BLIND IMAGE QUALITY ASSESSMENT VIA CASCaded
MULTI-TASK LEARNING

Zhengfang Duanmu

Department of Electrical and Computer Engineering
University of Waterloo

June. 5, 2017
Outline

1. Motivation
   - Image Quality Assessment (IQA)
   - Deep Learning-based IQA

2. Blind IQA via Cascaded Multi-task Learning
   - Cascaded Multi-task Learning
   - Training

3. Experimental Results
   - Evaluation Criteria
   - Experimental Results

4. Conclusion and Future Work
   - Conclusion
   - Future Work
Outline

1. **Motivation**
   - Image Quality Assessment (IQA)
   - Deep Learning-based IQA

2. **Blind IQA via Cascaded Multi-task Learning**
   - Cascaded Multi-task Learning
   - Training

3. **Experimental Results**
   - Evaluation Criteria
   - Experimental Results

4. **Conclusion and Future Work**
   - Conclusion
   - Future Work
Image Quality Assessment

Why IQA?

Denoiser A  Quality  Denoiser B
Image Quality Assessment (IQA)

IQA category

- Subjective IQA
  - Full-reference IQA
  - Reduce-reference IQA
  - No-reference/Blind IQA

- Objective IQA
  - Predict

IQA category
Image Quality Assessment (IQA)

Full-reference IQA

Reference image → Distortion → Test image → IQA → Quality score
Image Quality Assessment (IQA)

Reduced-reference IQA

Reference image → Feature extraction → Distortion → Test image → Feature extraction → IQA

Quality score
Image Quality Assessment (IQA)

No-reference/Blind IQA (BIQA)

Reference image

Distortion

Test image

IQA

Quality score
Challenge of Deep IQA

Current Status

- The TID2013 [Ponomarenko, 2015] database: \(3,000\) distorted images, \(25\) reference images;
Existing Deep IQA

Solutions

- Transfer learning;
- Patch score assignment;
- FR-IQA learning.
Transfer Learning

Transfer learning: idea

Source labels

Source data
E.g. ImageNet

Large amount of data/labels

Target labels

Target data
E.g. PASCAL

Small amount of data/labels

Transfer Learned Knowledge

[Image from Internet]
Transfer Learning

Limitation

Motorcycle

Classifier

IQA

10

Motorcycle

Classifier

IQA

8

Motorcycle

Classifier

IQA

5

Motorcycle

Classifier

IQA

2

Motorcycle

Classifier

IQA

12 / 42
Motivation
Blind IQA via Cascaded Multi-task Learning
Experimental Results
Conclusion and Future Work

Image Quality Assessment (IQA)
Deep Learning-based IQA

Patch Score Assignment

Limitation

Q = 3

CNN

Q = 3

Q = 3

Q = 3

Q = 3

Q = 3
FR-IQA Learning

Reference image → Distortion → CNN → IQA → Quality score

Test image
Motivation
- Image Quality Assessment (IQA)
- Deep Learning-based IQA

Blind IQA via Cascaded Multi-task Learning
- Cascaded Multi-task Learning
- Training

Experimental Results
- Evaluation Criteria
- Experimental Results

Conclusion and Future Work
- Conclusion
- Future Work
Motivation

Blind IQA via Cascaded Multi-task Learning

Experimental Results

Conclusion and Future Work

Cascaded Multi-task Learning

Training

Strategy

Problem Decomposition

- Subtask I: Identifying distortion type;
- Subtask II: Predicting quality score.

Motivation

- Infinite distorted image (distortion simulation);
- Limited human-labeled image.
Cascaded Multi-task Learning Framework

Traditional Multi-task Learning vs. Cascaded Multi-task Learning

(a)  
Shared layers  \rightarrow Subtask I  \rightarrow Subtask II

(b)  
Shared layers  \rightarrow Subtask I  \rightarrow Subtask II
IQA via Cascaded Multi-task Learning

**Architecture**

- **Shared layers**
  - Convolutional layers: `conv | 5 × 5 | 3 × 8 | 2 | 2`
  - GDN layers
  - Max pooling: `maxpool | 2 × 2`

- **Subtask I**
  - Fully connected layers: `fc | 64 × 128`
  - GDN layers
  - Softmax
  - Cross entropy
  - L1 norm

