Blind Image Quality Assessment with a Probabilistic Quality Representation

Hui Zeng[1], Lei Zhang[1], Alan C. Bovik[2]

[1]Department of Computing, The Hong Kong Polytechnic University

[2]Department of Electrical and Computer Engineering,

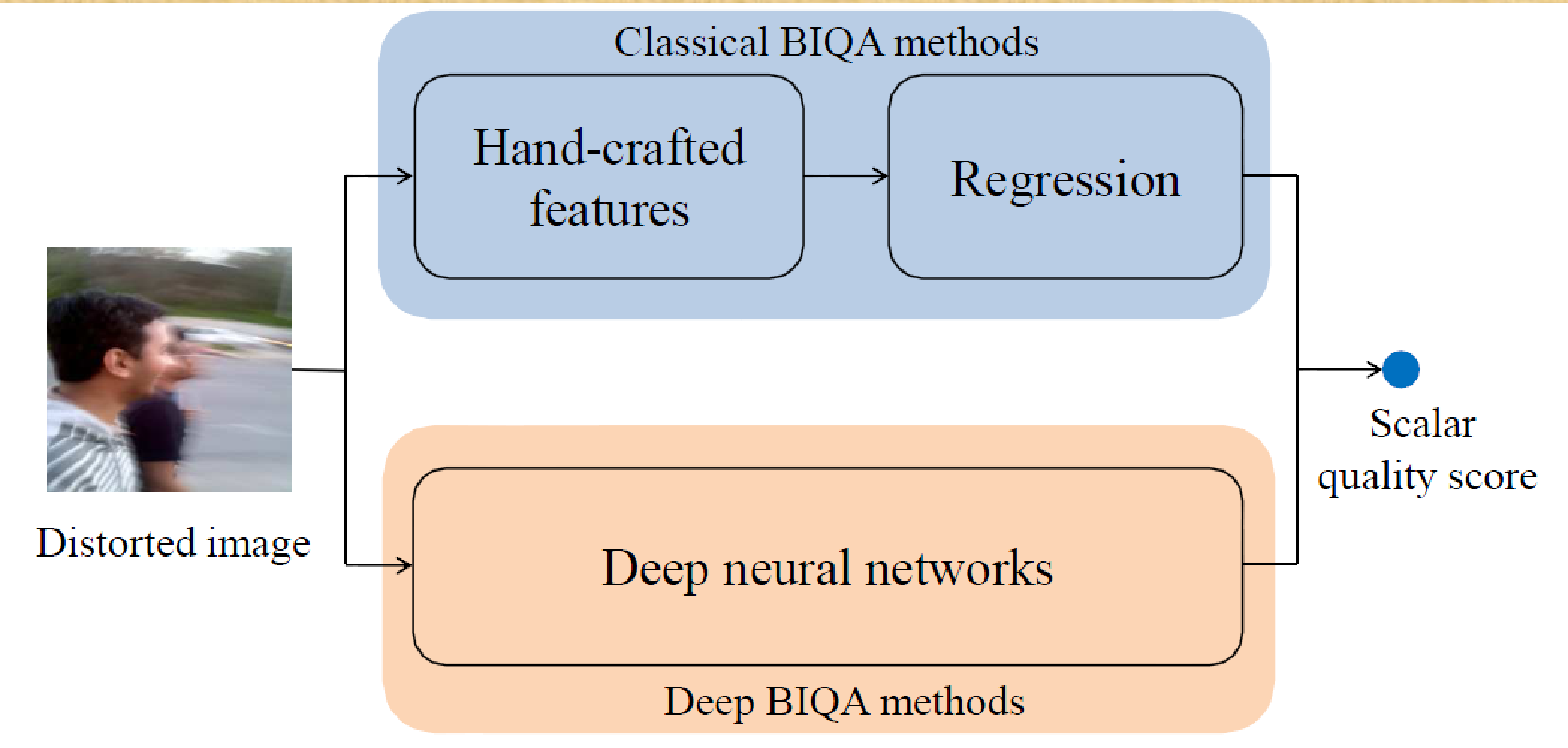
The University of Texas at Austin

Outline

- Introduction
- Probabilistic quality representation
- Experiments
- Q&A

Introduction

- Existing BIQA methods learn to regress scalar quality scores



