Algorithm Selection for Image Quality Assessment

M. Wagner[1], H. Lin[2], S. Li[3], D. Saupe[2]

Universities of Adelaide[1], Konstanz[2], Kent[3]

SFB-TRR 161
Quantitative Methods for Visual Computing

Subjective image quality assessment (ground truth)

MOS = 3.3 | MOS = 24.3 | MOS = 42.8 | MOS = 63.3 | MOS = 82.8

Mean opinion scores (MOS) of 120 observers on a scale from 0 (worst) to 100 (best). Five out of 10,073 images.

Objective image quality assessment (IQA)

Algorithm input:	rgb-image
Algorithm output:	estimation of MOS $\in [0, 100]$
Method:	image feature extraction, machine learning
Eight algorithms:	BIQI BLINDS-II BRISQUE CORNIA DIIVINE HOSA SSEQ KonCept512
KonIQ-10k dataset:	8,058 annotated images for training 2,015 annotated images for testing
Evaluation:	mean absolute error (MAE) Spearman rank order correlation (SROCC)

Method	Features	SROCC	MAE	Method best for images Rank 1	Rank 2	Rank 3
BIQI	18	0.559	8.339	187	188	240
BLIINDS-II	24	0.585	9.239	185	215	205
BRISQUE	36	0.705	8.224	176	205	253
CORNIA	20,000	0.780	7.308	217	263	286
DIIVINE	88	0.589	8.180	169	198	259
HOSA	14,700	**0.805**	**6.792**	**220**	324	316
SSEQ	12	0.604	9.403	179	227	168
KonCept512	1,536	**0.921**	**4.154**	**682**	395	288
Virtual best method	NA	**0.978**	**2.069**	**2,015**	0	0

Table 1: Performance of 8 IQA methods on the KonIQ-10k test set.

Result: The virtual best method (oracle) is far superior to the single best one!

Research question: Can methods of algorithm selection produce a hybrid method that comes close to the oracle?

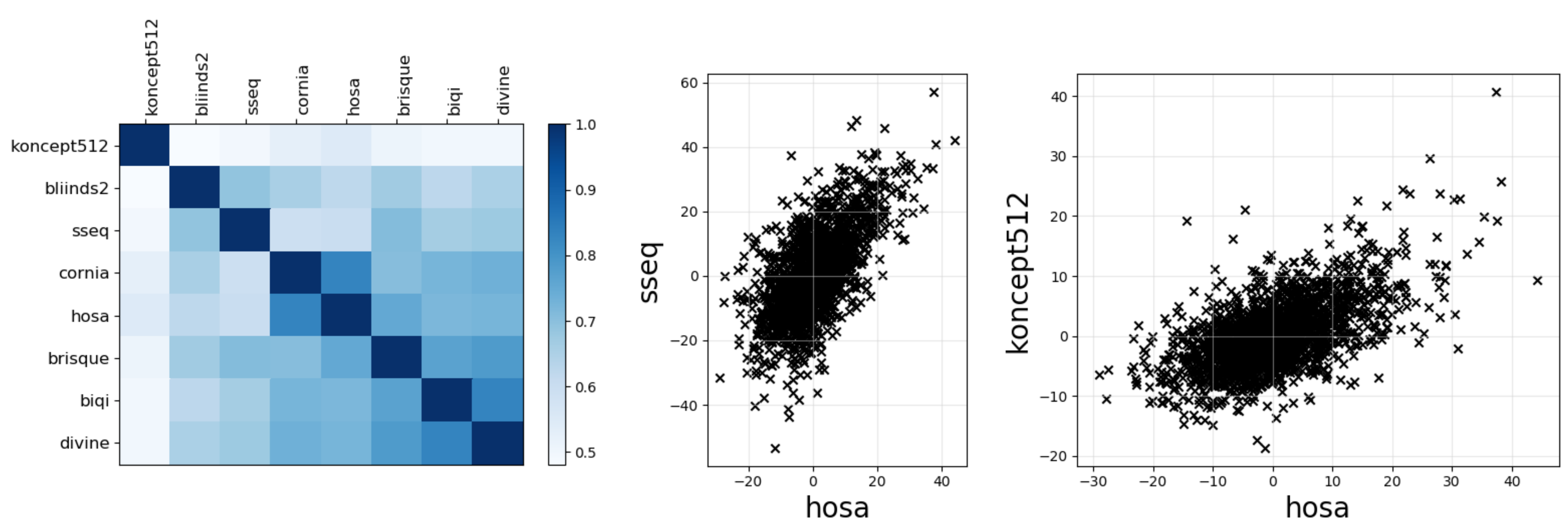


Figure 1: Left: The correlations (SROCC) between the predictions of the 8 selected methods. KonCept512's performance across all test instances is the most different from the other seven. Right: Two scatter plots showing the (signed) errors $M(I) - \text{MOS}(I)$ for two pairs of methods. Points clustered along the vertical axis imply that the method plotted on the horizontal axis has smaller errors, and vice versa. So HOSA is more accurate than SSEQ, but less than KonCept512.

