Blind Image Quality Assessment via Cascaded Multi-task Learning

Zhengfang Duanmu

Department of Electrical and Computer Engineering
University of Waterloo

June. 5, 2017

Outline

Outline

Image Quality Assessment

Why IQA?

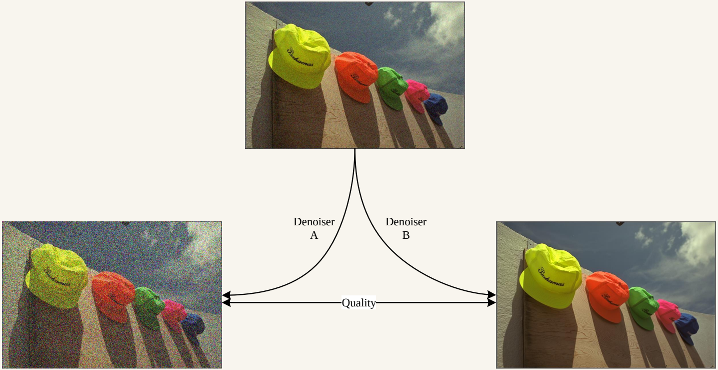


Image Quality Assessment (IQA)

IQA category

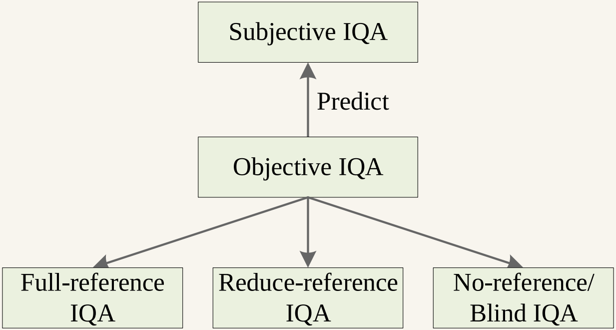


Image Quality Assessment (IQA)

Full-reference IQA

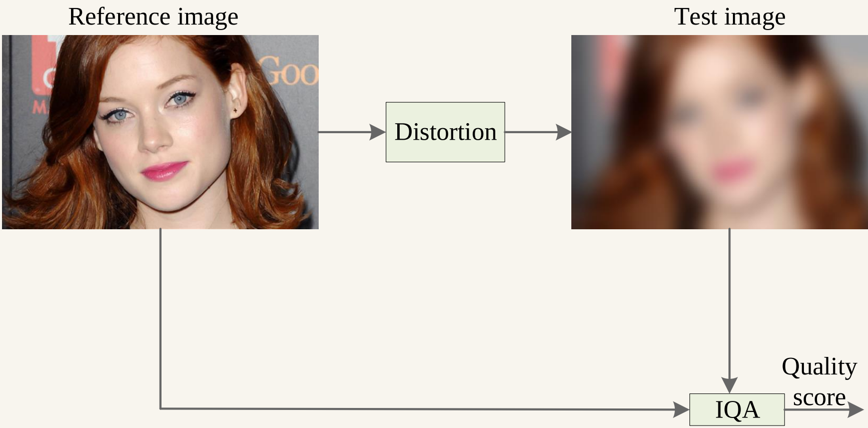


Image Quality Assessment (IQA)

