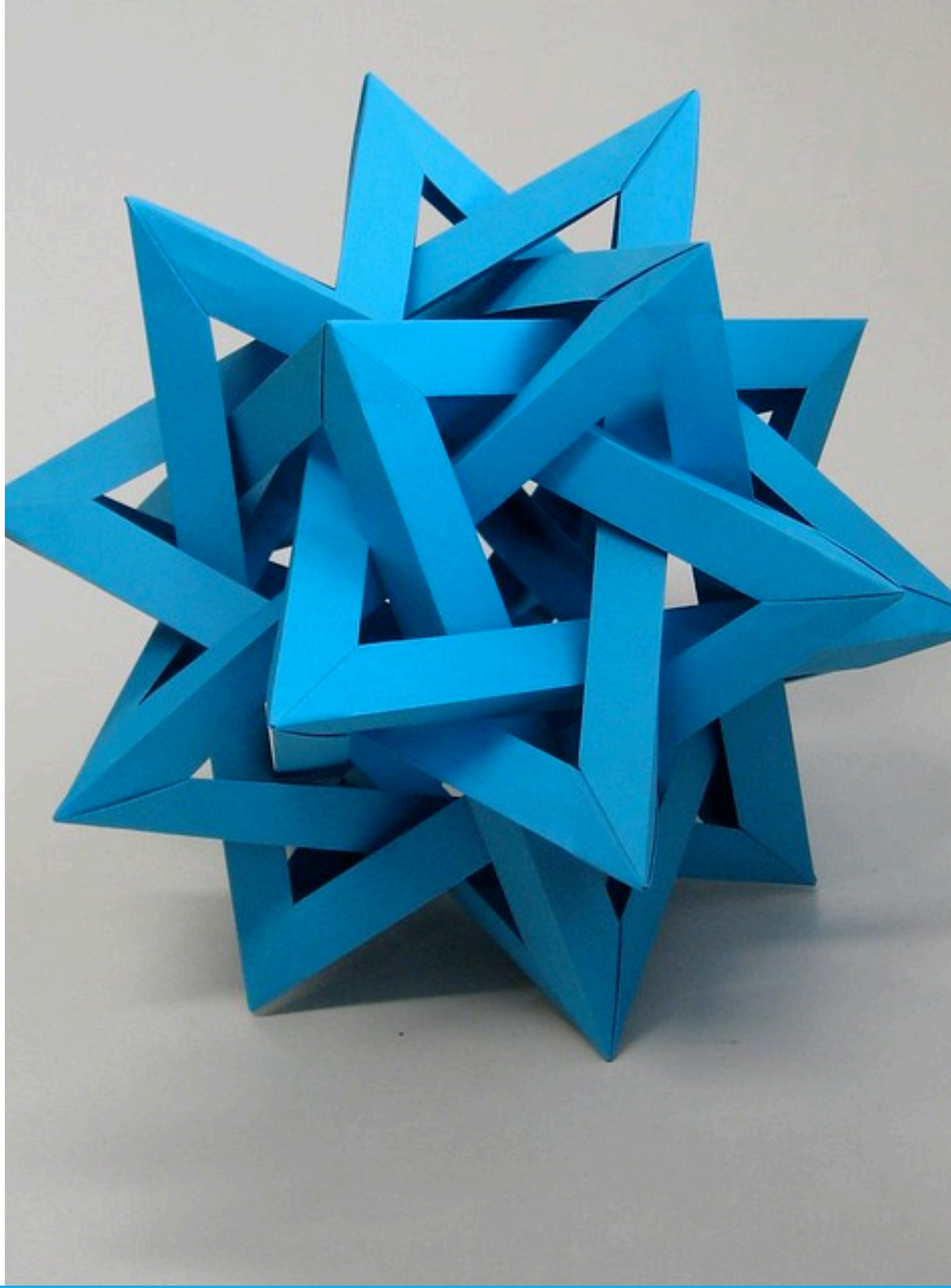
POLITECNICO DI TORINO
Dipartimento di Automatica e Informatica

Artificial Intelligence based Observers for Media Quality Assessment

LOHIC FOTIO TIOTSOP



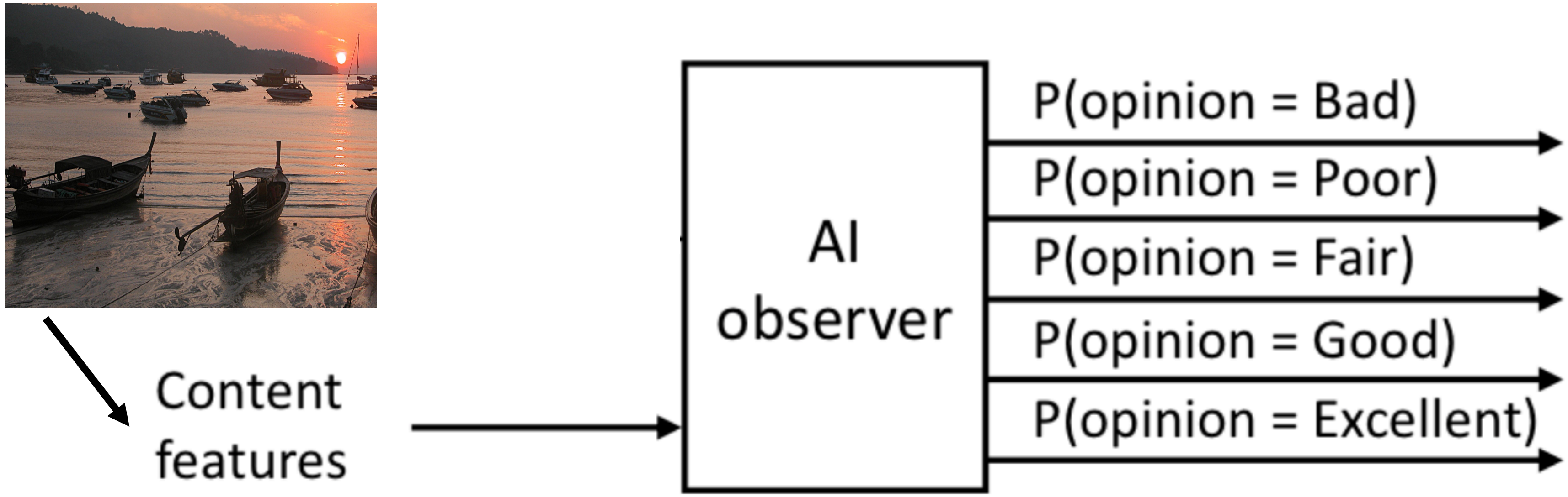
Artificial Intelligence based Observers (AIOs)

- Use Shallow/ Deep Neural network (NN) to learn single observers' quality perception
  - Train one NN for each observer
  - Ground truth data: individual opinion score