- **Subtask II**
  - Fully connected layers: `fc | 64 × 256`
  - GDN layers
  - GDN
  - Output: `g(p, s)`
Generalized Divisive Normalization

**Formulation [Balle, 2017]**

\[
y_i(m, n) = \frac{x_i(m, n)}{\left(\beta_i + \sum_{j=1}^{S} \gamma_{ij}x_j(m, n)^2\right)^{\frac{1}{2}}}.
\]

- **S**: depth dimension
- **x(m, n) = (x_1(m, n), \cdots, x_S(m, n))**: linear convolution activation at spatial location (m, n)
- **y(m, n) = (y_1(m, n), \cdots, y_S(m, n))**: normalized activation vector at spatial location (m, n)
- **γ**: symmetric trainable weight matrix
- **β**: trainable bias vector
Generalized Divisive Normalization

Comparison of Rectifiers

<table>
<thead>
<tr>
<th></th>
<th>Nonlinear</th>
<th>Spatially Adaptive</th>
<th>Trainable</th>
<th>Biologically-inspired</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relu [Nair, 2010]</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Batch Norm [Ioffe, 2015]</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>LRN [Krizhevsky, 2012]</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>GDN [Balle, 2017]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Cascaded Multi-task Learning

Training

Cascaded Layer

\[ g(\hat{p}^{(k)}, s^{(k)}) = \hat{p}^{(k)}T s^{(k)} = \sum_{i=1}^{d} \hat{p}_{i}^{(k)} \cdot s_{i}^{(k)}. \] (2)

- \( s^{(k)} \): score vector
- \( \hat{p}^{(k)} \): distortion type probability vector
Loss Functions

\[ \ell_1(\{X^{(k)}\}; W, w_1) = -\sum_{k=1}^{K} \sum_{i=1}^{C} p_i^{(k)} \log \hat{p}_i^{(k)} (X^{(k)}; W, w_1) . \] (3)

\[ \ell_2(\{X^{(k)}\}; W, w_2) = \|q - \hat{q}\|_1 = \sum_{k=1}^{K} |q^{(k)} - \hat{q}^{(k)}| . \] (4)

\[ \ell(\{X^{(k)}\}; W, w_1, w_2) = \ell_1 + \lambda \ell_2 , \] (5)

- \( \{X^{(k)}\} \): \( k \)-th raw input image
- \( W \): Shared parameters
- \( w_1 \): Subtask I-specific parameters
- \( w_2 \): Subtask II-specific parameters
Implementation Details

**Pre-train**
- Batch size: 40
- Learning rate: $10^{-2}$, lowered by a factor of 10 when the loss plateaus, until $10^{-4}$.

**Fine-tune**
- Learning rate: fixed to $10^{-4}$
- $\lambda$: 1
Figure: Sample source images used for pre-training. (a) Human. (b) Animal. (c) Plant. (d) Landscape. (e) Cityscape. (f) Still-life. (g) Transportation. All images are cropped for better visibility.
Distortion Simulation

Pre-train Images

840 (source) × 5 (distortion types) × 5 (distortion levels)
Fine-tune Images

LIVE Image Quality Database [Sheikh, 2006]

- 779 subject-rated images
- train(23 source images and their distorted version)/validate(6 source images and their distorted version)
- 1,000 random splitting, pick the best model
- 4 distortion types
Outline

1 Motivation
   - Image Quality Assessment (IQA)
   - Deep Learning-based IQA

2 Blind IQA via Cascaded Multi-task Learning
   - Cascaded Multi-task Learning
   - Training

3 Experimental Results
   - Evaluation Criteria
   - Experimental Results

4 Conclusion and Future Work
   - Conclusion
   - Future Work
Testing Database

- CSIQ [Larson, 2010]: 866 subject-rated images
- TID2013 [Ponomarenko, 2015]: 3000 subject-rated images
Spearman’s rank-order correlation coefficient (SRCC):

\[
SRCC = 1 - \frac{6 \sum_i d_i^2}{I(I^2 - 1)},
\]

(6)

where \( I \) is the test image number and \( d_i \) is the rank difference between the MOS and the model prediction of the \( i \)-th image.

