On the Use of Deep Learning for Blind Image Quality Assessment(DeepBIQ)

Simone Bianco,Luigi Celona,Paolo Napoletano,Raimondo Schettini

University of Milan-Bicocca

Arxiv,2017

2019/4/22

Main Contribution

- Investigate the use of different design choices on IQA problem
  - The choice of pre-trained model
  - The number of sub-regions and pooling strategies
  - Fine-tune of CNN

Datasets

- LIVE In the Wild Image Quality Challenge Database

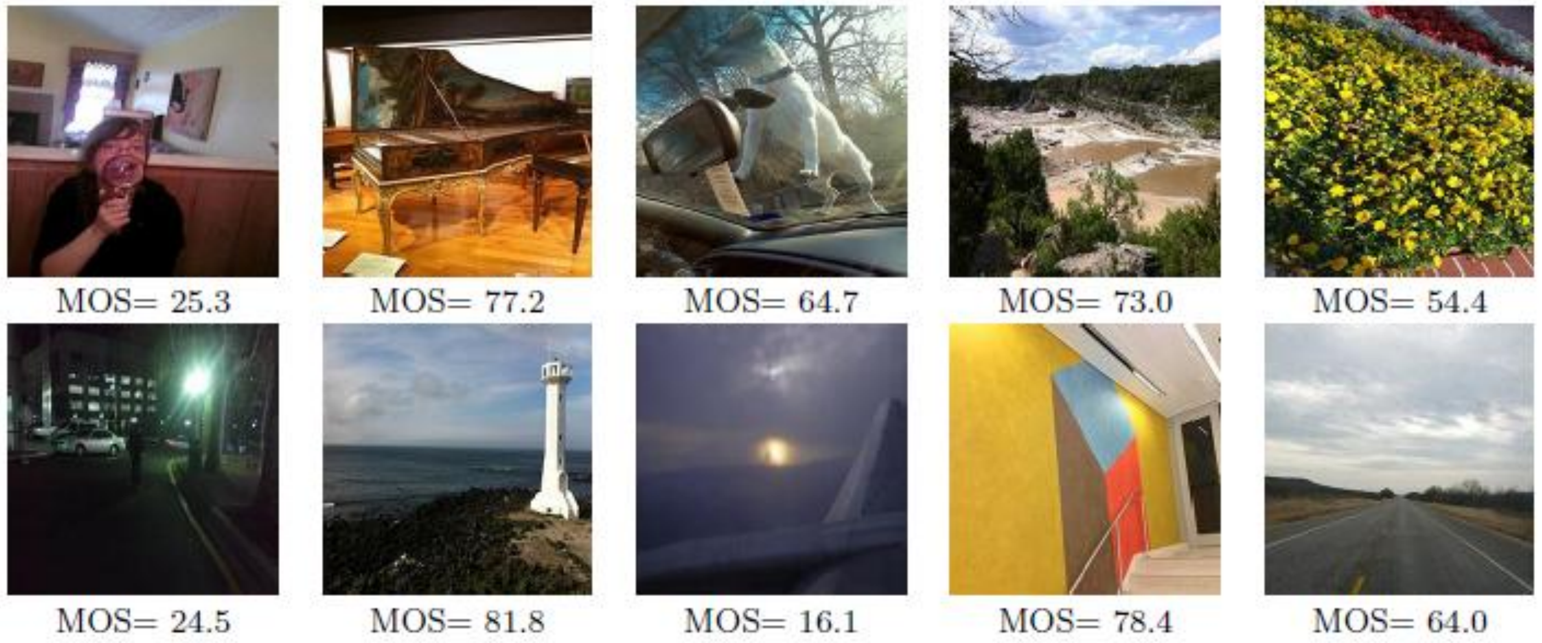
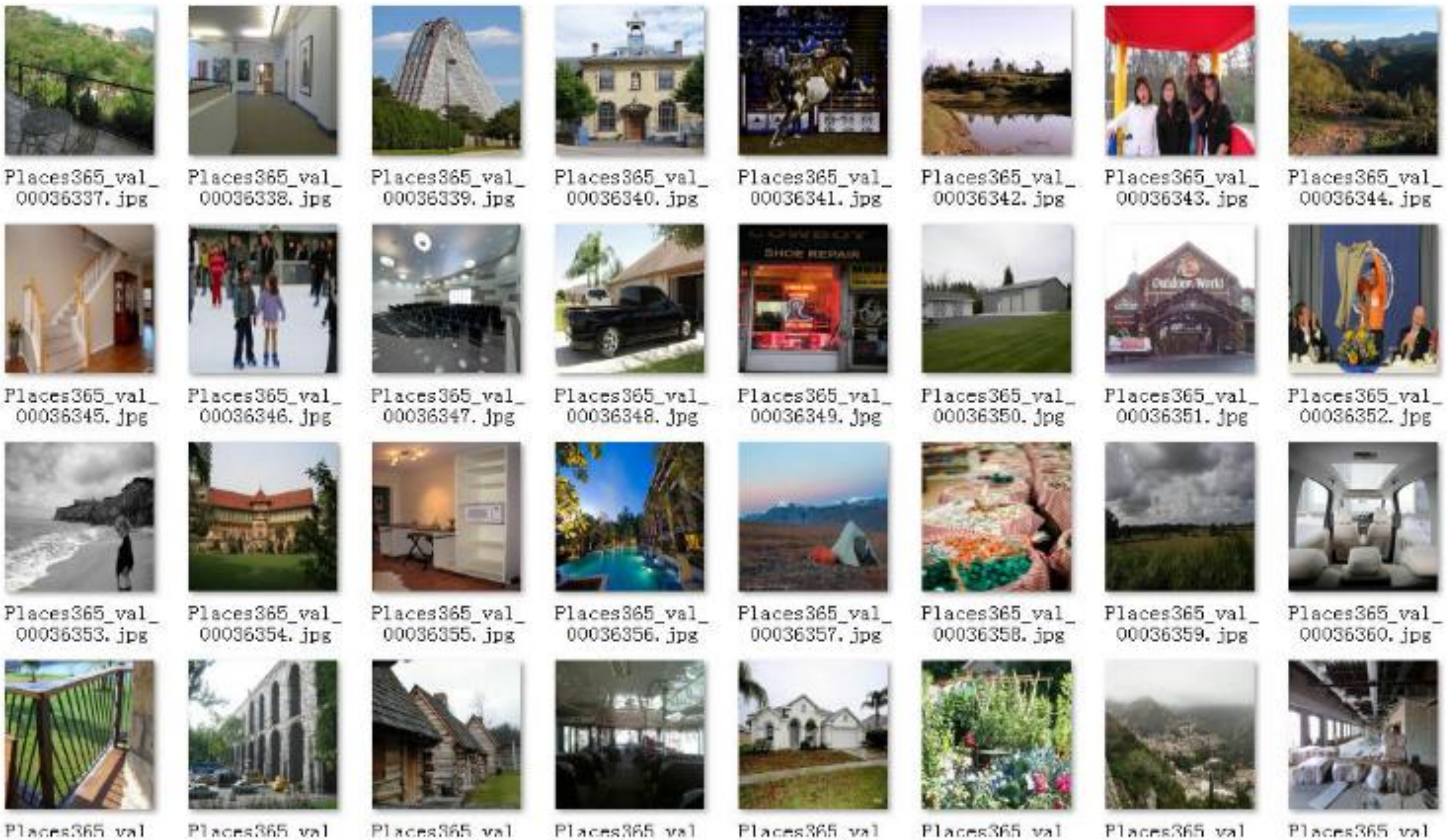