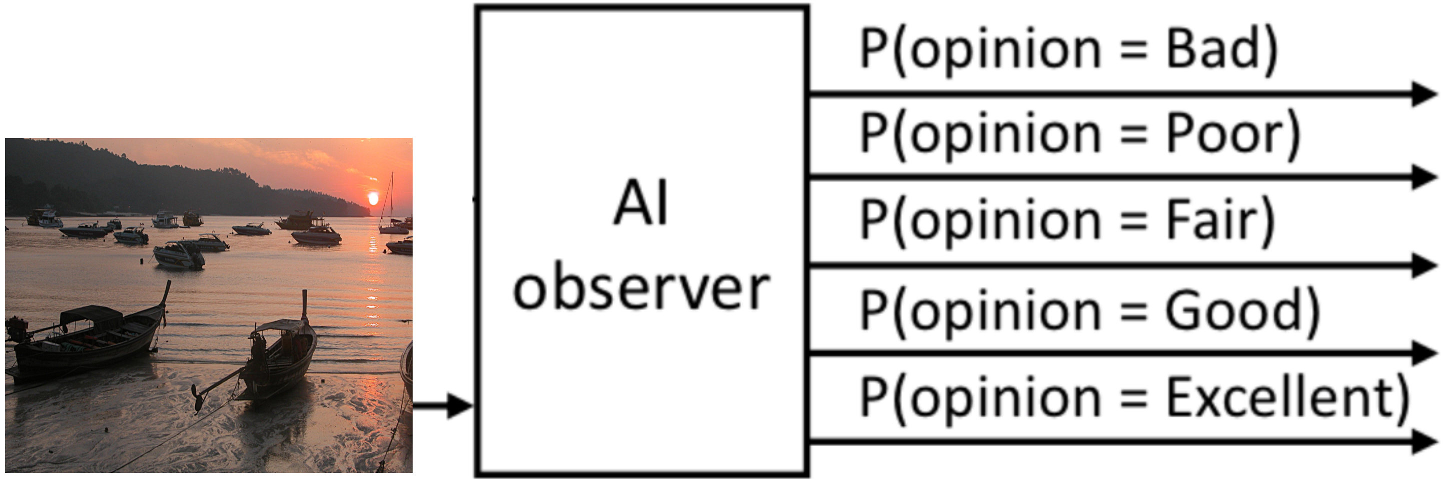
- Each neural network is looked at as a virtual observer (AIO)
  - It accounts for the characteristics and expectation of the corresponding observer
  - The subject's inconsistency is modeled

AIOs: The Implementation

Shallow NN based AIOs

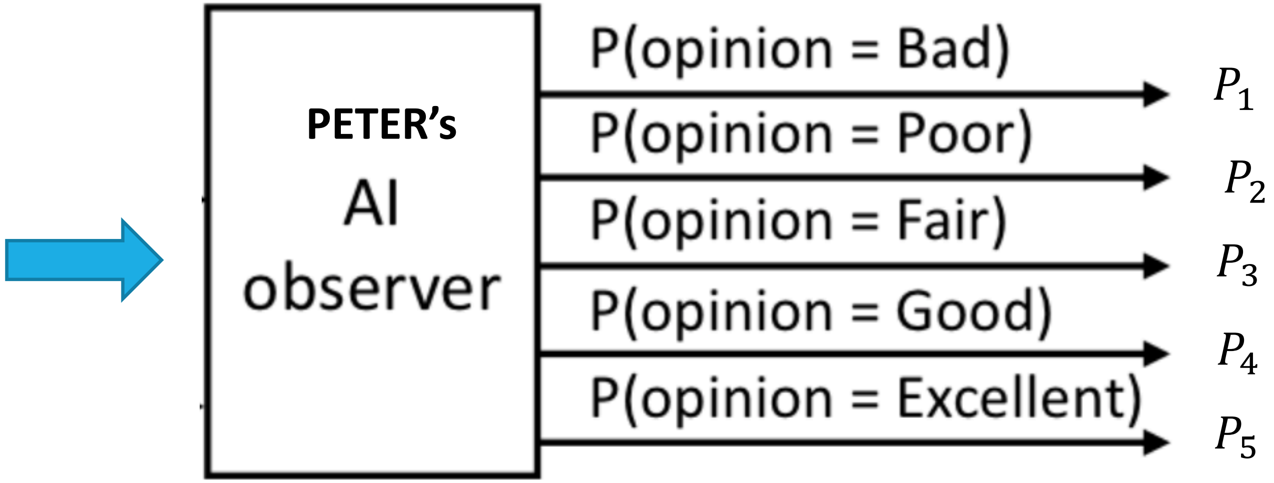


Deep NN based AIOs



AIOs: A measure of subject inconsistency

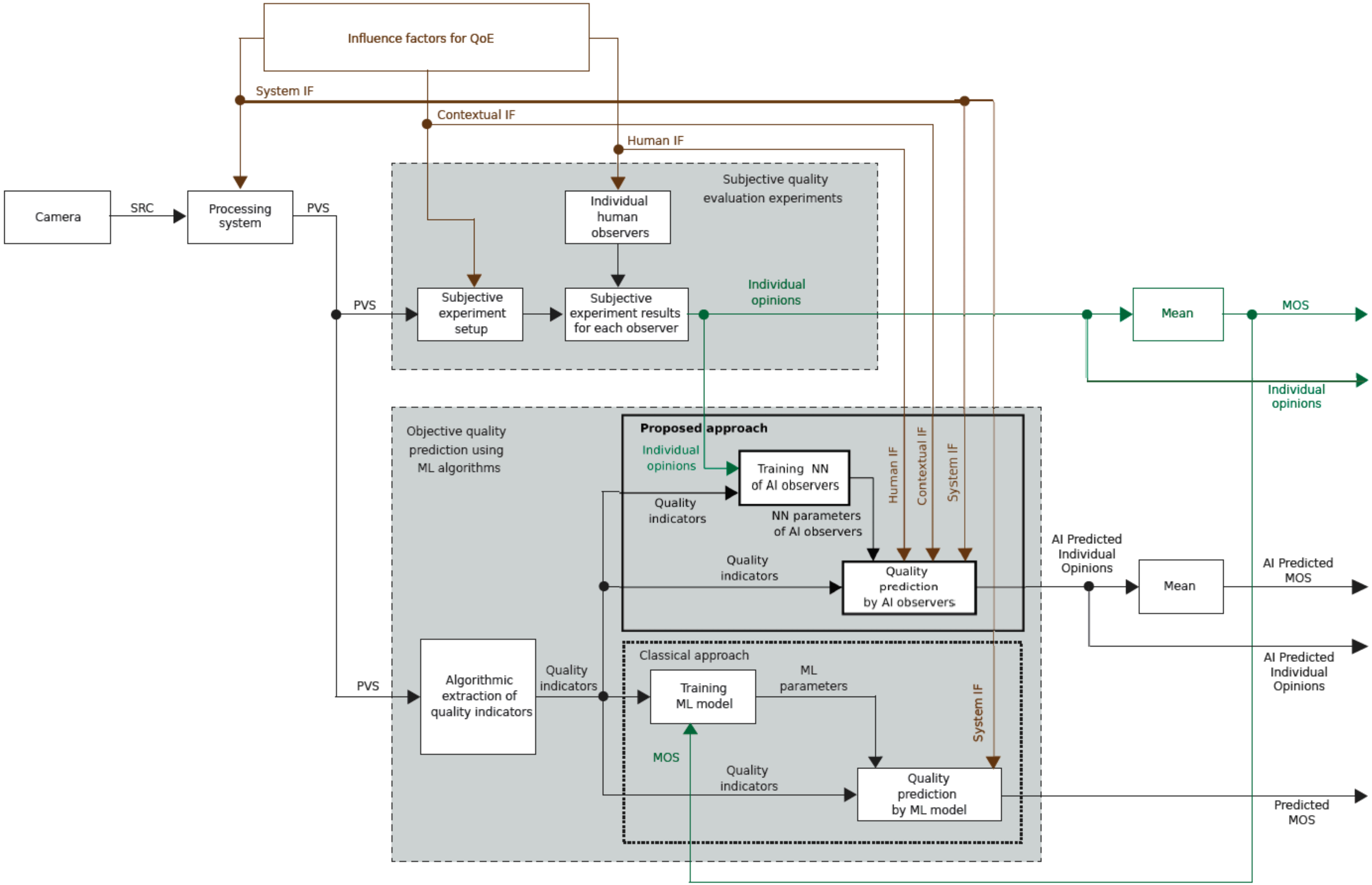
PVS



- Peter's inconsistency regarding the quality of the input PVS is predicted as:

