भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Generating Image Distortion Maps Using Convolutional Autoencoders with Application to No Reference Image Quality Assessment

Sumohana S. Channappayya
IIT Hyderabad

@ AIP-IITH Joint Workshop on Machine Learning and Applications
IIT Hyderabad

Acknowledgments

1. Students: Dendi Sathya Veera Reddy (EE PhD Scholar), Chander Dev (EE BTech), Narayan Kothari (EE BTech)
2. Drs. Srijith and Vineeth for the invitation

Introduction and Motivation

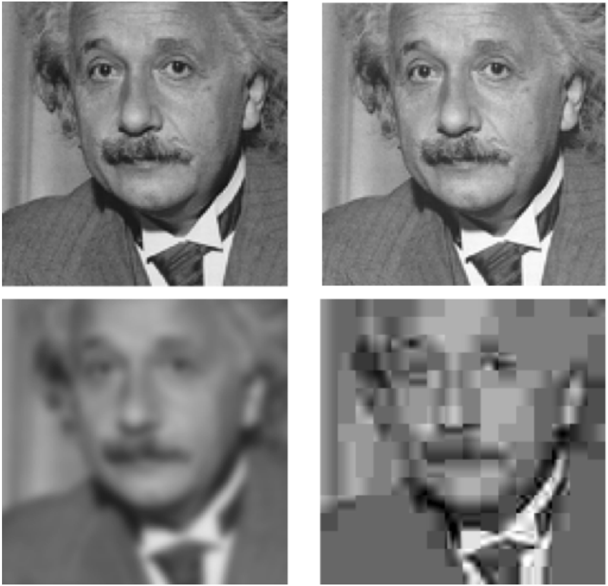


Image Quality Assessment – The Why

What's wrong with using MSE for IQA?

- **Poor correlation** with mean opinion score (MOS) of subjective evaluation.
- **Global** measure of error.

Why is MOS important?

- Majority of multimedia content intended for **human consumption**.
- **Gold standard** for quality evaluation.

Why not use MOS then?

- Expensive, time-consuming (non-real-time), large data volume.

Image Quality Assessment – The Why

An important problem for both the academia and the industry.

- An open research problem with several flavors!
- Immediate practical applications with economic impact.

Image Quality Assessment – The How

Flavors of Image Quality Assessment:

- *Full reference (FR):* **Pristine reference image** and **image under evaluation** are both available.
- *Reduced reference (RR):* **Partial information** about pristine reference image and **test image** available for comparison.
- *No reference/Blind (NR/B):* **Only test image available!**

Assumption: Working with **natural scenes** meant for **human consumption**.

Image Quality Assessment – The How

The turning point in FR – The Structural Similarity (SSIM) Index [1].

- *Hypothesis:* distortion affects local structure of images.
- Modern, successful approach: measure **loss** of **structure** in a distorted image.
- Basic idea: combine local measures of similarity of **luminance, contrast, structure** into local measure of quality.

$$\mathrm{SSIM}_{I,J}(i,j) = L_{I,J}(i,j)C_{I,J}(i,j)S_{I,J}(i,j) \text{ where}$$

- Perform **weighted average** of local measure across image.

Image Quality Assessment – SSIM Map

- Displaying SSIM(i, j) as an image is called an **SSIM Map**. It is an effective way of **visualizing** where the images I, J differ.
- The **SSIM map** depicts where the quality of one image differs from the other.

Correlation (SROCC) with DMOS on LIVE dataset – PSNR (L samples): 0.8754, SSIM: 9129.