Introduction

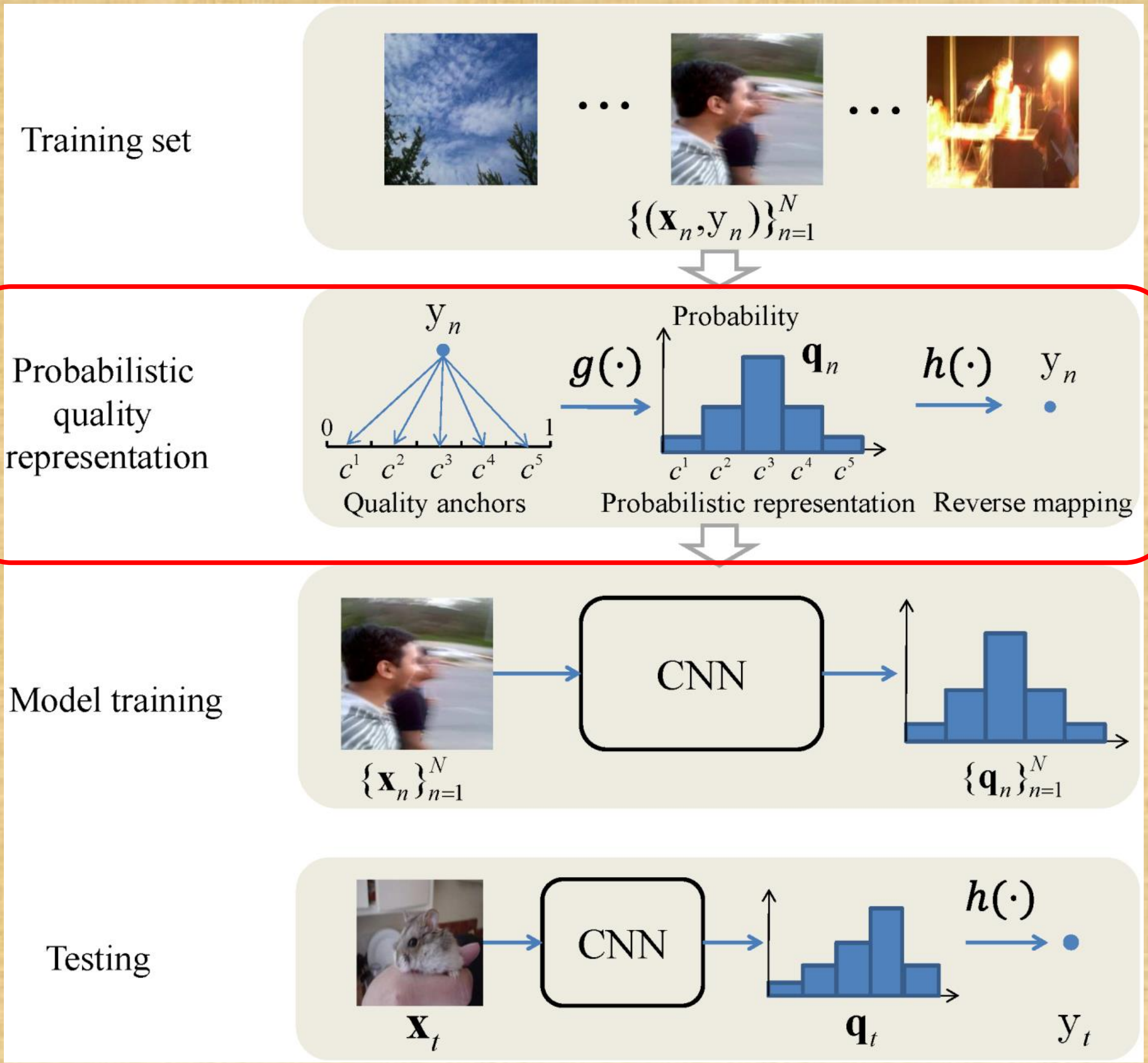
- An image will receive divergent subjective scores from different human raters.
    - The average standard deviation on the LIVE Challenge database is 19.27 on the MOS scale of [0,100].
- **Limitation:** The scalar MOS cannot adequately represent subjective property of human perception on image quality.
- **Challenge:** the score distributions are unavailable on existing IQA databases.

