Multiple Image Distortion DNN Modeling Individual Subject Quality Assessment

LOHIC FOTIO TIOTSOP
JEG-HYBRID GROUP

VQEG MEETING: JULY 2024

Artificial Intelligence-Based Observers: AIOs

- Training a Deep Neural Network (DNN) to mimic individual quality perception is a recent direction
- These DNNs are called **Artificial Intelligence-based Observers (AIOs)**
- An AIO outputs a five-class probability distribution on the five-point ACR scale
- **Aim and scope:** designing media processing systems that account for the characteristics of the targeted audience
  - Going beyond MOS and SOS
  - Toward users' characteristics-aware rendering systems

Research Challenges

- **Learning task with noisy labels:**
  - Individual opinion scores are noisier than the MOS
  - Need for complex models and large number of training samples
- **Limited number of training samples:**
  - Difficult to gather a very large number of opinion scores from the same subject
  - This hinders the model's capacity to make inference accurately
  - Mitigation strategies:
    - Transfer learning
    - Data augmentation
  - We have introduced a training approach named **“human in the loop”**

Human in the Loop Training (HLT) of DNN-based AIOs

- Placing the subject in an iterative procedure to derive their AIO
  1. Pretrain a DNN with many hidden convolutional layers on a **synthetically annotated large-scale dataset**
  2. Ask the subject to rate a selected set of stimuli (typically no more than 500)
  3. **Continue the training of the pretrained DNN** using the collected opinion scores.
  4. Use the obtained DNN **to make inference on a large-scale dataset**.
  5. **Select a new set of stimuli for the subject to rate**, specifically those with DNN-predicted quality prone to **high uncertainty**. Return to step 2 with the newly selected stimuli
- The DNN obtained at the last iteration represents the subject’s AIO

The Multi-Distortion ResNet50 (MDResNet50)

- **Step 1 of HLT:** We pretrained a DNN named **MDResNet50** on a synthetically annotated dataset
- The created synthetically annotated dataset includes **2 million images**
  - 100 000 high quality images were selected from the ImageNet dataset
  - 5 different levels of blur, noise, JPEG and JPEG2K compression were applied to each image
  - This yielded 2 million images, the quality of each annotated by the corresponding level of distortion
- We trained the MDResNet50 to recognize the five levels of impairment for all the four considered distortions
  - Network architecture: that of the ResNet50
  - Initial weights: those of the ResNet50

Table 1. Rules to synthetically annotate distorted images.

Distortion	Parameter	Distortion Levels and Corresponding Labels				
Noise	σ_{noise}	[0.21, 2.00]	[0.10, 0.21)	[0.05, 0.10)	[0.02, 0.05)	[0.00, 0.02)
Blur	σ_{filter}	[3.19, 6.00]	[2.13, 3.19)	[1.32, 2.13)	[0.66, 1.32)	[0.00, 0.66)
JPEG	q	[0.00, 12.00)	[12.00, 16.89)	[16.89, 28.72)	[28.72, 49.14)	[49.14, 100.00]
JP2K	br	[0.00, 0.05)	[0.05, 0.25)	[0.25, 0.50)	[0.50, 0.86)	[0.86, 3.00]
Quality labels		*Bad*	*Poor*	*Fair*	*Good*	*Excellent*

MDResNet50 vs ResNet50

- The MDResNet50 should be a better starting point for transfer leaning in image quality assessment tasks

