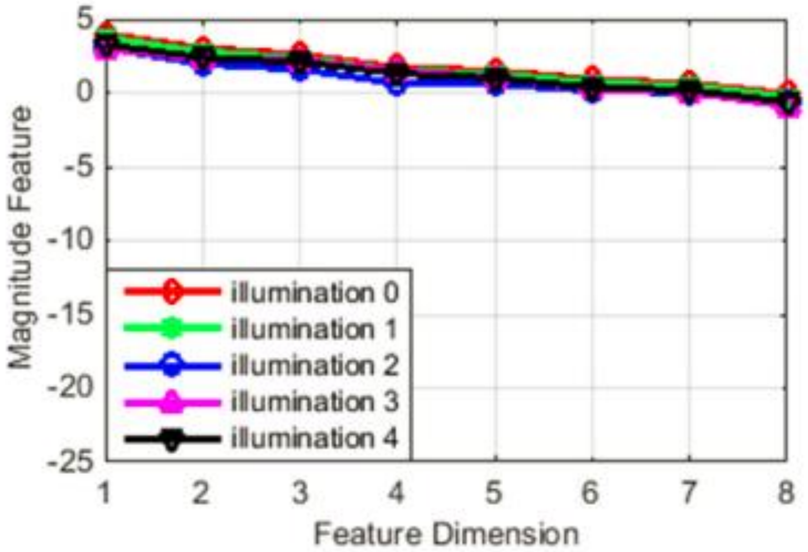


Fig. 5. Magnitude features of uneven illumination images, where different color lines represent different uneven illumination levels, and each line is the average result of ten images. The curves heavily overlap each other.

Method

- Blur distortion evaluation
  - Magnitude feature can estimate the blur degree even if there is illumination distortion
  - f_m is mapped to a level using a support vector regressor

Method

- Uneven illumination evaluation
  - Average gradient magnitude of the illumination distortion(AGIC)

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

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

$h_k(i,j)$ represent the average gray value of patch (i,j)

Method

- Uneven illumination evaluation
    - Even if there is blur distortion, AGIC works well

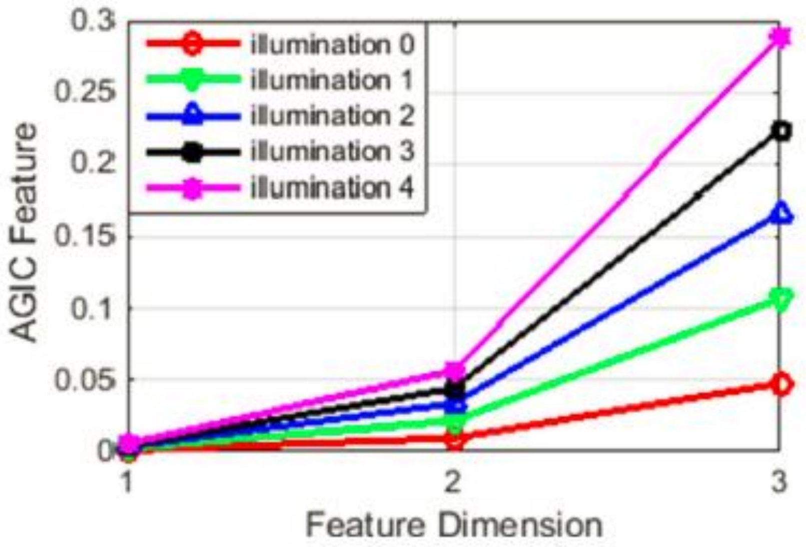


Fig. 7. AGIC features of uneven illumination images, where red, green, blue, black, and magenta lines represent uneven illumination levels ranging from level 0 to level 4 respectively. Each line is the average result of ten images.

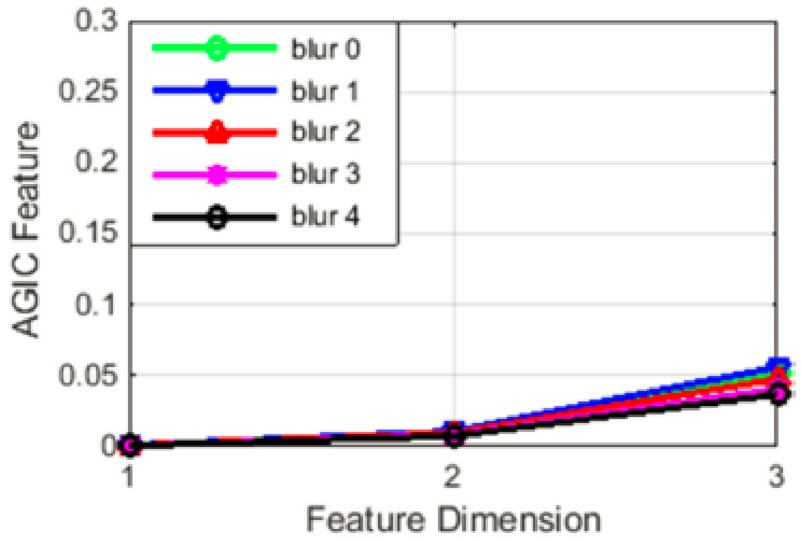


Fig. 8. AGIC features of blur images, where different color lines represent different blur levels. Each line is the average result of ten images. The curves heavily overlap each other.