Pearson linear correlation coefficient (PLCC):

\[
PLCC = \frac{\sum_i (q_i - \bar{q})(s_i - \bar{s})}{\sqrt{\sum_i (q_i - \bar{q})^2} \sqrt{\sum_i (s_i - \bar{s})^2}},
\]

(7)

where \( q_i \) and \( s_i \) stand for the MOS and the model prediction of the \( i \)-th image, respectively.
### Experimental Results on CSIQ

**Table: SRCC results on CSIQ**

<table>
<thead>
<tr>
<th>SRCC</th>
<th>JP2K</th>
<th>JPEG</th>
<th>WN</th>
<th>BLUR</th>
<th>ALL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIIVINE [Moorthy, 2011]</td>
<td>0.844</td>
<td>0.819</td>
<td>0.881</td>
<td>0.884</td>
<td>0.835</td>
</tr>
<tr>
<td>BRISQUE [Mittal, 2012]</td>
<td>0.894</td>
<td>0.916</td>
<td><strong>0.934</strong></td>
<td>0.915</td>
<td>0.909</td>
</tr>
<tr>
<td>CORNIA [Ye, 2012]</td>
<td>0.916</td>
<td>0.919</td>
<td>0.787</td>
<td><strong>0.928</strong></td>
<td>0.914</td>
</tr>
<tr>
<td>ILNIQE [Zhang, 2015]</td>
<td>0.924</td>
<td>0.905</td>
<td>0.867</td>
<td>0.867</td>
<td>0.887</td>
</tr>
<tr>
<td>BLISS [Ye, 2014]</td>
<td><strong>0.932</strong></td>
<td>0.927</td>
<td>0.879</td>
<td>0.922</td>
<td>0.920</td>
</tr>
<tr>
<td>HOSA [Xu, 2016]</td>
<td>0.920</td>
<td>0.918</td>
<td>0.895</td>
<td>0.915</td>
<td>0.918</td>
</tr>
<tr>
<td>dipIQ [Ma, 2017]</td>
<td><strong>0.944</strong></td>
<td><strong>0.936</strong></td>
<td>0.904</td>
<td><strong>0.932</strong></td>
<td><strong>0.930</strong></td>
</tr>
<tr>
<td>deepIQA [Bosse, 2017]</td>
<td>0.907</td>
<td>0.929</td>
<td>0.933</td>
<td>0.890</td>
<td>0.871</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.898</td>
<td><strong>0.948</strong></td>
<td><strong>0.951</strong></td>
<td>0.918</td>
<td><strong>0.932</strong></td>
</tr>
</tbody>
</table>
## Experimental Results on CSIQ

<table>
<thead>
<tr>
<th>PLCC</th>
<th>JP2K</th>
<th>JPEG</th>
<th>WN</th>
<th>BLUR</th>
<th>ALL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIIVINE [Moorthy, 2011]</td>
<td>0.898</td>
<td>0.818</td>
<td>0.903</td>
<td>0.909</td>
<td>0.855</td>
</tr>
<tr>
<td>BRISQUE [Mittal, 2012]</td>
<td>0.937</td>
<td>0.960</td>
<td>0.947</td>
<td>0.936</td>
<td>0.937</td>
</tr>
<tr>
<td>CORNIA [Ye, 2012]</td>
<td>0.947</td>
<td>0.960</td>
<td>0.777</td>
<td>0.953</td>
<td>0.934</td>
</tr>
<tr>
<td>ILNIQE [Zhang, 2015]</td>
<td>0.942</td>
<td>0.956</td>
<td>0.880</td>
<td>0.903</td>
<td>0.914</td>
</tr>
<tr>
<td>BLISS [Ye, 2014]</td>
<td>0.954</td>
<td>0.970</td>
<td>0.895</td>
<td>0.947</td>
<td>0.939</td>
</tr>
<tr>
<td>HOSA [Xu, 2016]</td>
<td>0.946</td>
<td>0.958</td>
<td>0.912</td>
<td>0.940</td>
<td>0.942</td>
</tr>
<tr>
<td>dipIQ [Ma, 2017]</td>
<td>0.959</td>
<td>0.975</td>
<td>0.927</td>
<td>0.958</td>
<td>0.949</td>
</tr>
<tr>
<td>deepIQA [Bosse, 2017]</td>
<td>0.931</td>
<td>0.951</td>
<td>0.933</td>
<td>0.906</td>
<td>0.891</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.925</td>
<td>0.979</td>
<td>0.958</td>
<td>0.946</td>
<td>0.944</td>
</tr>
</tbody>
</table>
### Experimental Results on TID2013