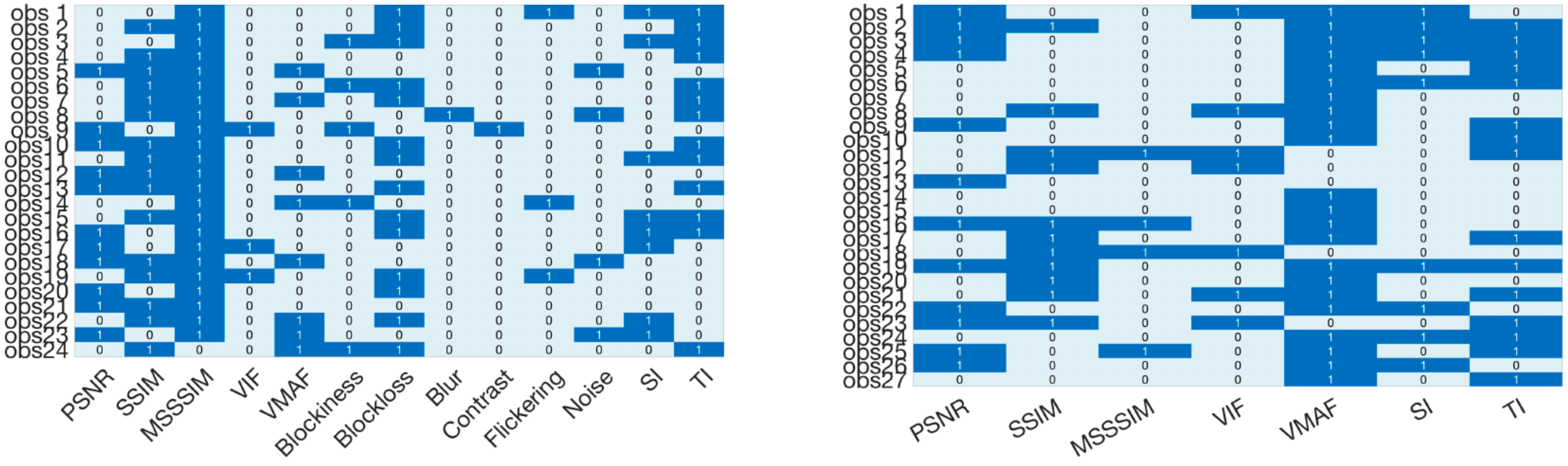
$$\boxed{\sigma_{PETER}^{pvs} = \sum_{i=1}^{5} i^2 \cdot p_i - \left(\sum_{i=1}^{5} i \cdot p_i\right)^2}$$

AIOs vs Traditional media quality assessment approach



Shallow NN based AIOs: Results

- The optimal set of features changes from an observer to another



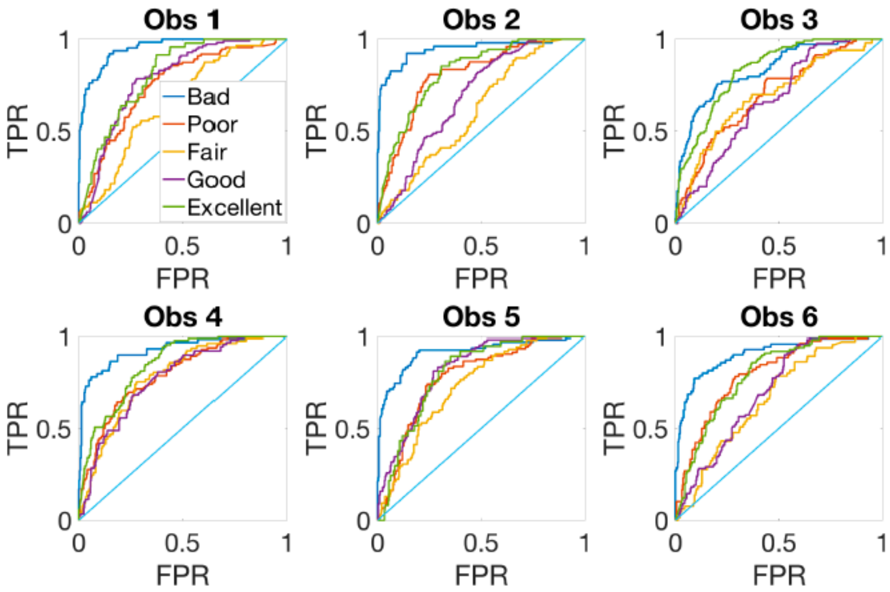
(a) VQEG-HD dataset

(b) ITS4S dataset

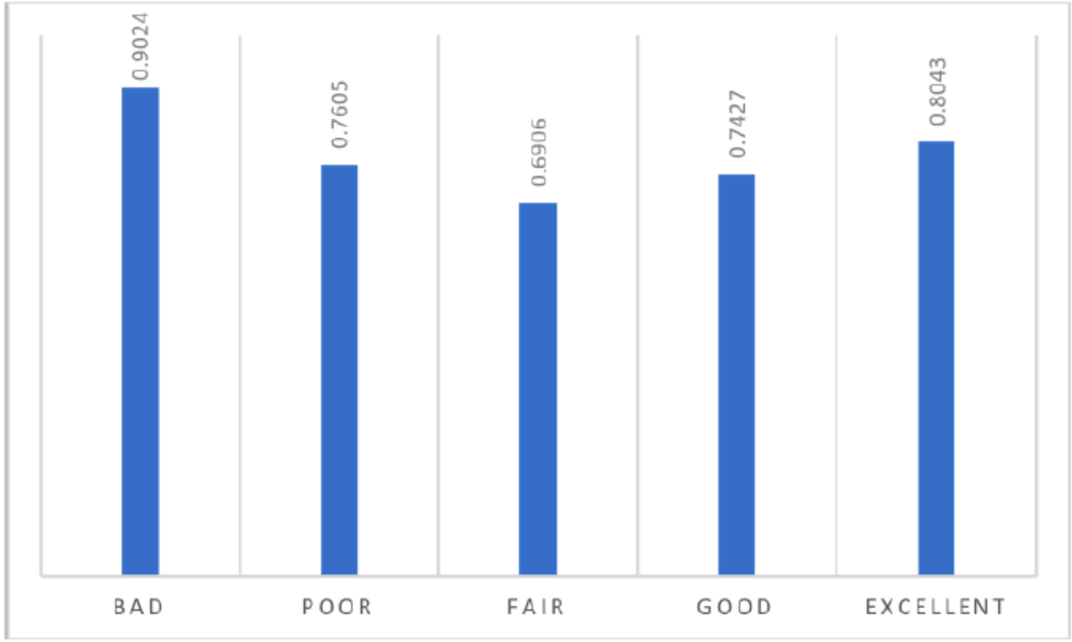
Fig. 3. Optimal set of the features to model each observer, where 1 indicates that the feature is selected, and 0 that it is not. The heterogeneity of the rows suggests that users rate the quality on the basis of different criteria.