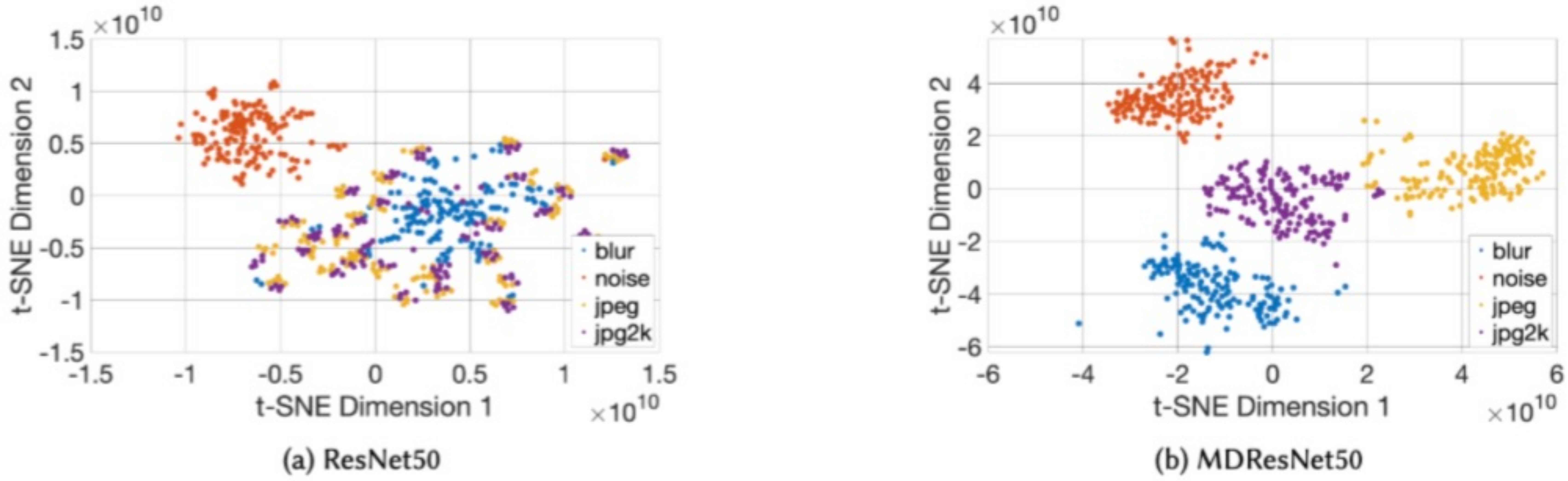


Fig. 12. 2D t-SNE maps of the features extracted by the ResNet50 (left) and the MDResNet50 (right). It can be noticed that the MDResNet50 better distinguishes among the different image distortions.

From the MDResNet50 to AIOs

- In [1] we applied three iterations of HLT, creating and releasing a new subjectively annotated datasets as well as 5 AIOs.

- In [2] a single iteration of HLT was applied using existing datasets to train 19 AIOs
  - Impossible to do more than one iteration as modelled subjects were unknown

[1] Majer, Pavel; Fotio Tiotsop, Lohic and Borkowski, Marcus. **"Training the DNN of a Single Observer by Conducting Individualized Subjective Experiments."** In 2023 15th International Conference on Quality of Multimedia Experience (QoMEX), pp. 103-106. IEEE, 2023

[2] Fotio Tiotsop, Lohic; Servetti, Antonio; Pocta, Peter; Van Wallendael, Glenn; Barkowsky, Marcus and Masala, Enrico. **"Multiple Image Distortion DNN Modeling Individual Subject Quality Assessment."** ACM Transactions on Multimedia Computing, Communications and Applications (2024)

Some Results

Trained AIOs Accuracy

- The predicted opinion scores by AIOs correlate with the MOS as good as the opinion scores of the mimicked subjects in many testing conditions

Table 3. The Spearman Rank Order Correlation Coefficient (SROCC) between the opinion scores of each one of the mimicked real subjects and the MOS on the training set, i.e., the MD-LIVE-IQA dataset.

Subjects	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
SROCC	0.84	0.82	0.91	0.83	0.88	0.88	0.83	0.83	0.84	0.87	0.85	0.93	0.85	0.60	0.85	0.81	0.93	0.84	0.82