#### Table: SRCC results on TID2013

<table>
<thead>
<tr>
<th>SRCC</th>
<th>JP2K</th>
<th>JPEG</th>
<th>WN</th>
<th>BLUR</th>
<th>ALL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIIVINE [Moorthy, 2011]</td>
<td>0.857</td>
<td>0.680</td>
<td>0.879</td>
<td>0.859</td>
<td>0.795</td>
</tr>
<tr>
<td>BRISQUE [Mittal, 2012]</td>
<td>0.906</td>
<td>0.894</td>
<td>0.889</td>
<td>0.886</td>
<td>0.883</td>
</tr>
<tr>
<td>CORNIA [Ye, 2012]</td>
<td>0.907</td>
<td>0.912</td>
<td>0.798</td>
<td><strong>0.934</strong></td>
<td>0.893</td>
</tr>
<tr>
<td>ILNIQE [Zhang, 2015]</td>
<td>0.912</td>
<td>0.873</td>
<td>0.890</td>
<td>0.815</td>
<td>0.881</td>
</tr>
<tr>
<td>BLISS [Ye, 2014]</td>
<td>0.906</td>
<td>0.893</td>
<td>0.856</td>
<td>0.872</td>
<td>0.836</td>
</tr>
<tr>
<td>HOSA [Xu, 2016]</td>
<td><strong>0.933</strong></td>
<td>0.917</td>
<td>0.843</td>
<td>0.921</td>
<td><strong>0.904</strong></td>
</tr>
<tr>
<td>dipIQ [Ma, 2017]</td>
<td>0.926</td>
<td><strong>0.932</strong></td>
<td>0.905</td>
<td><strong>0.922</strong></td>
<td>0.877</td>
</tr>
<tr>
<td>deepIQA [Bosse, 2017]</td>
<td><strong>0.948</strong></td>
<td><strong>0.921</strong></td>
<td><strong>0.938</strong></td>
<td>0.910</td>
<td>0.885</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.911</td>
<td>0.919</td>
<td><strong>0.908</strong></td>
<td>0.891</td>
<td><strong>0.912</strong></td>
</tr>
</tbody>
</table>
### Experimental Results

#### Experimental Results on TID2013

**Table:** PLCC results on TID2013

<table>
<thead>
<tr>
<th>PLCC</th>
<th>JP2K</th>
<th>JPEG</th>
<th>WN</th>
<th>BLUR</th>
<th>ALL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIIIVINE [Moorthy, 2011]</td>
<td>0.901</td>
<td>0.696</td>
<td>0.882</td>
<td>0.860</td>
<td>0.794</td>
</tr>
<tr>
<td>BRISQUE [Mittal, 2012]</td>
<td>0.919</td>
<td>0.950</td>
<td>0.886</td>
<td>0.884</td>
<td>0.900</td>
</tr>
<tr>
<td>CORNIA [Ye, 2012]</td>
<td>0.928</td>
<td>0.960</td>
<td>0.778</td>
<td></td>
<td>0.934</td>
</tr>
<tr>
<td>ILNIQE [Zhang, 2015]</td>
<td>0.929</td>
<td>0.944</td>
<td>0.899</td>
<td>0.816</td>
<td>0.890</td>
</tr>
<tr>
<td>BLISS [Ye, 2014]</td>
<td>0.930</td>
<td>0.963</td>
<td>0.863</td>
<td>0.872</td>
<td>0.862</td>
</tr>
<tr>
<td>HOSA [Xu, 2016]</td>
<td><strong>0.952</strong></td>
<td>0.949</td>
<td>0.842</td>
<td>0.921</td>
<td><strong>0.918</strong></td>
</tr>
<tr>
<td>dipIQ [Ma, 2017]</td>
<td>0.948</td>
<td><strong>0.973</strong></td>
<td>0.906</td>
<td><strong>0.928</strong></td>
<td>0.894</td>
</tr>
<tr>
<td>deepIQA [Bosse, 2017]</td>
<td><strong>0.963</strong></td>
<td>0.960</td>
<td><strong>0.943</strong></td>
<td>0.897</td>
<td><strong>0.913</strong></td>
</tr>
<tr>
<td>Proposed</td>
<td>0.924</td>
<td><strong>0.969</strong></td>
<td><strong>0.911</strong></td>
<td>0.899</td>
<td>0.912</td>
</tr>
</tbody>
</table>
Confusion Matrix

Table: The confusion matrices produced by our method on CSIQ and TID2013. The column and the raw contain ground truth and predicted distortion types, respectively.