Shallow NN based AIOs: Results

- AIOs are more accurate on the boundaries of the quality scale just like actual observers



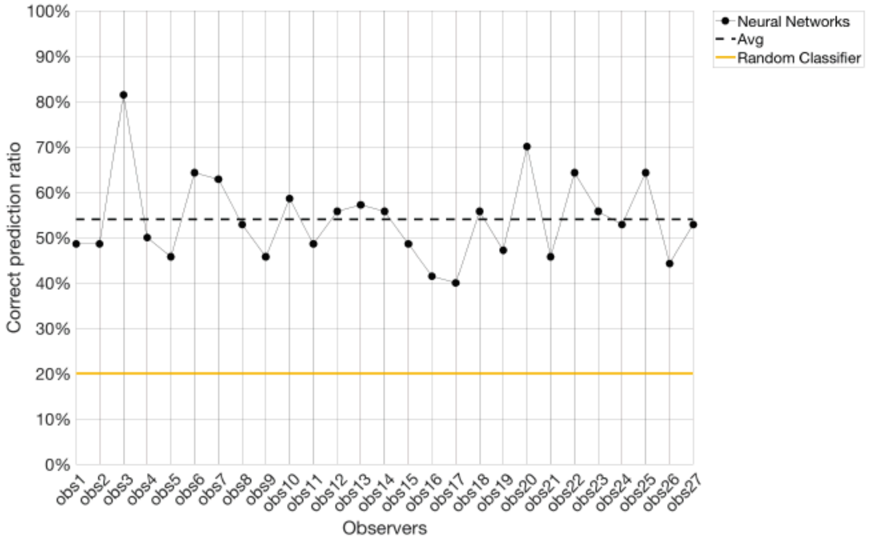
(a) Observer #1 to #6



(b) Average AUC

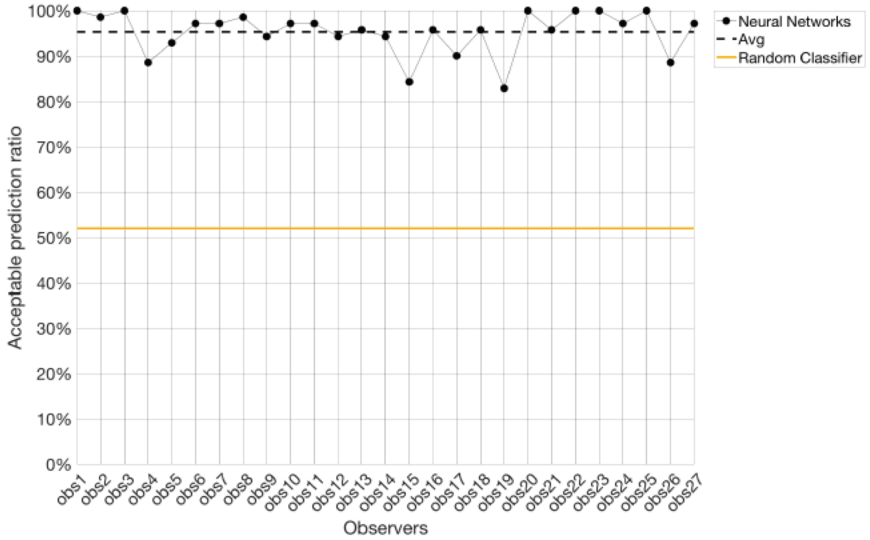
Shallows NN based AIOs: Results

- Actual observers vs their corresponding AIOs
- 27 AIOs trained using the ITS4S dataset
- The PVSs in the ”Public Safety” category are used as test set



(c) Correct prediction ratio (Public Safety category)

Predicted = actual

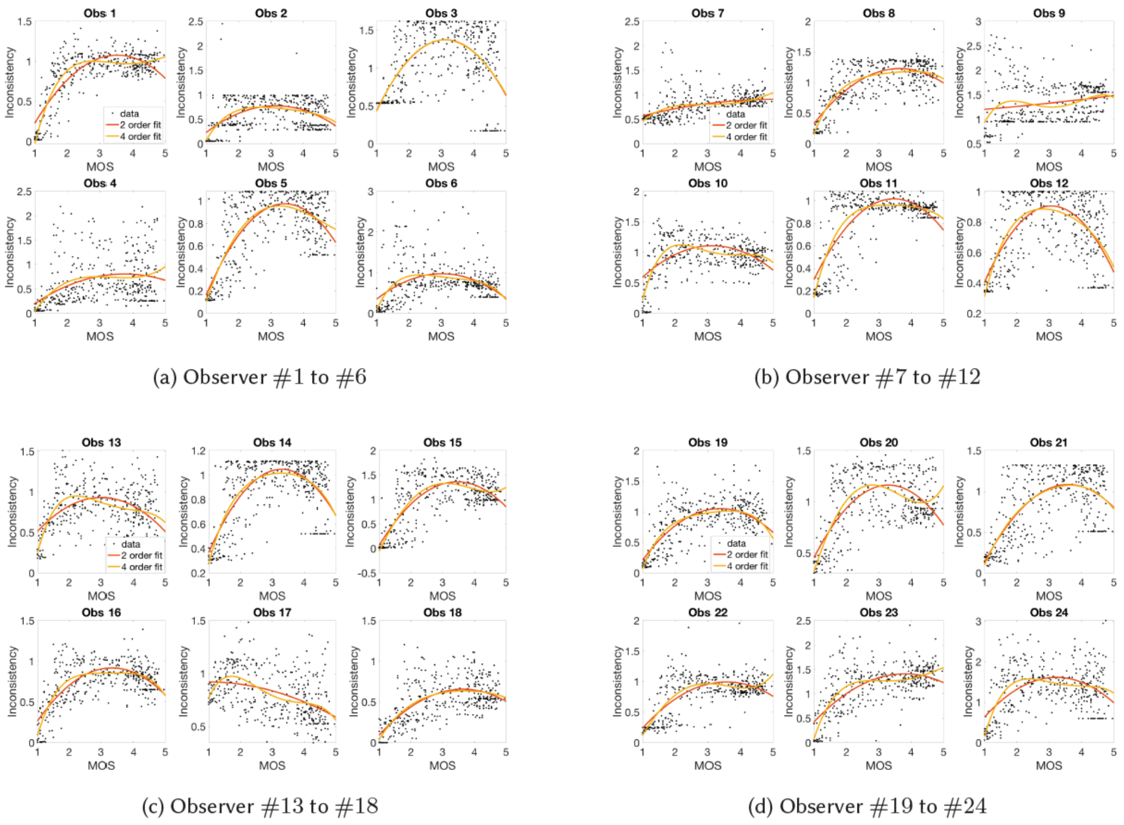


(d) Acceptable prediction ratio (Public Safety category)