Introduction

- **Our idea:**
  - Propose a probabilistic quality representation (PQR) to approximately describe the subjective score distribution
  - Train model using the vectorized PQR instead of scalar MOS

- **Expected outcomes**:
  - Faster convergence speed
  - higher prediction accuracy

Overall framework



Probabilistic quality representation

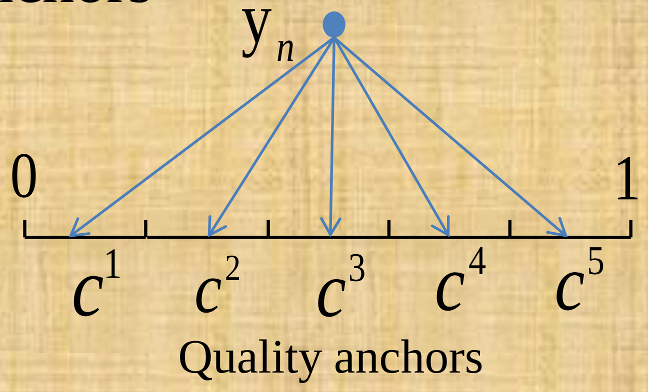
- **Step 1: quality anchors**

➢ Normalize the MOS to [0,1]

➢ Divide the score range into M equal bins

➢ Define the midpoints of bins as quality anchors

$$\{c^m\}_{m=1}^{M}$$



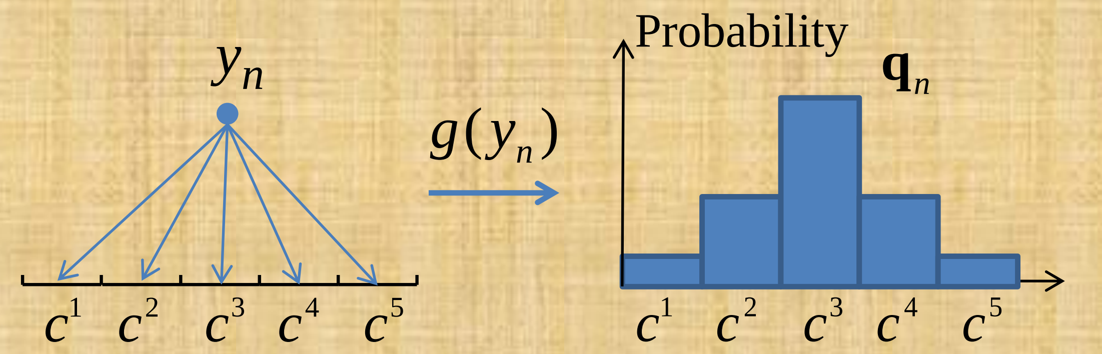
Probabilistic quality representation

- **Step 2: probabilistic representation**

➢ Two constraints:

1. Smaller distance → higher probability
2. All probabilities sum to 1

$$q_n^m = \frac{\exp(-\beta \| y_n - c^m \|^2)}{\sum_{i=1}^{M} \exp(-\beta \| y_n - c^i \|^2)}$$



Quality anchors

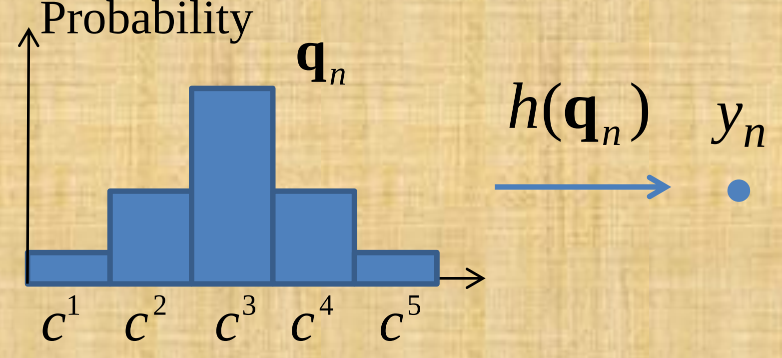
Probabilistic representation

Probabilistic quality representation

- **Step 3: reverse mapping**

- PQR vectors should be able to mapped back to scalar quality scores for evaluation
- Linear SVM minimization: $err = \frac{1}{N}\sum_{n=1}^{N} \| h(\mathbf{q}_n) - y_n \|^2$
- Negligible error:

$$err = \frac{1}{N}\sum_{n=1}^{N} | h(\mathbf{q}_n) - y_n | < 0.01$$



Probabilistic representation

Probabilistic quality representation

- Network training with PQR

- Loss function: $$\min \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} -q_n^m \log q_n^m$$

- **Benefits:**

1. Richer information (vectorized representation), stabilizing the training process, better generalization
2. Enable to use softmax cross-entropy for network training: faster convergence than MSE