Fig. 1 Examples from the LIVE In the Wild IQ Chall.DB.

Experiments

- Experiment I: the choice of pre-trained CNNs
  - ImageNet-CNN, trained on 1.2 million images of ImageNet Database
  - Places-CNN, trained on 2.5 million images of the Places Database
  - ImageNet+Places-CNN, trained on 3.5 million images of the mergence of the scene categories from Places Database and the object categories from ImageNet

Places



Places365_val_00036337.jpg
Places365_val_00036338.jpg
Places365_val_00036339.jpg
Places365_val_00036340.jpg
Places365_val_00036341.jpg
Places365_val_00036342.jpg
Places365_val_00036343.jpg
Places365_val_00036344.jpg
Places365_val_00036345.jpg
Places365_val_00036346.jpg
Places365_val_00036347.jpg
Places365_val_00036348.jpg
Places365_val_00036349.jpg
Places365_val_00036350.jpg
Places365_val_00036351.jpg
Places365_val_00036352.jpg
Places365_val_00036353.jpg
Places365_val_00036354.jpg
Places365_val_00036355.jpg
Places365_val_00036356.jpg
Places365_val_00036357.jpg
Places365_val_00036358.jpg
Places365_val_00036359.jpg
Places365_val_00036360.jpg

Experiments

- Experiment I: the choice of pre-trained CNNs
    - Resize from 500*500 to 256*256, crop out the central 227*227 part
    - Extract feature by CNN, then use SVR to predict the quality score
    - Use 80 percent for training and 20 percent for testing, 10 times

Experiments

Table 2 Median LCC and SROCC across 10 train-test random splits of the LIVE In the Wild Image Quality Challenge Database considering only the central crop of the subsampled image as input for the pre-trained CNNs considered.