|Predicted-actual|<=1

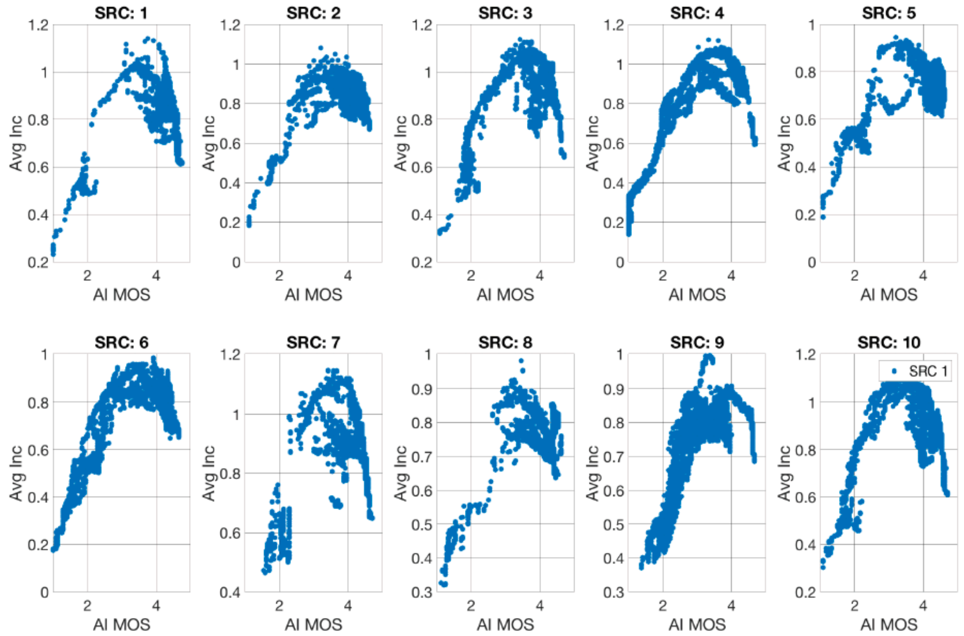
Shallow NN based AIOs: Results

- Investigating the properties of the proposed inconsistency measure



VQEG-HD

For almost all observers, larger inconsistency is observed in the middle of the quality scale

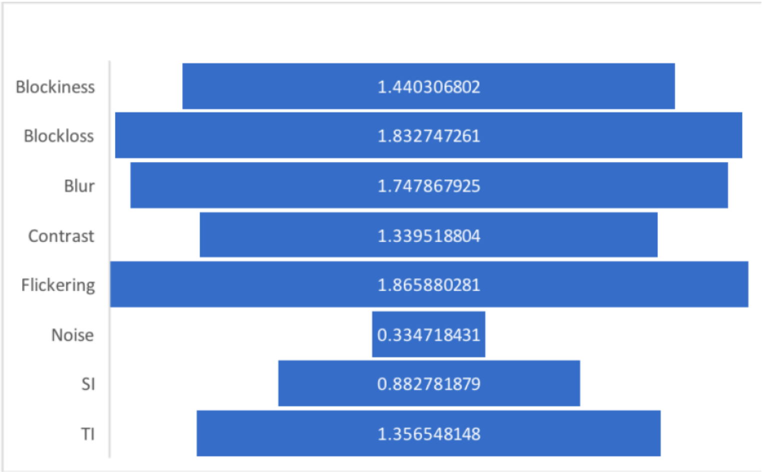


JEG-HYBRID DATASET

Independently from the source, the AOIs showed higher consistency when evaluating low quality PVSs

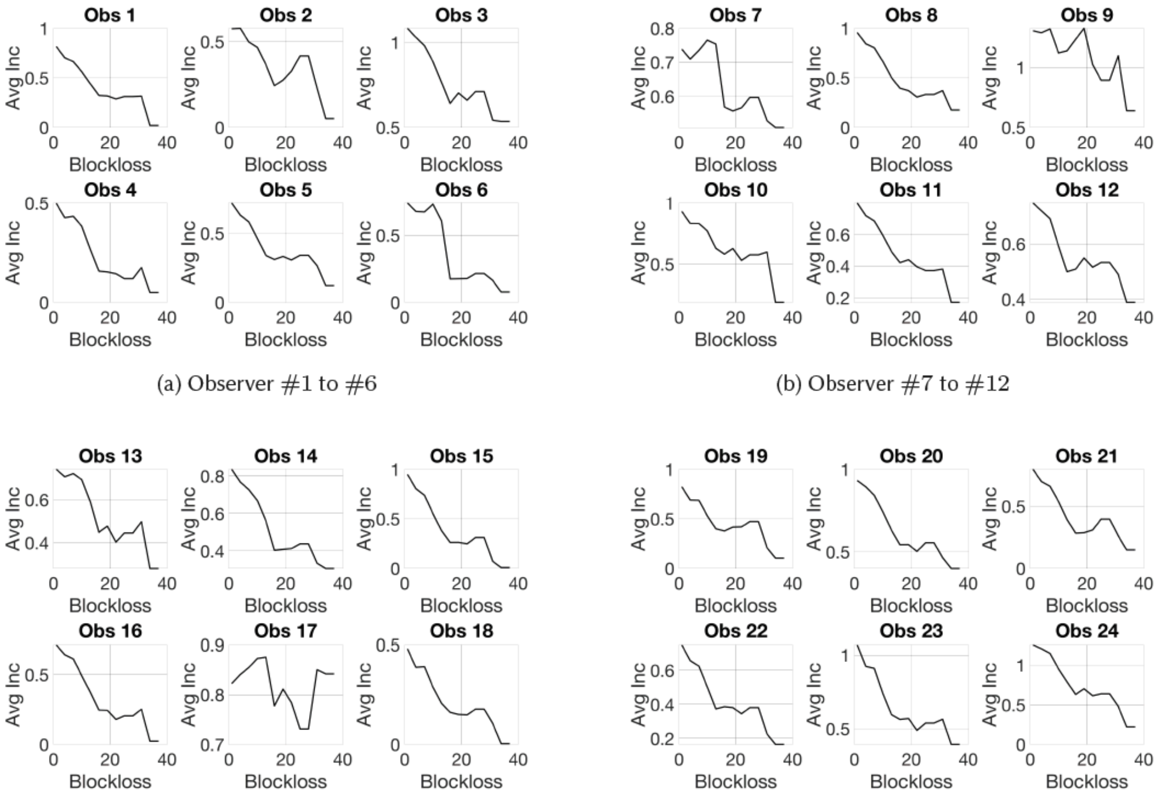
Shallow NN based AIOs: Results

- Towards understanding peculiarities of the human vision system



(b) Average importance of features

The importance of each feature for each observer is obtained using the neighborhood component analysis feature selection algorithm.