Experiments

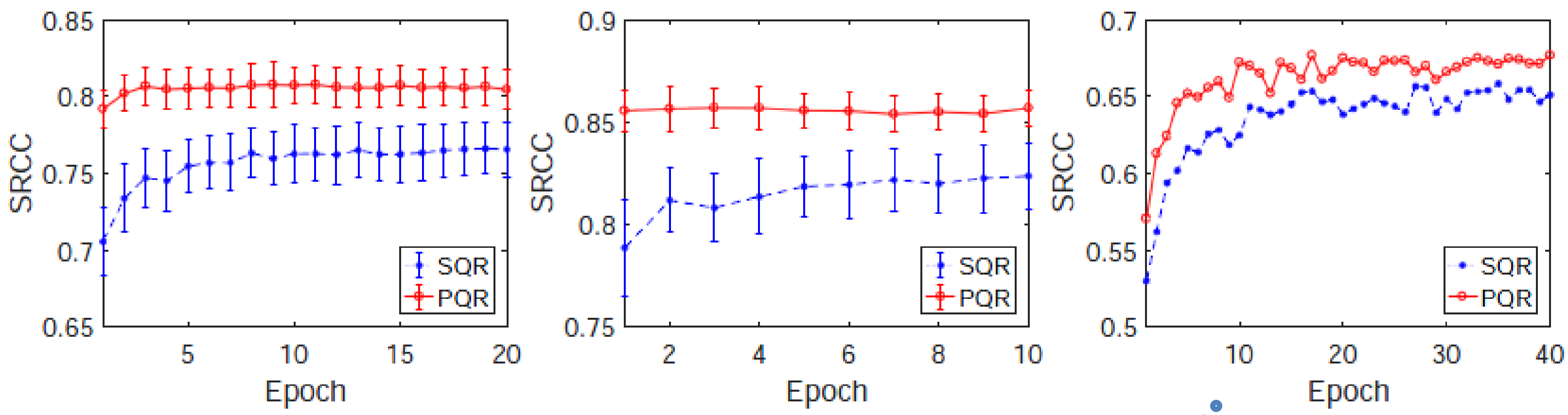
- Four databases:
  - LIVE Challenge, LIVE, CSIQ, TID2013
- Three base CNN models:
  - AlexNet
  - ResNet50
  - A shallow CNN (S_CNN) training from scratch
- Two metrics:
  - median SRCC and PCC over 10 times repetitions
- Random crops for data augmentation

Experiments

- Comparison Against Scalar Regression

Higher accuracy

CNN model	Representation	LIVE-C		LIVE		CSIQ		TID2013	
		SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
AlexNet	SQR	0.7658	0.8074	0.9319	0.9462	0.7965	0.8405	0.5362	0.6136
	PQR	0.8075	0.8357	0.9554	0.9638	0.8713	0.8958	0.5742	0.6687
ResNet50	SQR	0.8236	0.8680	0.9468	0.9527	0.8217	0.8713	0.6406	0.7068
	PQR	**0.8568**	**0.8822**	**0.9653**	**0.9714**	0.8728	0.9010	**0.7399**	**0.7980**
S_CNN	SQR	0.6582	0.6729	0.9450	0.9455	0.8787	0.8987	0.6526	0.6921
	PQR	0.6766	0.7032	0.9637	0.9656	**0.9080**	**0.9267**	0.6921	0.7497



(a) AlexNet (b) ResNet50 (c) S_CNN

Faster convergence and smaller std.

Experiments

- Pre-trained models (AlexNet and ResNet) have no advantages on the three legacy databases (LIVE, CSIQ and TID2013)
- Two possible reasons:
  1. The legacy databases: synthetic distortion, limited degradation levels, homogeneous
     LIVE Challenge: complex combinations of distortion, inhomogeneous
  2. Domain transferability: The LIVE Challenge and ImageNet have more statistical similarity than the legacy databases