Image Quality Assessment – SSIM Map Example

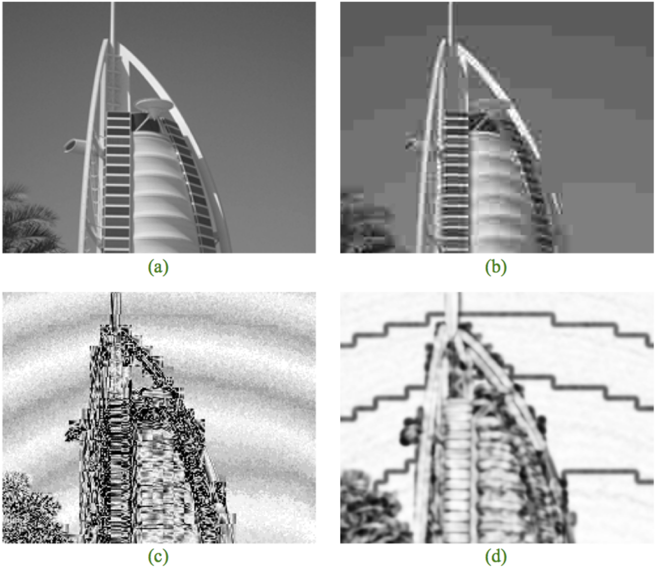


Figure: a: Reference; b: JPEG; c: Absolute diff; d: SSIM map

Image Quality Assessment – SSIM Map Example

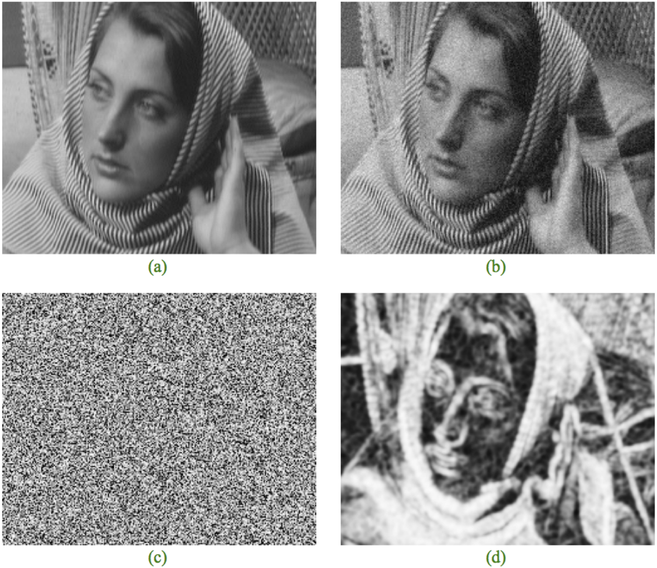
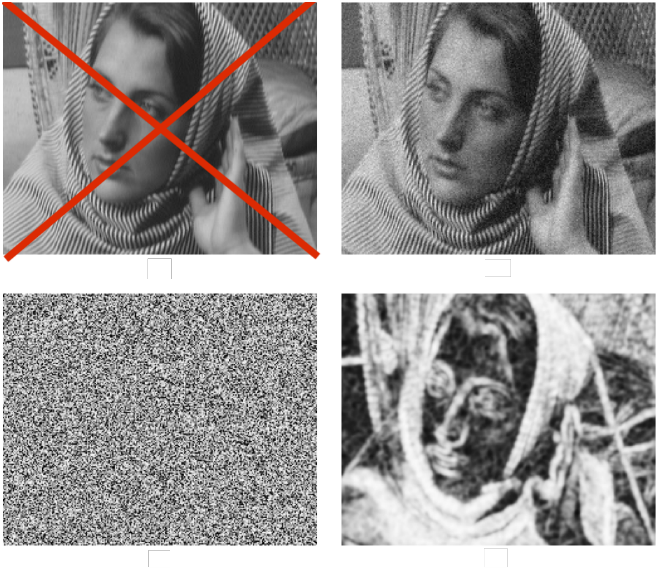


Figure: a: Reference; b: AWGN; c: Absolute diff; d: SSIM map

No-reference Image Quality Assessment



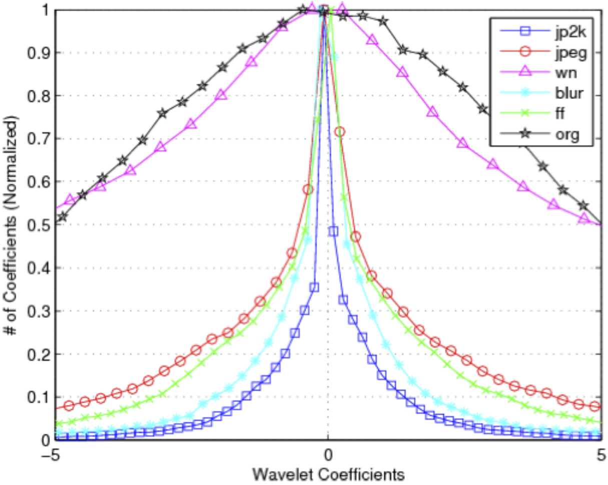
No-reference or Blind Image Quality Assessment (NR/BIQA)

- **Pristine reference image not available** for comparison.
- **Distortion information** used.
- **Opinion information** used.
- An open problem

Representative Examples of No-reference Image Quality Assessment

- Unsupervised Learning: Natural Image Quality Evaluator (NIQE) [2]
- Supervised Learning: Convolutional Neural Networks for No-Reference Image Quality Assessment [3]

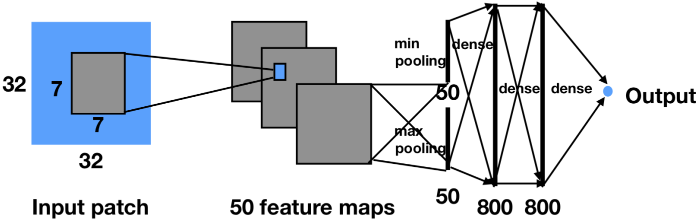
Unsupervised Learning: Natural Image Quality Evaluator (NIQE) [2]



1

[1] Source: Moorthy and Bovik, IEEE TIP 2011.

