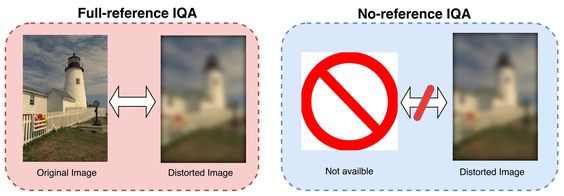
RankIQA: Learning from Rankings for No-reference Image Quality Assessment

Xialei Liu, Joost van de Weijer, Andrew D Bagdanov

Centre de Visió per Computador

Image Quality Assessment (IQA)

Definition: IQA algorithms must estimate objective image quality consistent with human evaluation.



Problem: Hard to train CNN-based methods due to *absence of large datasets for IQA*.

Main observation: While human-annotated IQA data is difficult to obtain, it is easy to *generate* images *ranked* according to image quality.

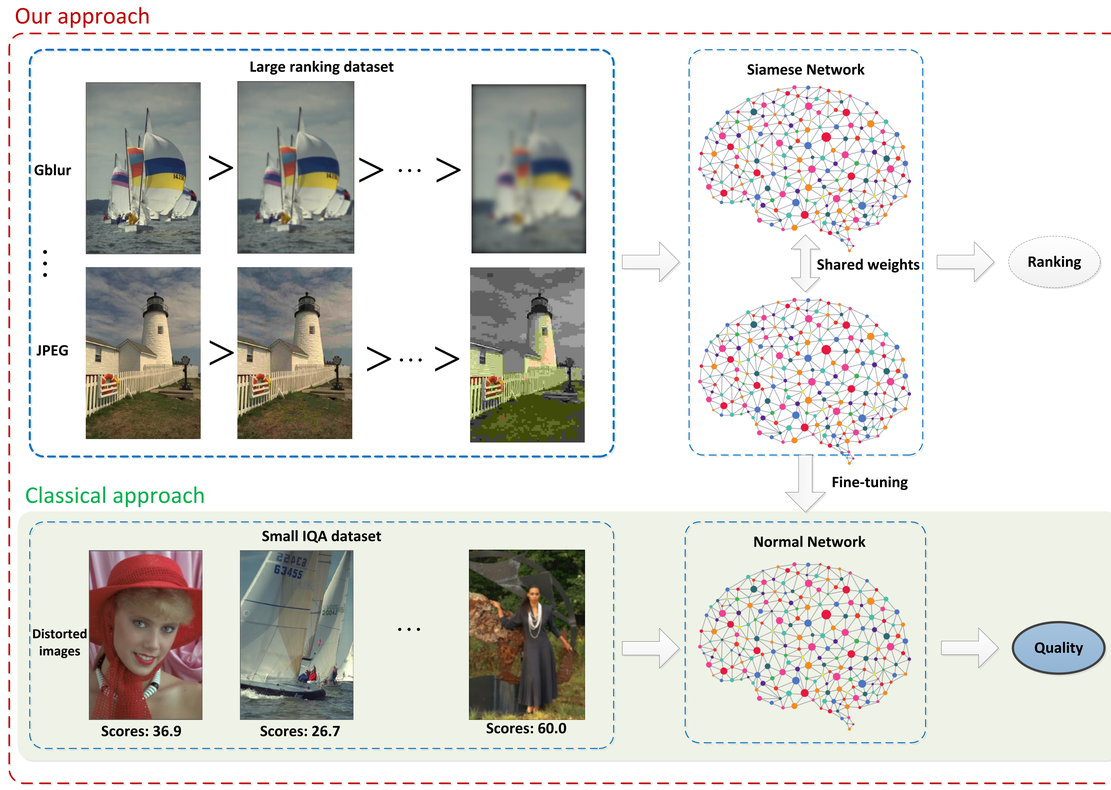
Datasets

- **LIVE**: 808 images generated from 29 originals using five distortion types.
- **TID2013**: 3000 images generated from 25 originals using twenty-four distortion types.
- **Waterloo** and **Places2**: High quality images used to generate ranked image sets.

Evaluation protocols

- **Linear Correlation Coefficient**: A measure of the linear correlation relationship.
- **Spearman Rank Order Correlation Coefficient**: A measure of the monotonic relationship.

RankIQA: Learning from rankings

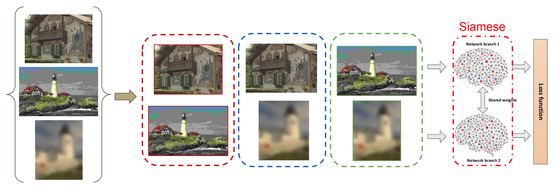


- **Classical Approach**: Train a deep CNN regressor *directly* on the ground-truth.
- **Our approach**: Train network from an image *ranking* dataset (ranked images can be generated by applying distortions of varying intensities), then fine-tune a regression network on labeled IQA data.