	LCC	SROCC
Imagenet-CNN	0.6782	0.6381
Places-CNN	0.6267	0.6055
ImageNet+Places-CNN	**0.7215**	**0.7021**

Experiments

- Experiment II: feature and prediction pooling
  - The resize operation could have reduced the effect of some artifacts of each image, such as noise
  - Randomly crops different amounts of sub-regions(227*227), the number of which ranges from 5 to 50
  - Fusion strategies:
    - feature pooling: minimum, average, and maximum
    - feature concatenation: longer feature vector
    - prediction pooling: minimum, average, and maximum

Experiments

Table 3 Median LCC and SROCC across 10 train-test random splits of the LIVE In the Wild IQ Chall. DB considering randomly selected crops as input for the ImageNet+Places-CNN and three different fusion approaches.

	LCC	SROCC
Feature pooling (avg-pool,@30crops)	**0.7938**	**0.7828**
Feature concatenation (@35crops)	0.7864	0.7724
Prediction pooling (avg-pool,@20crops)	0.7873	0.7685

Experiments

- Experiment III: fine-tuned CNN
  - Substitute a new fully connected layer initialized with random values
  - During training, classify image sub-regions into five disjoint sets(bad, poor, fair, good, excellent)
  - Use the trained CNN as a feature extractor, and then predict the quality score by SVR
  - Still use the sub-regions strategy

Experiments

- Experiment III: fine-tuned CNN
  - ImageNet+Places-CNN
  - Average-pooling

Table 4 Median LCC and SROCC across 10 train-test random splits of the LIVE In the Wild Image Quality Challenge Database considering randomly selected crops as input for the fine-tuned CNN and two different fusion approaches.

	LCC	SROCC
Feature pooling (avg-pool,@20crops)	0.9026	0.8851
Prediction pooling (avg-pool,@25crops)	**0.9082**	**0.8894**

Experiments

Table 5 Median LCC and median SROCC across 10 train-test random splits of the LIVE In the Wild IQ Chall. DB.

	LCC	SROCC
DIIVINE [34]	0.56	0.51
BRISQUE [31]	0.61	0.60
BLIINDS-II [39]	0.45	0.40
S3 index [47]	0.32	0.31
NIQE [32]	0.48	0.42
C-DIIVINE [51]	0.66	0.63
FRIQUEE [12,14]	0.71	0.68
HOSA [49]	-	0.65
DeepBIQ (Exp. I)	0.72	0.70
DeepBIQ (Exp. II)	0.79	0.79
DeepBIQ (Exp. III)	**0.91**	**0.89**

Experiments

Table 6 Median LCC and median SROCC across 100 random splits of the legacy LIVE Image Quality Assessment DB.

Method	LCC	SROCC
DIIVINE [34]	0.93	0.92
BRISQUE [31]	0.94	0.94
BLIINDS-II [39]	0.92	0.91
NIQE [32]	0.92	0.91
C-DIIVINE [51]	0.95	0.94
FRIQUEE [12,14]	0.95	0.93
ShearletIQM [29]	0.94	0.93
MGMSD [1]	0.97	**0.97**
Low Level Features [21]	0.95	0.94
Rectifier Neural Network [45]	–	0.96
Multi-task CNN [20]	0.95	0.95
Shallow CNN [19]	0.95	0.96
DLIQA [16]	0.93	0.93
HOSA [49]	0.95	0.95
CNN-Prewitt [27]	0.97	0.96
CNN-SVR [26]	0.97	0.96
DeepBIQ	**0.98**	**0.97**

Experiments

Table 7 Median LCC and median SROCC across 100 train-val-test random splits of the CSIQ.

Method	LCC	SROCC
DIIVINE [34]	0.90	0.88
BRISQUE [31]	0.93	0.91
BLIINDS-II [39]	0.93	0.91
Low Level Features [21]	0.94	0.94
Multi-task CNN [20]	0.93	0.94
HOSA [49]	0.95	0.93
DeepBIQ	**0.97**	**0.96**

Experiments

Table 8 Median LCC and median SROCC across 100 train-val-test random splits of the TID2008.

Method	LCC	SROCC
DIIVINE [34]	0.90	0.88
BRISQUE [31]	0.93	0.91
BLIINDS-II [39]	0.92	0.90
MGMSD [1]	0.88	0.89
Low Level Features [21]	0.89	0.88
Multi-task CNN [20]	0.90	0.91
Shallow CNN [19]	0.90	0.92
DeepBIQ	**0.95**	**0.95**

Experiments

Table 9 Median LCC and median SROCC across 100 train-val-test random splits of the TID2013.

Method	LCC	SROCC
DIIVINE [34]	0.89	0.88
BRISQUE [31]	0.92	0.89
BLIINDS-II [39]	0.91	0.88
Low Level Features [21]	0.89	0.88
HOSA [49]	**0.96**	0.95
DeepBIQ	**0.96**	**0.96**

My own thinking

- The training and settings mentioned in this paper is useful
  - Fine-tune CNN is necessary when data is not so small
  - The choice of pre-trained model may affect the result
  - Sub-images and pooling might help, maybe in natural images
  - Different pooling strategies during convolutional layers may lead to different results

Thanks!