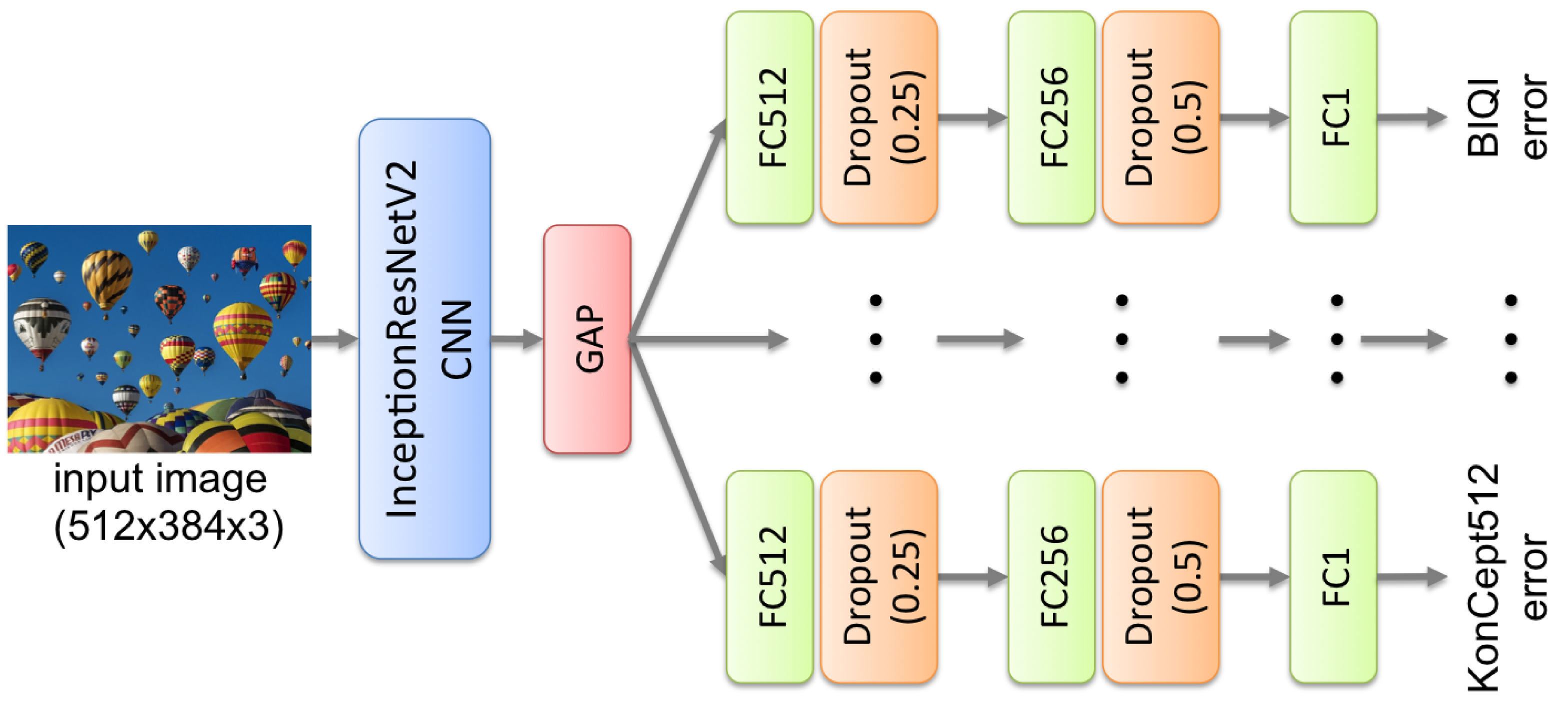
Algorithm selection for IQA

AutoFolio (AF)

- AutoFolio uses algorithm configuration to optimize the performance of algorithm selection systems by determining the best selection approach and its hyperparameters.

Deep learning (DL)

- Error function for method M and image I: $f_M(I) = |M(I) - \text{MOS}(I)|$
- Siamese network learns output of the 8 IQA methods
- CNN base network: pretrained by ImageNet, finetuned on the KonIQ-10k training set
- Selects method with smallest predicted error



Method	Using all methods SBM	VBM	AF	DL	w/o KonCept512 SBM	VBM	AF
BIQI	–	187	0	0	–	263	51
BLIINDS-II	–	185	0	0	–	277	32
BRISQUE	–	176	0	1	–	256	140
CORNIA	–	217	0	58	–	329	512
DIIVINE	–	169	0	0	–	241	252
HOSA	–	220	0	0	2015	363	918
SSEQ	–	179	0	0	–	286	110
KonCept512	2015	682	2015	1956	–	–	–
MAE	4.154	2.069	4.154	6.447	6.792	3.063	6.665
SROCC	0.921	0.978	0.921	0.871	0.805	0.954	0.784

Table 2: Performance of single best method (SBM), virtual best method (VBM), and algorithm selection (AS) by AutoFolio and deep learning (DL).

Result: Algorithm selection does not outperform the single best method!

Conjecture / research questions:

- Is the failure of AS due do intrinsic "noisyness" of IQA methods?
- How can one quantitatively assess the noisiness of IQA methods?
- Does "denoising" of IQA methods improve their performance?
- Does denoising remove the gap between the SBM and the VBM?
- Are denoised IQA methods better suited for the algorithm selection?

Download paper:

Acknowledgment:

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 251654672 TRR 161.