Siamese network for ranking

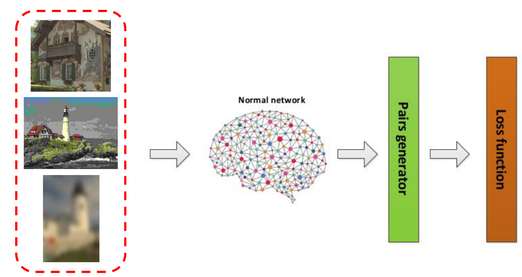
We use the pairwise ranking hinge loss:

$$L(x_1, x_2; \theta) = \max\left(0, f(x_2; \theta) - f(x_1; \theta) + \varepsilon\right)$$

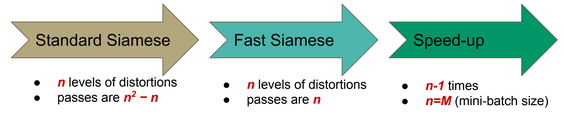


To compute the backpropagation of 3 pairs, 3 images are passed through the network ***twice***.

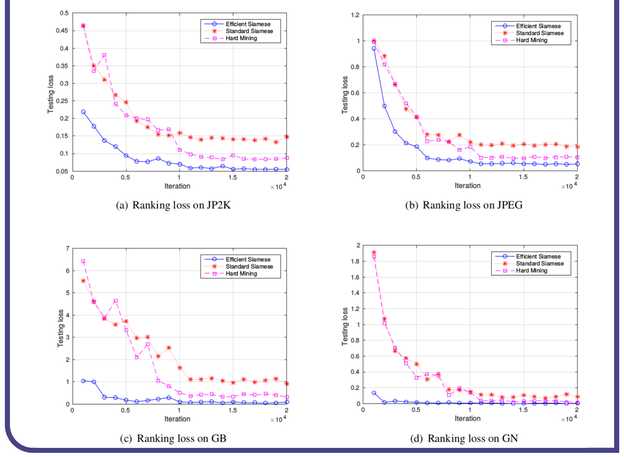
Efficient Siamese back-propagation



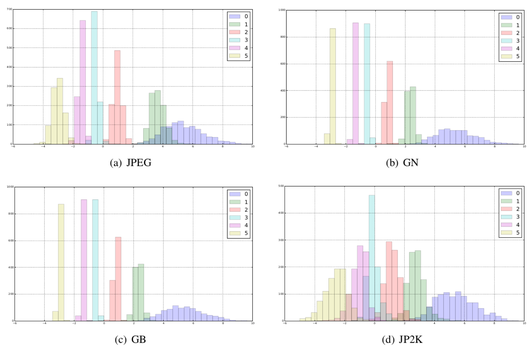
To compute the backpropagation of 3 pairs, 3 images are passed through the network ***once***.



Comparison of sampling methods is as follows,



Experimental results



Outputs of RankIQA for different types.

	LCC	JP2K	JPEG	GN	GB	FF	ALL
FR-IQA	PSNR	0.873	0.876	0.926	0.779	0.87	0.856
	SSIM [35]	0.921	0.955	0.982	0.893	0.939	0.906
	FSIM [42]	0.91	0.985	0.976	0.978	0.912	0.96
	DCNN [17]	–	–	–	–	–	0.977
NR-IQA	DIVINE [21]	0.922	0.921	0.988	0.923	0.888	0.917
	BLIINDS-II [23]	0.935	0.968	0.98	0.938	0.896	0.93
	BRISQUE [19]	0.923	0.973	0.985	0.951	0.903	0.942
	CORNIA [40]	0.951	0.965	0.987	0.968	0.917	0.935
	CNN [11]	0.953	0.981	0.984	0.953	0.933	0.953
	SOM [43]	0.952	0.961	0.991	0.974	0.954	0.962
	DNN [2]	–	–	–	–	–	0.972
	RankIQA+FT	**0.975**	**0.986**	**0.994**	**0.988**	**0.960**	**0.982**

Table 1. Evaluation (LCC) on LIVE dataset.

LCC	JP2K	JPEG	GN	GB	ALL
RankIQA+FT (Waterloo)	0.975	0.986	0.994	0.988	0.982
RankIQA+FT (Places2)	0.983	0.983	0.993	0.990	0.981

Table 2. RankIQA on different datasets.

Method	#01	#02	#03	#04	...	#24	ALL
BLIINDS-II	0.714	0.728	0.825	0.358	...	0.856	0.550
BRISQUE	0.630	0.424	0.727	0.321		0.800	0.562
CORNIA-10K	0.341	-0.196	0.689	0.184		0.903	0.651
HOSA	0.853	0.625	0.782	0.368		**0.905**	0.728
Baseline	0.605	0.468	0.761	0.232	...	0.742	0.612
RankIQA	**0.891**	**0.799**	**0.911**	**0.644**		0.768	0.623
RankIQA+FT	0.667	0.620	0.821	0.365		0.859	**0.780**

Table 3. Evaluation (SROCC) on TID2013 dataset.

Comparisons with the state-of-the-art on different datasets.