Method

- Uneven illumination evaluation
  - Even if there is blur distortion, AGIC works well
  - AGIC is mapped to a level using a support vector regressor

Method

- Final Image quality prediction

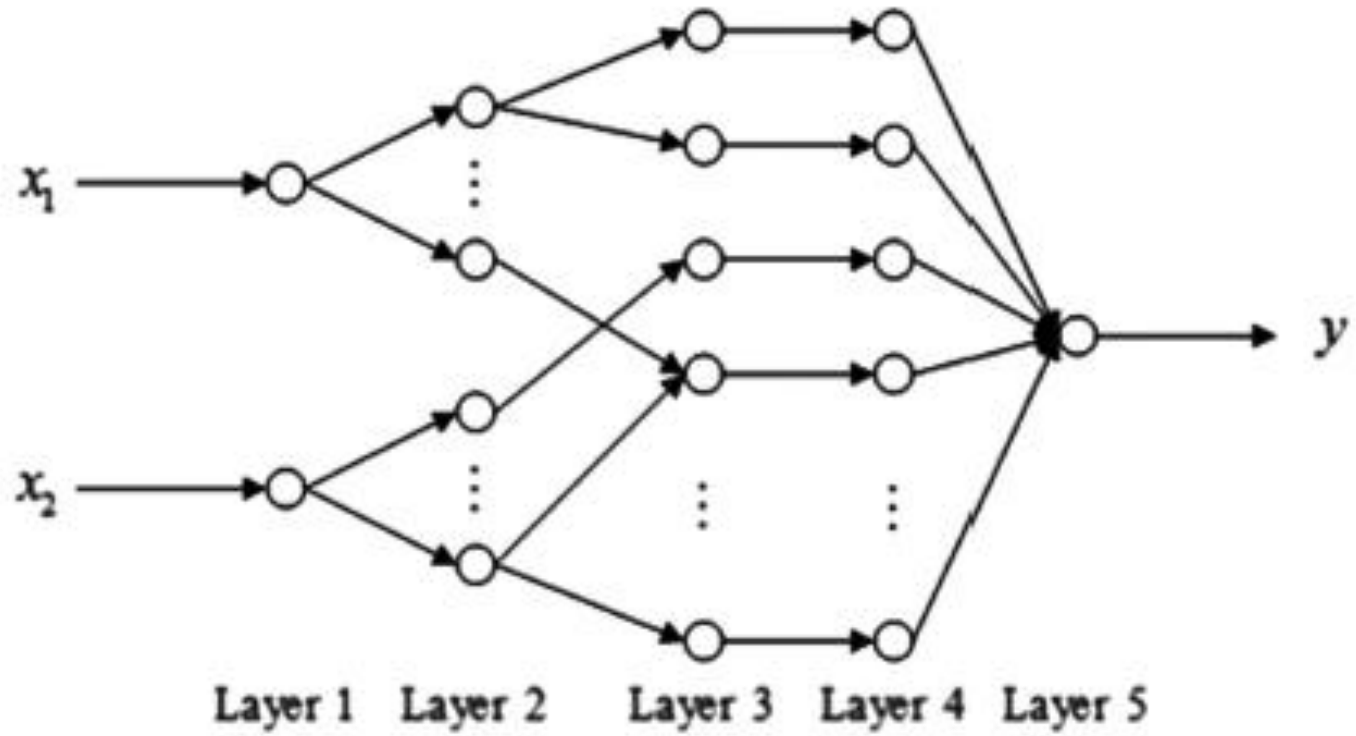


Fig. 9. Schematic diagram of the fuzzy neural network.