Experiments

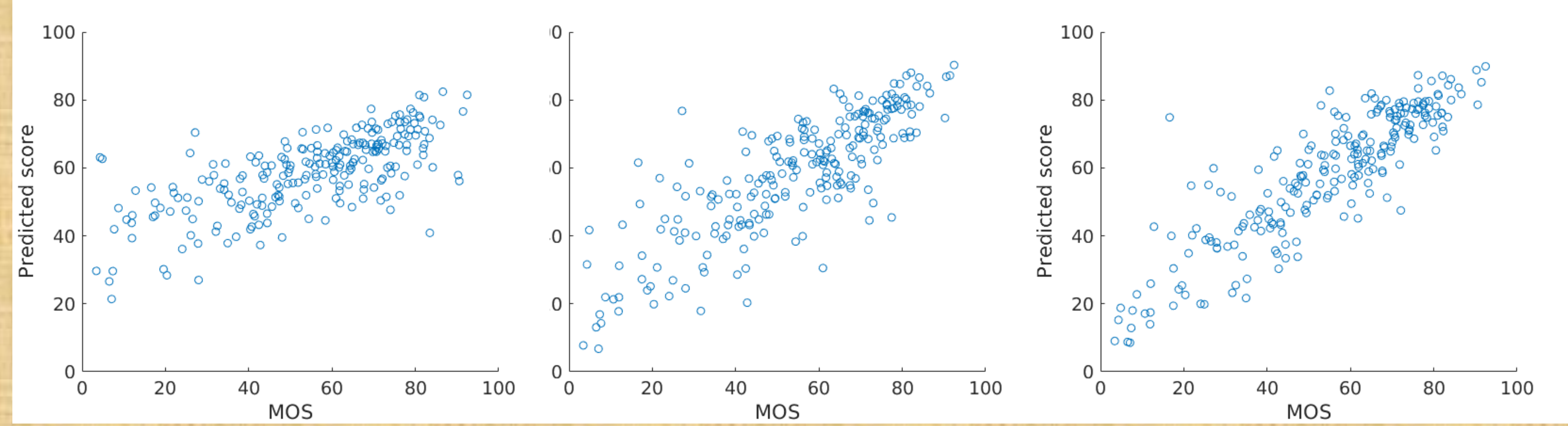
- Comparison with the state-of-the-art on LIVE Challenge

Methods	SRCC	PLCC
DIIVINE [1]	0.58 ± 0.03	0.60 ± 0.03
CORNIA [2]	0.63 ± 0.04	0.66 ± 0.04
BRISQUE [3]	0.61 ± 0.03	0.65 ± 0.04
NIQE [22]	0.43 ± 0.03	0.48 ± 0.03
HOSA [8]	0.66 ± 0.04	0.68 ± 0.03
FRIQUEE-ALL [5]	0.69 ± 0.03	0.71 ± 0.03
Bosse *et al.* [10]*	0.67	0.68
PQR (AlexNet)	0.81 ± 0.01	0.84 ± 0.01
PQR (ResNet50)	$\mathbf{0.86} \pm \mathbf{0.01}$	$\mathbf{0.88} \pm \mathbf{0.01}$
PQR (S_CNN)	0.68 ± 0.03	0.70 ± 0.03

Leading performance on the LIVE Challenge database

Experiments

- Scatter plot of prediction vs. MOS on LIVE-Challenge



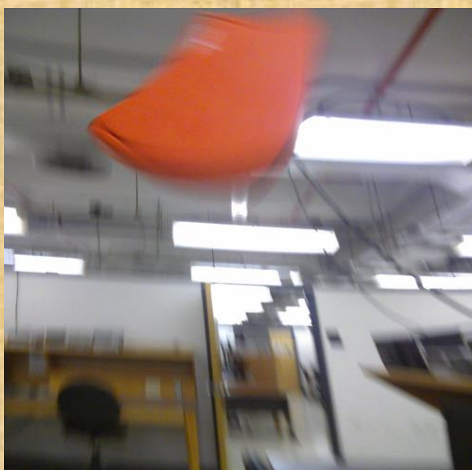
S_CNN
SRCC: 0.6892

AlexNet
SRCC: 0.8165

ResNet50
SRCC: 0.8648

Experiments

- Comparisons on images having very low or high MOS



MOS: 7.4
FRIQUEE: 44.1
S_CNN: 29.6
AlexNet: 16.9
ResNet50: 12.9

MOS: 10.7
FRIQUEE: 48.4
S_CNN: 44.8
AlexNet: 21.3
ResNet50: 17.1

MOS: 74.0
FRIQUEE: 49.2
S_CNN: 47.7
AlexNet: 62.3
ResNet50: 70.9

MOS: 83.5
FRIQUEE: 64.3
S_CNN: 40.9
AlexNet: 84.0
ResNet50: 84.4

Our ResNet50 model makes the most accurate predictions.

More details and toolbox

1. More details and analysis can be found in our technical report (12 page version):

   https://arxiv.org/abs/1708.08190

2. A BIQA Matlab toolbox for easy implementation and fair comparison, containing 8 methods, 3 CNN models and 4 databases:

   https://github.com/HuiZeng/BIQA_Toolbox

Thanks for your time!

Q&A

Technical Report

Toolbox