(a) Observer #1 to #6

(b) Observer #7 to #12

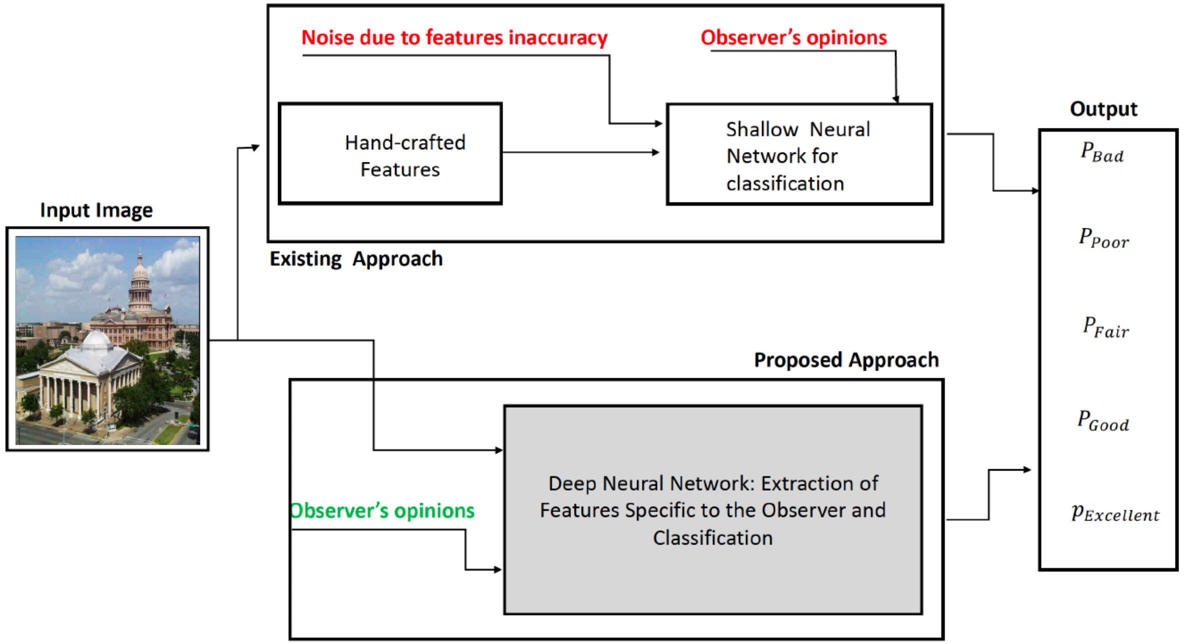
(c) Observer #13 to #18

(d) Observer #19 to #24

A significant loss of blocks leads to a distortion perceptible in a deterministic way by observers.

From Shallow NN to Deep NN based AIOs

- Hand-crafted features might not fully characterize the PVS
- There is the need to extract the optimal set of features that really model each subject's individual quality perception



- Deep NNs solve both issues

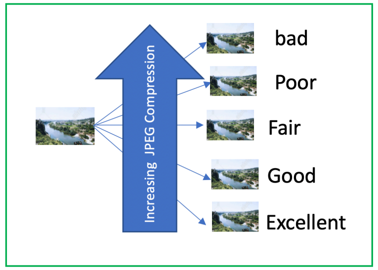
- How to overcome the lack of large-scale training set which are needed to train Deep NNs based AIOs

Deep NN based AIOs: Training Process

- Synthetic training set + Transfer learning

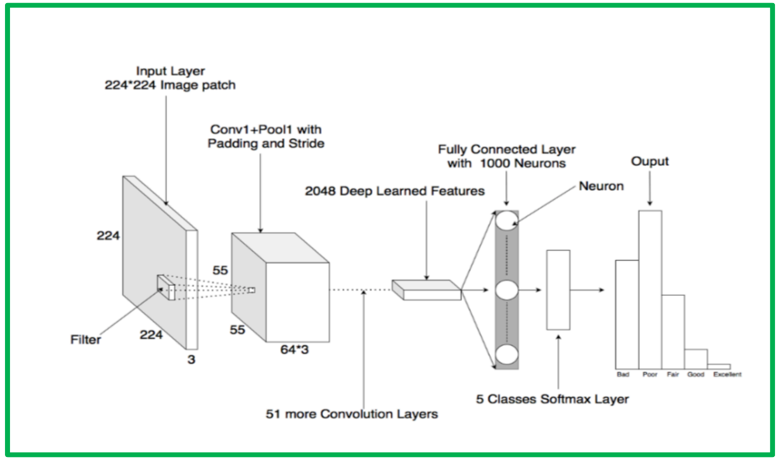
JPEG Quality parameter interval	Opinion score	Image label
[2, 10]	1	Bad
[11, 18]	2	Poor
[19, 25]	3	Fair
[26, 50]	4	Good
[51, 100]	5	Excellent

Synthetic data



Training

JPEGRestNet50



Image/Video	Hand-crafted Features	Alice Ratings
	$f_1^{Video_1}, f_2^{Video_1}, \dots, f_m^{Video_1}$	Fair
	$f_1^{Video_2}, f_2^{Video_2}, \dots, f_m^{Video_2}$	Excellent
...	...	...
	$f_1^{Video_n}, f_2^{Video_n}, \dots, f_m^{Video_n}$	Bad

Features recognizing JPEG compression are updated to capture Alice's quality perception

Alice's AIO

JPEGResNet50
+
Transfer Learning

Deep NN based AIOs: Results

- The JPEGResNet50 learned Image quality Assessment features

True Class \ Predicted Class	1=Bad	2=Poor	3=Fair	4=Good	5=Excellent		
1=Bad	489	10	1			97.8%	2.2%
2=Poor	3	452	45			90.4%	9.6%
3=Fair		25	440	35		88.0%	12.0%
4=Good			17	431	52	86.2%	13.8%
5=Excellent	1		1	33	465	93.0%	7.0%
	99.2%	92.8%	87.3%	86.4%	89.9%		
	0.8%	7.2%	12.7%	13.6%	10.1%		

Performance of the JPEGResNet50 when used to classify 2500 JPEG distorted Images

Deep NN based AIOs: Results

- The JPEGResNet50 output five probabilities values, i.e. $p_i \;\; i = 1,2, \ldots, 5$
- $MOS_{Res} = \sum_{i=1}^{5} i \cdot p_i$
- MOS_{AI} is the mean of AIOs opinions

DATASET	DISTORTION	BRISQUE	PSNR	SSIM	MOS_{res}	MOS_{AI}
CSIQ [47]	JPEG	0.86	0.89	0.94	0.95	0.91
MICT [48]	JPEG	0.90	0.64	0.64	0.88	0.75
SDIVL [49]	JPEG	0.56	0.73	0.77	0.82	0.43
TID2013 [50]	JPEG	0.81	0.91	0.92	0.94	0.84
VCL-FER [51]	JPEG	0.76	0.57	0.82	0.93	0.76
LIVE-IQA-r1 [45]	JPEG	0.94	0.85	0.96	0.96	0.92
LIVE-IQA-r2 [52]	JPEG	0.96 (T)	0.95	0.92	0.91	0.86
MICT [48]	JP2K	0.87	0.84	0.84	0.46	0.69
LIVE-IQA-r1 [45]	JP2K	0.91	0.85	0.88	0.59	0.83
LIVE-MD-ph1 [13]	BLUR + JPEG	0.12	0.37	0.36	0.25	0.83 (T)
LIVE-MD-ph2 [13]	BLUR + NOISE	0.01	0.53	0.42	0.02	0.52