<table>
<thead>
<tr>
<th></th>
<th>JP2K</th>
<th>JPEG</th>
<th>WN</th>
<th>BLUR</th>
<th>Pristine</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSIQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JP2K</td>
<td>0.847</td>
<td>0.007</td>
<td>0.000</td>
<td>0.093</td>
<td>0.053</td>
</tr>
<tr>
<td>JPEG</td>
<td>0.040</td>
<td>0.820</td>
<td>0.000</td>
<td>0.027</td>
<td>0.113</td>
</tr>
<tr>
<td>WN</td>
<td>0.000</td>
<td>0.000</td>
<td>0.947</td>
<td>0.013</td>
<td>0.040</td>
</tr>
<tr>
<td>BLUR</td>
<td>0.067</td>
<td>0.006</td>
<td>0.000</td>
<td>0.827</td>
<td>0.100</td>
</tr>
<tr>
<td>Pristine</td>
<td>0.067</td>
<td>0.000</td>
<td>0.100</td>
<td>0.166</td>
<td>0.667</td>
</tr>
<tr>
<td>TID2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JP2K</td>
<td>0.944</td>
<td>0.016</td>
<td>0.000</td>
<td>0.040</td>
<td>0.000</td>
</tr>
<tr>
<td>JPEG</td>
<td>0.032</td>
<td>0.968</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>WN</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>BLUR</td>
<td>0.088</td>
<td>0.008</td>
<td>0.000</td>
<td>0.848</td>
<td>0.056</td>
</tr>
<tr>
<td>Pristine</td>
<td>0.160</td>
<td>0.000</td>
<td>0.040</td>
<td>0.000</td>
<td>0.800</td>
</tr>
</tbody>
</table>
### Ablation Experiments

**Table:** SRCC results of ablation experiments on CSIQ and TID2013

<table>
<thead>
<tr>
<th>Experiment</th>
<th>CSIQ</th>
<th>TID2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single task w/o pre-training</td>
<td>0.844</td>
<td>0.850</td>
</tr>
<tr>
<td>Traditional multi-task w/o pre-training</td>
<td>0.885</td>
<td>0.871</td>
</tr>
<tr>
<td>Cascaded multi-task w/o pre-training</td>
<td>0.894</td>
<td>0.880</td>
</tr>
<tr>
<td>Single-task with pre-training</td>
<td>0.923</td>
<td>0.911</td>
</tr>
<tr>
<td>Traditional multi-task with pre-training</td>
<td>0.930</td>
<td>0.905</td>
</tr>
<tr>
<td>Proposed</td>
<td><strong>0.932</strong></td>
<td><strong>0.912</strong></td>
</tr>
</tbody>
</table>
Effect of GDN

Table: Analyzing the effect of GDN on reducing the model complexity in comparison with ReLU.

<table>
<thead>
<tr>
<th>Model</th>
<th>CSIQ</th>
<th>TID2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>ReLU</td>
<td>0.922</td>
<td>0.891</td>
</tr>
<tr>
<td>ReLU + double layers</td>
<td>0.924</td>
<td>0.900</td>
</tr>
<tr>
<td>ReLU + double layers + BN</td>
<td>0.930</td>
<td>0.918</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.932</td>
<td>0.912</td>
</tr>
</tbody>
</table>
Motivation
- Image Quality Assessment (IQA)
- Deep Learning-based IQA

Blind IQA via Cascaded Multi-task Learning
- Cascaded Multi-task Learning
- Training

Experimental Results
- Evaluation Criteria
- Experimental Results

Conclusion and Future Work
- Conclusion
- Future Work
**Conclusion**

**Contributions**

- Presented a cascaded multi-task learning framework for BIQA
- Demonstrated state-of-the-art performance
Future Work

- More distortion types
- More image databases
- Video quality assessment
Thank you
References


References