Experiments

- Effectiveness of the single distortion metrics
- Effectiveness of the overall quality assessment model
- Sensitivity in Relation to Training Set Size
- Performance for real distorted dermoscopy images

- LCC
- SROCC

Experiments

- Effectiveness of the single distortion metrics

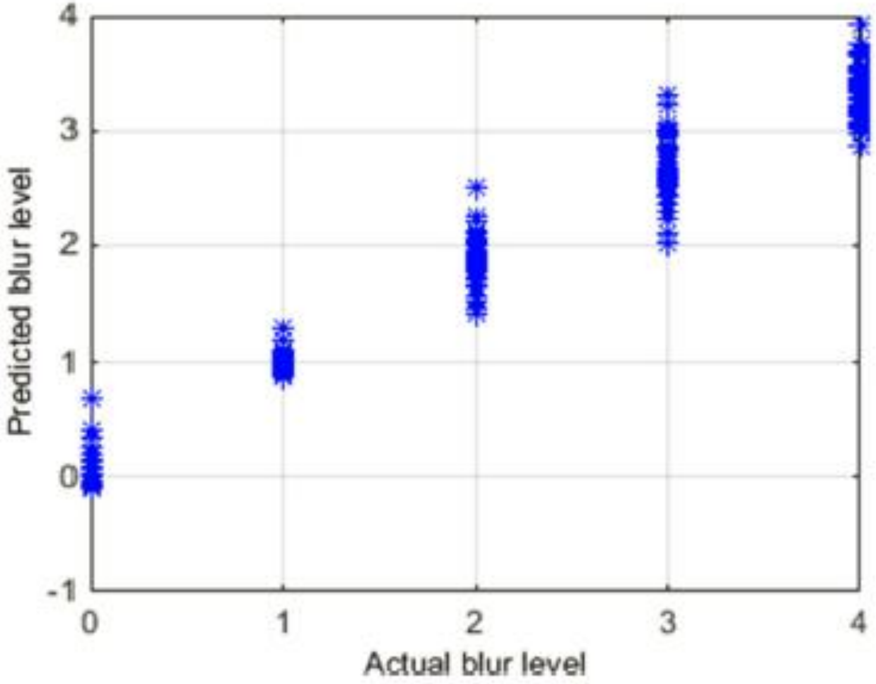


Fig. 10. Blur prediction.

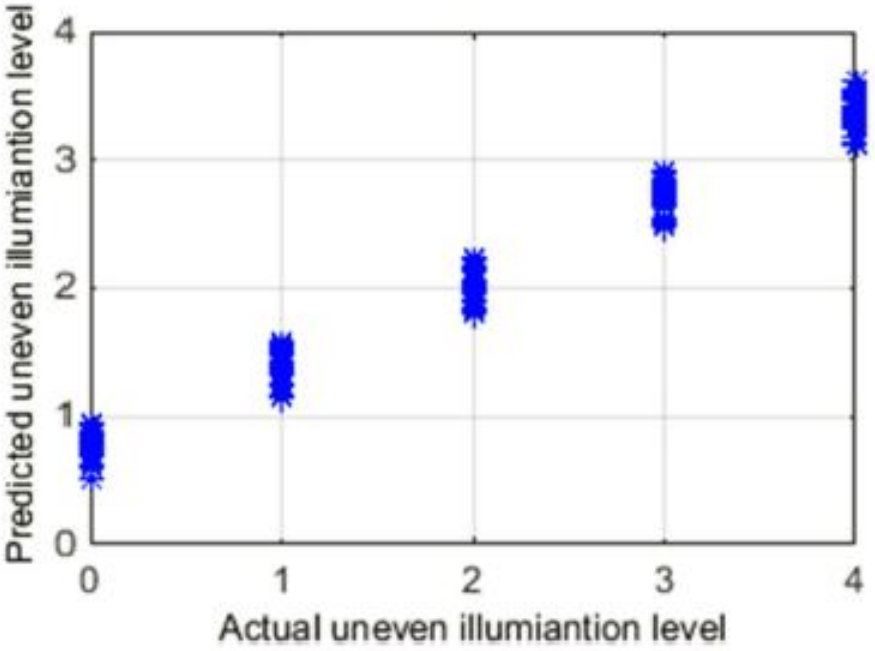


Fig. 11. Uneven illumination prediction.

Experiments

- Effectiveness of the single distortion metrics

TABLE I
AVERAGE LCC AND SROCC OF THE SINGLE DISTORTION METRICS

	Blur	Uneven Illumination
LCC	0.9643	0.9838
SROCC	0.9534	0.9753

Experiments

- Effectiveness of the Overall Quality Assessment

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

Experiments

- Sensitivity in Relation to Training Set Size

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

Experiments

- Performance for Real Distorted Dermoscopy Images

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

My own thinking

- It’s important to prove that poor image quality can influence the analysis(the value of this problem)
- Other visual features can be used for IQA(Some specific layes of CNNs)
- Classification results can be an important feature for IQA(The ratio between it and visual features)
- Still hand-crafted features even though CNN is used
- Too much levels, paired training , single image testing