Reduced-reference IQA

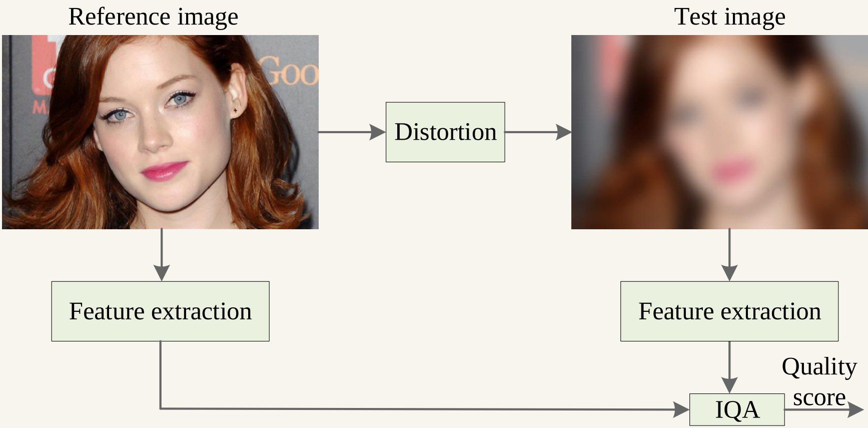
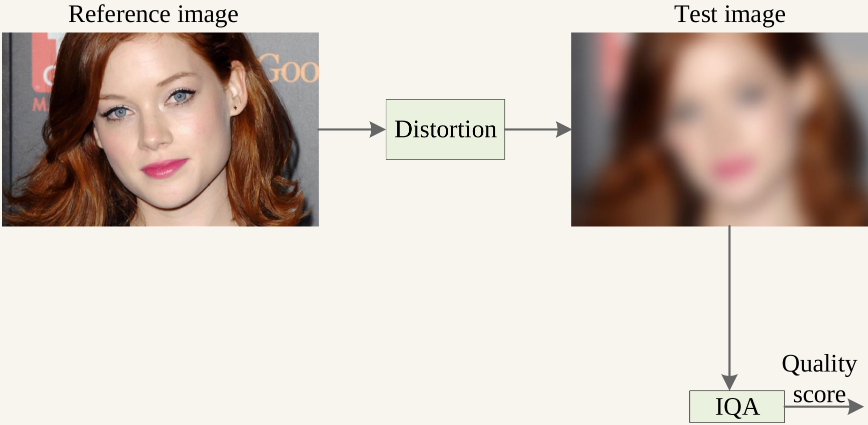


Image Quality Assessment (IQA)

No-reference/Blind IQA (BIQA)



Challenge of Deep IQA

Current Status

- The TID2013 [Ponomarenko, 2015] database: 3, 000 distorted images, 25 reference images;
- The ImageNet [Deng, 2009] database: 14, 197, 122 images.

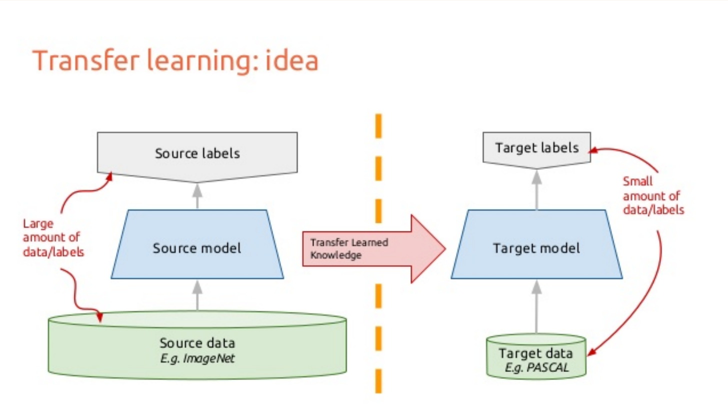
Existing Deep IQA

Solutions

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

Transfer Learning

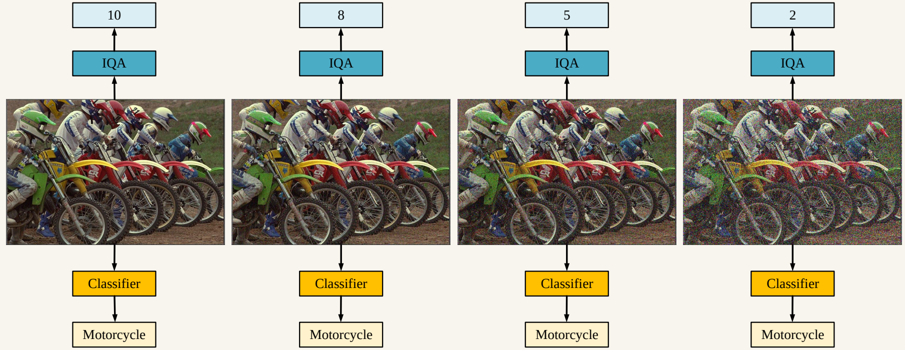
Transfer Learning



[Image from Internet]