Supervised Learning: Convolutional Neural Networks for No-Reference Image Quality Assessment [3]



Architecture of the CNN used for NRIQA

Challenges in NRIQA

- **Databases are small** compared to typical computer vision databases
- Constructing large databases is **challenging**
- Standard databases employ **synthetic distortions**
- Databases with **realistic distortions are few**
- Realistic distortions mean reference images (and scores) not available
- Generation of **localized distortion maps**

Proposed Approach: Distortion Map Generation

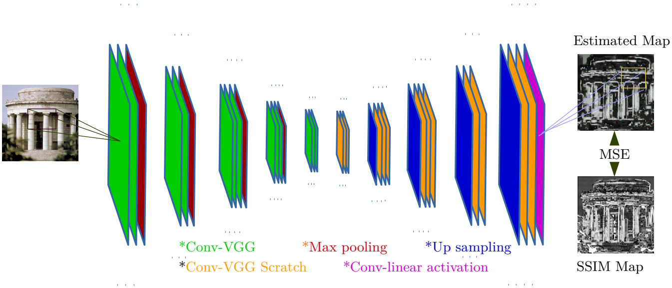
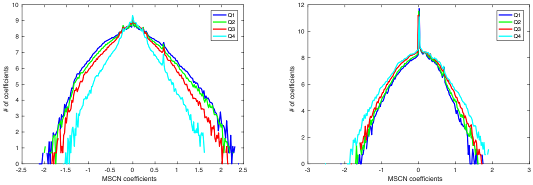


Figure: Architecture of DistNet

Proposed Approach: NRIQA using Distortion Map

- Approach 1: Simple weighted averaging
- Approach 2: Statistical modeling of normalized map coefficients and supervised learning

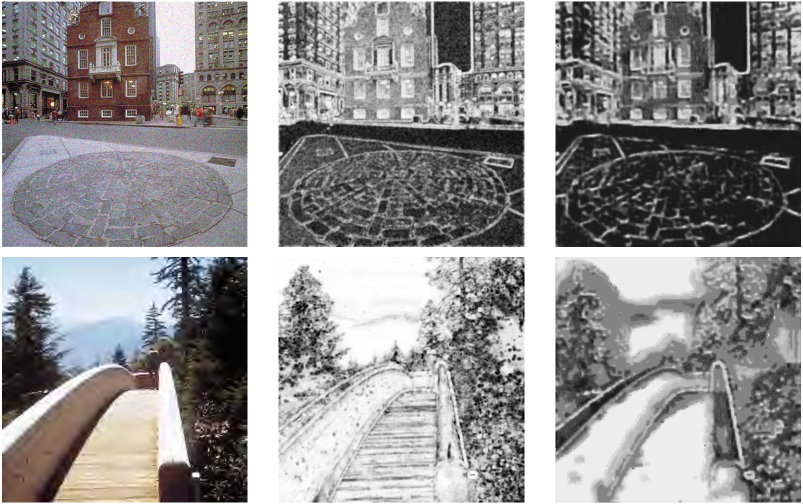


- Approach 3: Supervised learning using spatial statistics [11] plus average map score

Implementation Details

- DistNet
  - 120 natural images
  - Distortions: JPEG, JP2K, AWGN, Gaussian blur. 5 levels each
  - 2400 distorted images and corresponding SSIM maps used for training and validation (80:20)
  - Preprocessing: mean subtraction and variance normalization
- NRIQA
  - Evaluated over 7 IQA databases: 5 synthetic distortions and 2 authentic distortions
  - Performance evaluated using linear correlation coefficient (LCC) and rank ordered correlation coefficient (SROCC)