PLCC

DATASET	DISTORTION	BRISQUE	PSNR	SSIM	MOS_{res}	MOS_{AI}
CSIQ	JPEG	0.85	0.90	0.93	0.93	0.87
MICT	JPEG	0.92	0.60	0.66	0.87	0.75
SDIVL	JPEG	0.54	0.76	0.82	0.71	0.29
TID2013	JPEG	0.83	0.93	0.90	0.92	0.83
VCL-FER	JPEG	0.79	0.58	0.82	0.94	0.74
LIVE-IQA-r1	JPEG	0.92	0.93	0.94	0.92	0.85
LIVE-IQA-r2	JPEG	0.97 (T)	0.94	0.95	0.90	0.86
MICT	JP2K	0.90	0.88	0.88	0.52	0.67
LIVE-IQA-r1	JP2K	0.92	0.92	0.91	0.69	0.78
LIVE-MD-ph1	BLUR+JPEG	0.12	0.37	0.36	0.27	0.83 (T)
LIVE-MD-ph2	BLUR+NOISE	0.16	0.52	0.37	0.01	0.53

SROCC

DATASET	DISTORTION	BRISQUE	PSNR	SSIM	MOS_{res}	MOS_{AI}
CSIQ	JPEG	0.63	0.56	0.43	0.37	0.51
MICT	JPEG	0.51	0.89	0.90	0.55	0.76
SDIVL	JPEG	0.77	0.64	0.60	0.54	0.85
TID2013	JPEG	0.40	0.28	0.26	0.24	0.38
VCL-FER	JPEG	0.56	0.70	0.49	0.31	0.56
LIVE-IQA-r1	JPEG	0.33	0.49	0.25	0.26	0.35
LIVE-IQA-r2	JPEG	0.26 (T)	0.31	0.38	0.42	0.50
MICT	JP2K	0.60	0.64	0.65	1.06	0.87
LIVE-IQA-r1	JP2K	0.35	0.45	0.41	0.69	0.47
LIVE-MD-ph1	BLUR+JPEG	0.49	0.45	0.46	0.47	0.27 (T)
LIVE-MD-ph2	BLUR+NOISE	0.54	0.46	0.49	0.54	0.46

RMSE

Deep NN based AIOs: Results

TABLE V
RESULTS OF THE STATISTICAL TEST PERFORMED FOR COMPARING THE PLCC VALUES PROVIDED BY THE DIFFERENT METRICS ON ALL THE DATASETS. CONSIDERING THE DATASETS ORDERED AS THEY APPEAR IN TABLE II, THE K-TH DIGIT OF THE BINARY SEQUENCE IN THE I-TH ROW AND J-TH COLUMN IS 1 IF AND ONLY IF ON THE K-TH DATASET, THE I-TH METRIC PERFORMED SIGNIFICANTLY BETTER THAN THE J-TH ONE WITH 95% OF CONFIDENCE. FOR INSTANCE, ON THE TID2013 DATASET (K=4) THE MOS_{RES} PERFORMED SIGNIFICANTLY BETTER THAN THE BRISQUE.

	BRISQUE	PSNR	SSIM	MOS_{res}	MOS_{AI}	Total
BRISQUE	- - - - - - - - - - -	0 1 0 0 1 1 0 0 1 0 0	0 1 0 0 0 0 1 0 0 0 0	0 0 0 0 0 0 1 1 1 0 0	0 1 0 0 0 0 1 1 1 0 0	13
PSNR	0 0 1 1 0 0 0 0 0 1 1	- - - - - - - - - - -	0 0 0 0 0 0 1 0 0 0 0	0 0 0 0 0 0 1 1 1 0 1	0 0 1 1 0 0 1 1 0 0 0	13
SSIM	1 0 1 1 0 1 0 0 0 1 1	1 0 0 0 1 1 0 0 0 0 0	- - - - - - - - - - -	0 0 0 0 0 0 0 1 1 0 1	1 0 1 1 0 1 1 1 1 0 0	19
MOS_{res}	1 0 1 1 1 1 0 0 0 0 0	1 1 1 0 1 1 0 0 0 0 0	0 1 0 0 1 0 0 0 0 0 0	- - - - - - - - - - -	1 1 1 1 1 1 1 0 0 0 0	18
MOS_{AI}	1 0 0 0 0 0 0 0 0 1 1	0 0 0 0 1 1 0 0 0 1 0	0 0 0 0 0 0 0 0 0 1 0	0 0 0 0 0 0 0 1 1 1 1	- - - - - - - - - - -	10

Deep NN based AIOs: Results

- Deep NNs based AIOs vs Actual ones
- Is the SROCC between an AIO and an actual observer similar to that between two actual observers?

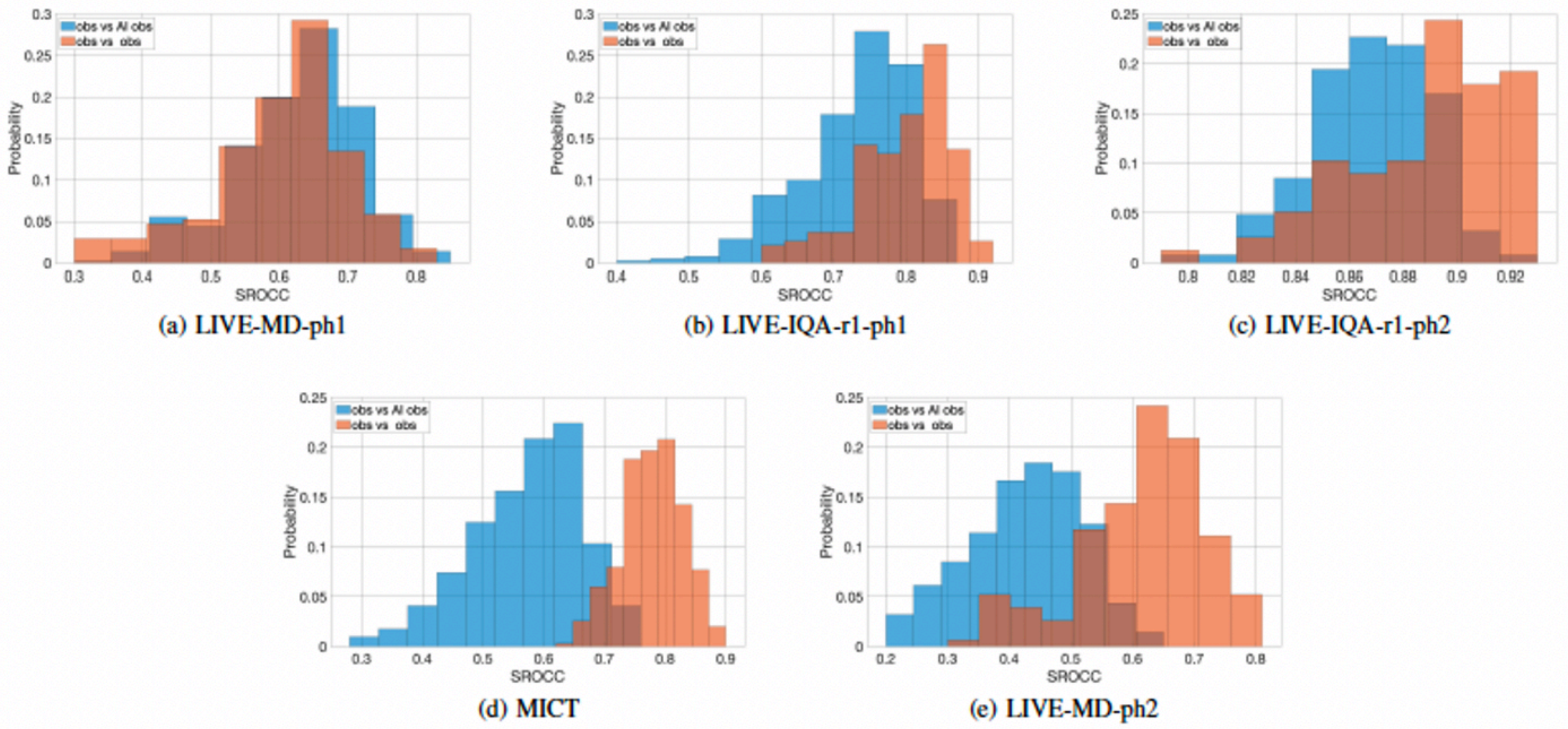
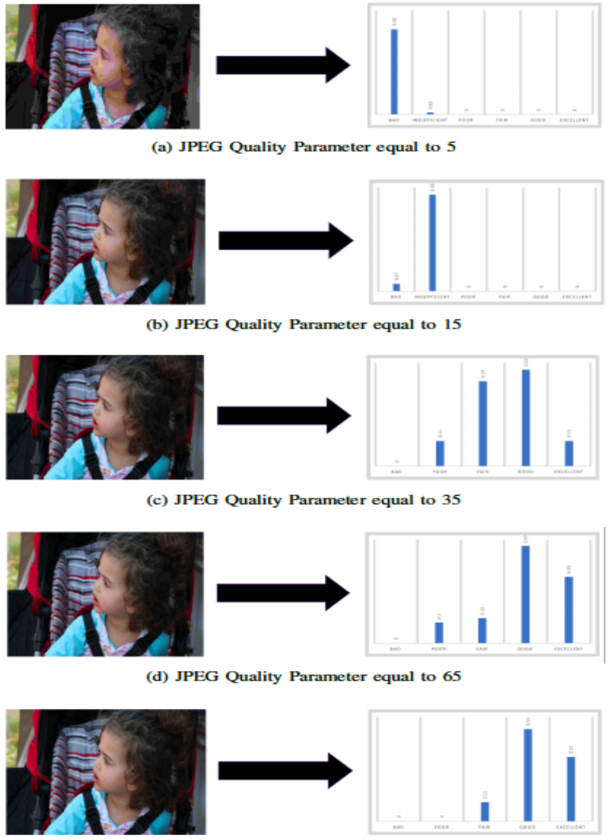