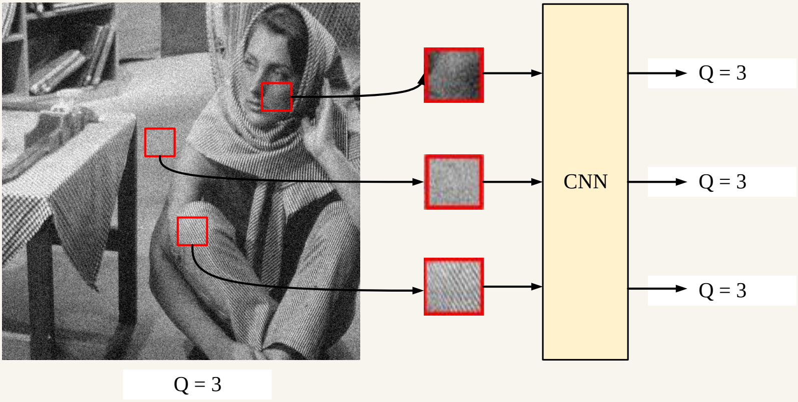
Transfer Learning

Limitation



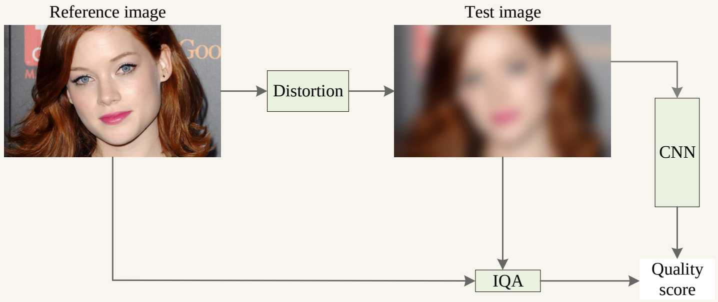
Patch Score Assignment

Limitation



FR-IQA Learning

FR-IQA Learning



Outline

Strategy

Problem Decomposition

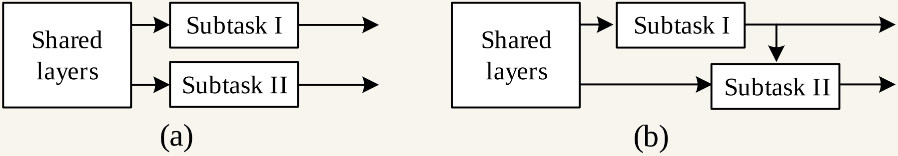
- Subtask I: Identifying distortion type;
- Subtask II: Predicting quality score.

Motivation

- Infinite distorted image (distortion simulation);
- Limited human-labeled image.

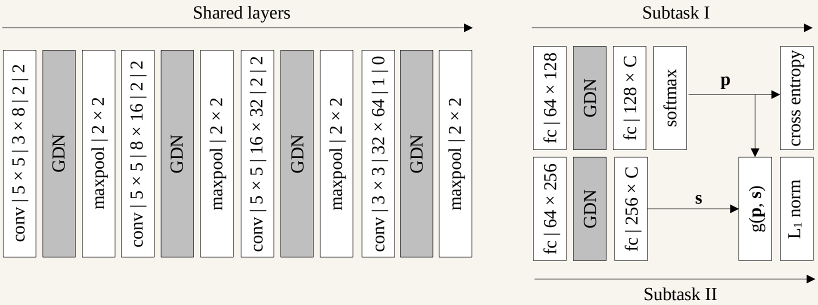
Cascaded Multi-task Learning Framework

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



IQA via Cascaded Multi-task Learning

Architecture



Generalized Divisive Normalization

Formulation [Balle, 2017]

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

- S: depth dimension
- $\mathbf{x}(m,n) = (x_1(m,n), \cdots, x_S(m,n))$: linear convolution activation at spatial location (m,n)
- $\mathbf{y}(m,n) = (y_1(m,n), \cdots, y_S(m,n))$: normalized activation vector at spatial location (m,n)
- $\boldsymbol{\gamma}$: symmetric trainable weight matrix
- $\boldsymbol{\beta}$: trainable bias vector

Generalized Divisive Normalization

Comparison of Rectifiers

Table: Comparison of Rectifiers

	Nonlinear	Spatially Adaptive	Trainable	Biologically-inspired
Relu [Nair, 2010]	✓	✗	✗	✗
Batch Norm [Ioffe, 2015]	✗	✗	✓	✗
LRN [Krizhevsky, 2012]	✓	✓	✗	✓
GDN [Balle, 2017]	✓	✓	✓	✓

Cascaded Layer

Cascaded Layer

$$g(\hat{\mathbf{p}}^{(k)}, \mathbf{s}^{(k)}) = \hat{\mathbf{p}}^{(k)T}\mathbf{s}^{(k)} = \sum_{i=1}^{d} \hat{p}_i^{(k)} \cdot s_i^{(k)} . \tag{2}$$