Results: DistNet



Results: NRIQA

	LIVE II [4]		CSIQ [5]		TID 2013 [6]		LIVE MD [7]		MDID 2013 [8]	
	LCC	**SRCC**	**LCC**	**SRCC**	**LCC**	**SRCC**	**LCC**	**SRCC**	**LCC**	**SRCC**
NFERM [9]	**0.95**	0.94	0.78	0.70	0.50	0.36	**0.94**	**0.92**	0.90	0.89
BLIINDS-II [10]	0.93	0.92	0.83	0.78	0.61	0.53	0.92	0.91	**0.92**	**0.91**
BRISQUE [11]	0.94	0.94	0.82	0.77	0.54	0.47	0.93	0.90	0.89	0.87
DIIVINE [12]	0.89	0.88	0.79	0.76	0.60	0.51	0.72	0.66	0.45	0.45
NIQE [2]	0.91	0.91	0.71	0.62	0.43	0.32	0.77	0.84	0.57	0.57
IL-NIQE [13]	0.91	0.90	0.85	0.81	0.65	0.52	0.88	0.89	0.51	0.52
QAC [3]	0.87	0.87	0.66	0.55	0.49	0.39	0.66	0.47	0.15	0.19
DistNet-Q1	0.88	0.86	0.80	0.79	0.30	0.30	0.60	0.55	0.44	0.38
DistNet-Q2	0.91	0.92	0.87	0.85	0.69	0.62	0.91	0.84	0.87	0.85
DistNet-Q3	**0.95**	**0.95**	**0.91**	**0.88**	**0.82**	**0.79**	0.89	0.84	0.90	0.89

Results: NRIQA Performance on Authentic Distortions

	LIVE Wild [14]		KonIQ-10K [15]	
	LCC	**SRCC**	**LCC**	**SRCC**
NFERM [9]	0.42	0.32	0.25	0.24
BLIINDS-II [10]	0.48	0.45	0.58	0.57
BRISQUE [11]	**0.60**	0.56	0.70	**0.70**
DIIVINE [12]	0.47	0.43	0.62	0.58
NIQE [2]	0.47	0.45	0.55	0.54
IL-NIQE [13]	0.51	0.43	0.53	0.50
QAC [3]	0.32	0.24	0.37	0.34
DistNet-Q1	0.30	0.24	0.25	0.21
DistNet-Q2	0.51	0.48	0.60	0.59
DistNet-Q3	**0.60**	**0.57**	**0.71**	**0.70**

Results: NRIQA

Dataset	Distortion Type	NIQE [2]	QAC [3]	IL-NIQE [13]	DistNet -Q1
TID13 [6]	AWGN	0.82	0.74	**0.88**	0.86
	AWGNC	0.67	0.72	**0.86**	0.78
	SCN	0.67	0.17	**0.92**	0.71
	MN	**0.75**	0.59	0.51	0.56
	HFN	0.84	0.86	0.87	**0.87**
	IN	0.74	**0.80**	0.75	0.72
	QN	0.85	0.71	**0.87**	0.58
	GB	0.79	**0.85**	0.81	0.84
	ID	0.59	0.34	**0.75**	0.32
	JPEG	0.84	0.84	0.83	**0.89**
	JP2K	**0.89**	0.79	0.86	0.77

Concluding Remarks

- Reference-less distortion map estimation
- Application to NRIQA
- Opens up several other potential applications such as NRVQA
- Better distortion map estimation techniques can be explored
- Accepted to *IEEE Signal Processing Letters*

Key References

1. Wang et al., Image Quality Assessment: From Error Visibility to Structural Similarity, IEEE Transactions on Image Processing, 2004
2. Kang et al., Convolutional Neural Networks for No-Reference Image Quality Assessment, IEEE CVPR 2014.
3. Mittal et al., Making a 'Completely Blind' Image Quality Analyzer, IEEE Signal Processing Letters, 2013

CNNs for NRIQA Explained

- Relies on the ability of neural networks to capture *non-linearities*
- The convolutional layer directly accepts image input
  $\hat{I}(i,j) = \frac{I(i,j)-\mu(i,j)}{\sigma(i,j)+1}$
  $I(i,j)$: pixel at location (i,j)
  $\mu(i,j)$: local mean
  $\sigma(i,j)$: local standard deviation
- Input patch size: 32×32
- Convolutional layer size: 26×26 (50)
- Dimensionality reduction: *min* pooling and *max* pooling
- Non-linearity: Rectified Linear Unit (ReLU)
  $g = \max(0, \sum_i w_i a_i)$

SROCC with DMOS on LIVE dataset – PSNR: 0.8636, SSIM: 9129, RRED: 0.9343, CNN: 9202.

Unsupervised Learning: Natural Image Quality Evaluator (NIQE) [Mittal et al. 2013]

- **Statistical modeling** of **normalized pixels**
- *Hypothesis:* distortion affects pixel statistics of natural scenes
- **Measure this change to estimate distortion**
- Models normalized pixel statistics using a **Generalized Gaussian Density (GGD)**
- **Modeling model** parameters using a **Multivariate Gaussian Density (MVD)**

SROCC with DMOS on LIVE dataset – PSNR: 0.8636, SSIM: 9129, RRED: 0.9343, NIQE: 9135.

NIQE Highlights

- Completely unsupervised algorithm: *opinion unaware* and *distortion unaware*
- Features based on a *fundamental property* of natural scenes
- Operates in in the *pixel domain*
- Delivers *excellent performance* and is *very fast*