Fig: Comparing the correlation values observed between the actual observers and the ones of the actual observers and AIOs. The higher is the overlap, the better it is.

Deep NN based AIOs: Results

- AIOs enable the estimation of the user's opinion distribution



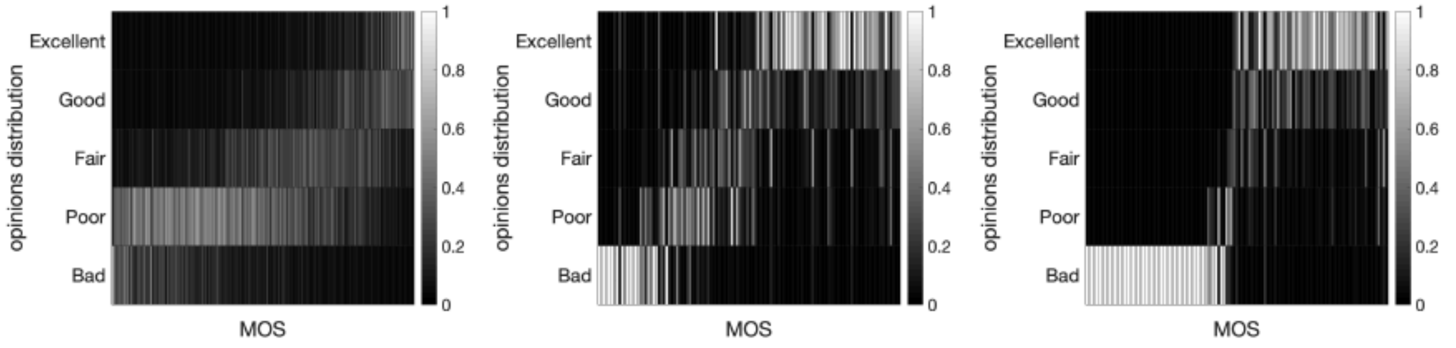
(a) JPEG Quality Parameter equal to 5

(b) JPEG Quality Parameter equal to 15

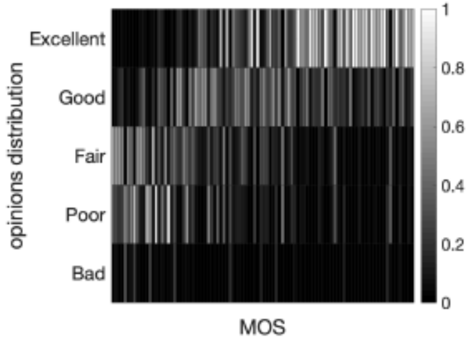
(c) JPEG Quality Parameter equal to 35

(d) JPEG Quality Parameter equal to 65

Performance of the deep NN based AIOs on 5 images obtained by progressively compressing a source image



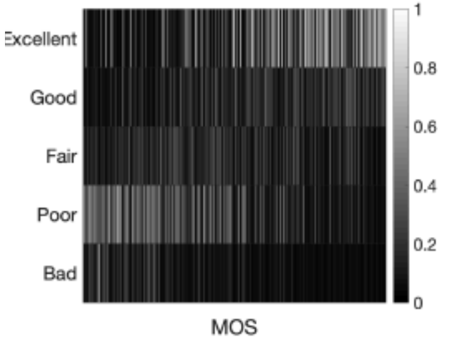
(a) LIVE-MD-ph1 (b) LIVE-IQA-r1-ph1 (c) LIVE-IQA-r1-ph2



(d) MICT

Dataset	Percentage of images
LIVE-MD-ph1	100%
LIVE-IQA-ph1	66%
LIVE-IQA-ph2	90%
MICT	50%
LIVE-MD-ph2	68%
Average	75%

Kolmogorov–Smirnov test of distribution similarity



(e) LIVE-MD-ph2

The predicted distribution of the user opinions for each image as a function of its MOS. Note that the mode of the distribution tends to increase as the MOS increases. Furthermore, as expected, the distribution is concentrated around the value of the mode in most of the cases.

Publications

- L. Fotio Tiotsop, T. Mizdos, M. Uhrina, P. Pocta, M. Barkowsky, E. Masala, "Predicting Single Observer's Votes from Objective Measures using Neural Networks". In Proceedings of Human Vision and Electronic imaging (HVEI) Conference 2020.
- L. Fotio Tiotsop, T. Mizdos, M. Barkowsky, P. Pocta, A. Servetti, E. Masala, " Mimicking individual media quality perception with neural network based artificial observers ". Submitted to the ACM Transactions on Multimedia computing communications and applications journal.
- L. Fotio Tiotsop, A. Servetti, T. Mizdos, M. Uhrina, P. Pocta, G.Van Wallendael, M. Barkowsky, E. Masala, "Deep Neural Networks based Artificial Observers for No Reference Image Quality Assessment". Submitted to the IEEE Transactions on Image Processing journal.

Thank you for your attention