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

Loss Functions

Loss Functions

$$\ell_1(\{\mathbf{X}^{(k)}\};\mathbf{W},\mathbf{w}_1) = -\sum_{k=1}^{K}\sum_{i=1}^{C} p_i^{(k)} \log \hat{p}_i^{(k)}(\mathbf{X}^{(k)};\mathbf{W},\mathbf{w}_1)\,. \tag{3}$$

$$\ell_2(\{\mathbf{X}^{(k)}\};\mathbf{W},\mathbf{w}_2) = \|\mathbf{q}-\hat{\mathbf{q}}\|_1 = \sum_{k=1}^{K} |q^{(k)} - \hat{q}^{(k)}|\,. \tag{4}$$

$$\ell(\{\mathbf{X}^{(k)}\};\mathbf{W},\mathbf{w}_1,\mathbf{w}_2) = \ell_1 + \lambda\ell_2\,, \tag{5}$$

- $\{\mathbf{X}^{(k)}\}$: k-th raw input image
- $\mathbf{W}$: Shared parameters
- $\mathbf{w}_1$: Subtask I-specific parameters
- $\mathbf{w}_2$: 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}
- λ: 1

Pre-train Images

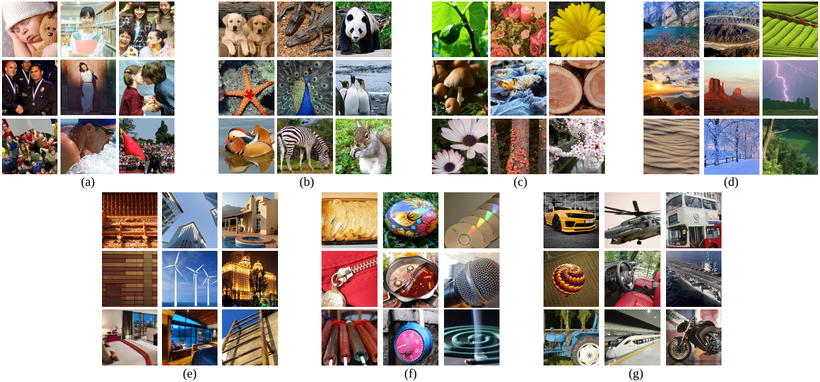
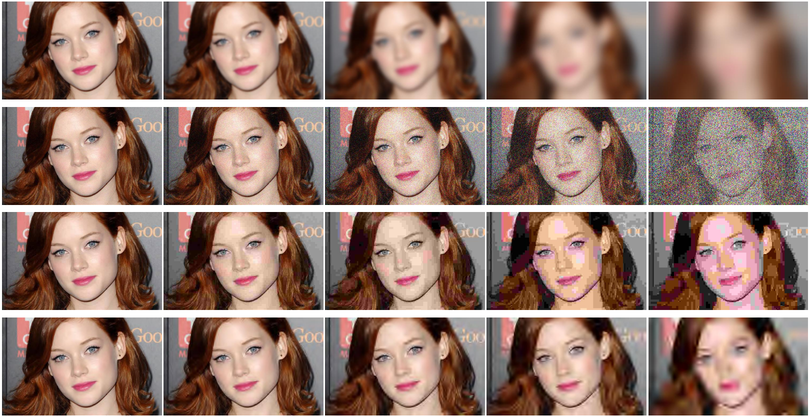


Figure: Sample source images used for pre-training. (a) Human. (b) Animal. (c) Plant. (d) Landscape. (e) Cityscape. (f) Still-life. (g) Transportation. All images are cropped for better visibility.

Distortion Simulation



Pre-train Images

840 (source) $\times$ 5 (distortion types) $\times$ 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

Testing Database

Testing Database

- CSIQ [Larson, 2010]: 866 subject-rated images
- TID2013 [Ponomarenko, 2015]: 3000 subject-rated images

Evaluation Criteria

Evaluation Criteria

- Spearman's rank-order correlation coefficient (SRCC):

$$\mathrm{SRCC} = 1 - \frac{6\sum_i d_i^2}{I(I^2-1)}, \tag{6}$$

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):

$$\mathrm{PLCC} = \frac{\sum_i (q_i - \bar{q})(s_i - \bar{s})}{\sqrt{\sum_i (q_i - \bar{q})^2}\sqrt{\sum_i (s_i - \bar{s})^2}}, \tag{7}$$

where q_i and s_i stand for the MOS and the model prediction of the i-th image, respectively.

Experimental Results

Experimental Results on CSIQ

Table: SRCC results on CSIQ

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

Experimental Results

Experimental Results on CSIQ

Table: PLCC results on CSIQ

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

Experimental Results

Experimental Results on TID2013

Table: SRCC results on TID2013

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

Experimental Results

Experimental Results on TID2013

Table: PLCC results on TID2013

PLCC	JP2K	JPEG	WN	BLUR	ALL4
DIIVINE [Moorthy, 2011]	0.901	0.696	0.882	0.860	0.794
BRISQUE [Mittal, 2012]	0.919	0.950	0.886	0.884	0.900
CORNIA [Ye, 2012]	0.928	0.960	0.778	**0.934**	0.904
ILNIQE [Zhang, 2015]	0.929	0.944	0.899	0.816	0.890
BLISS [Ye, 2014]	0.930	0.963	0.863	0.872	0.862
HOSA [Xu, 2016]	**0.952**	0.949	0.842	0.921	**0.918**
dipIQ [Ma, 2017]	0.948	**0.973**	0.906	**0.928**	0.894
deepIQA [Bosse, 2017]	**0.963**	0.960	**0.943**	0.897	**0.913**
Proposed	0.924	**0.969**	**0.911**	0.899	0.912

Experimental Results

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

Accuracy		JP2K	JPEG	WN	BLUR	Pristine
CSIQ	JP2K	**0.847**	0.007	0.000	0.093	0.053
	JPEG	0.040	**0.820**	0.000	0.027	0.113
	WN	0.000	0.000	**0.947**	0.013	0.040
	BLUR	0.067	0.006	0.000	**0.827**	0.100
	Pristine	0.067	0.000	0.100	0.166	**0.667**
TID2013	JP2K	**0.944**	0.016	0.000	0.040	0.000
	JPEG	0.032	**0.968**	0.000	0.000	0.000
	WN	0.000	0.000	**1.000**	0.000	0.000
	BLUR	0.088	0.008	0.000	**0.848**	0.056
	Pristine	0.160	0.000	0.040	0.000	**0.800**

Experimental Results

Ablation Experiments

Table: SRCC results of ablation experiments on CSIQ and TID2013

	CSIQ	TID2013
Single task w/o pre-training	0.844	0.850
Traditional multi-task w/o pre-training	0.885	0.871
Cascaded multi-task w/o pre-training	0.894	0.880
Single-task with pre-training	0.923	0.911
Traditional multi-task with pre-training	0.930	0.905
Proposed	**0.932**	**0.912**

Experimental Results

Effect of GDN

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

	CSIQ	TID2013
ReLU	0.922	0.891
ReLU + double layers	0.924	0.900
ReLU + double layers + BN	0.930	**0.918**
Proposed	**0.932**	0.912

Outline

Conclusion

Contributions

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

Future Work

Future Work

- More distortion types
- More image databases
- Video quality assessment

Thank you

References

N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Processing:Image Communication*, 2015.

J. Deng *et al.*, "ImageNet: A large-scale hierarchical image database," In *CVPR*, 2009.

A. Moorthy and A. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality,"*IEEE TIP*, 2011.

A. Mittal *et al.*, "Making a "completely blind" image quality analyzer," *IEEE SPL*, 2013.

P. Ye *et al.*, "Unsupervised feature learning framework for no-reference image quality assessment," In *CVPR*, 2012.

L. Zhang *et al.*, "A feature-enriched completely blind image quality evaluator," *IEEE TIP*, 2015.

H. R. Sheikh *et al.*, Image and video quality assessment research at LIVE [Online]. Available: http://live.ece.utexas.edu/research/quality/.

E. Larson and D. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *JEI*, 2010.

A. Mittal *et al.*, "No-reference image quality assessment in the spatial domain," *IEEE TIP*, 2012.

References

P. Ye *et al.* "Beyond human opinion scores: Blind image quality assessment based on synthetic scores," *CVPR*, 2014.

P. Ye *et al.* "Blind image quality assessment based on high order statistics aggregation," *IEEE TIP*, 2016.

K. Ma *et al.* "dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE TIP*, 2017.

S. Bosse *et al.* "Deep neural networks for no-reference and full-reference image quality assessment," *CoRR*, 2016.

J. Balle *et al.* "End-to-end optimized image compression," *ICLR*, 2017.

S. Ioffe *et al.* "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, 2015.

V. Nair *et al.* "Rectified linear units improve restricted boltzmann machines," *ICML*, 2010.

A. Krizhevsky *et al.* "ImageNet classification with deep convolutional neural networks," *NIPS*, 2012.