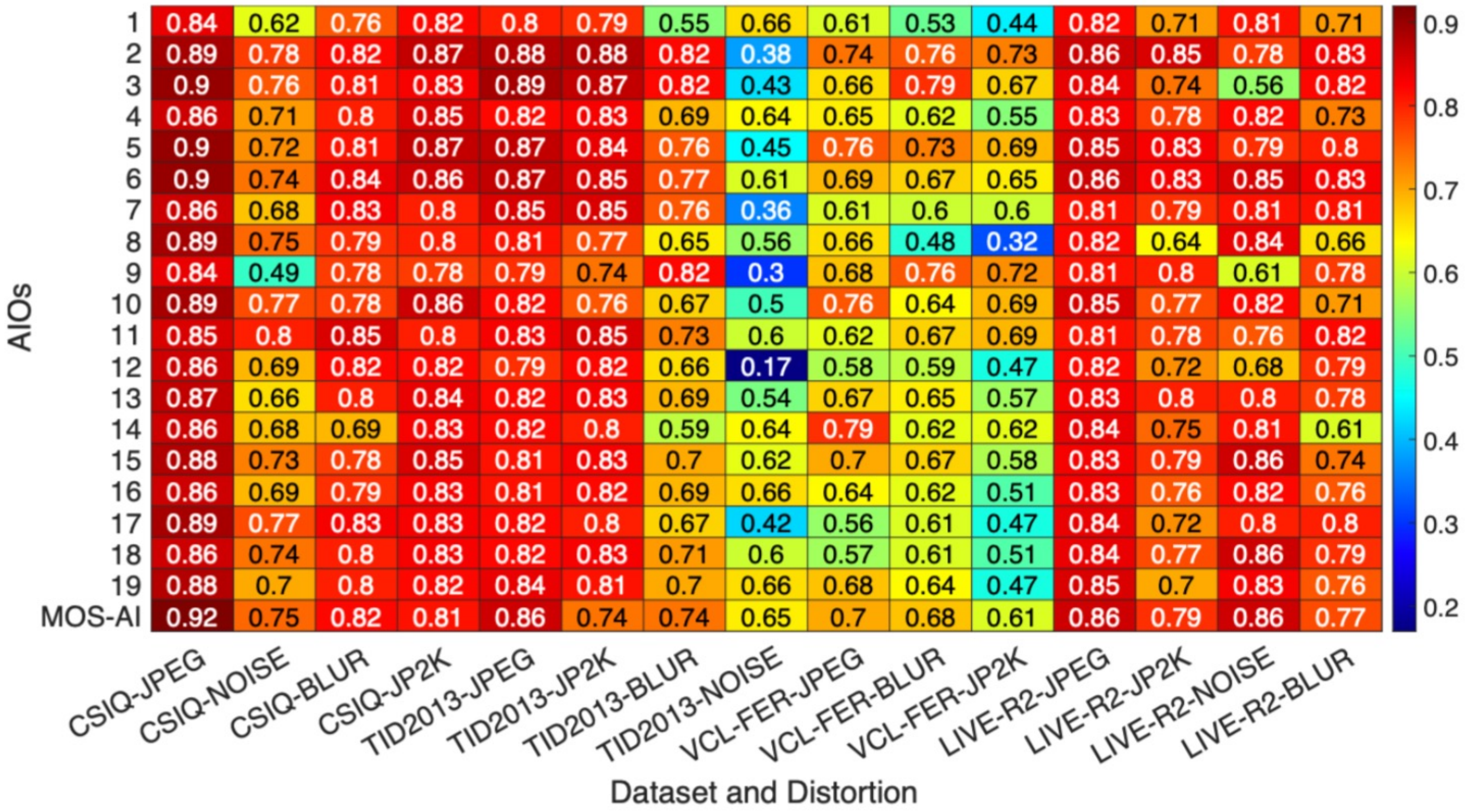


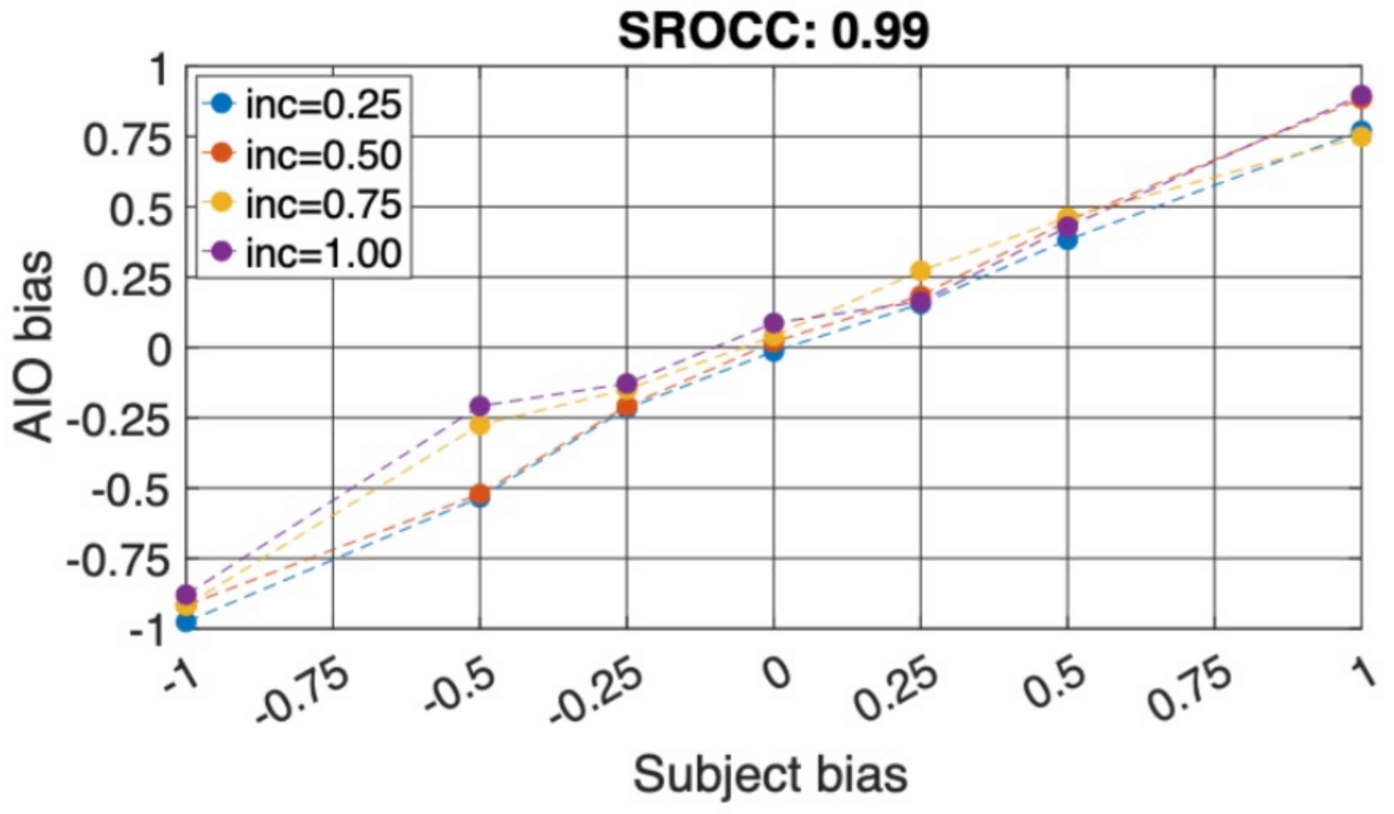
Fig. 6. The SROCC between the prediction of each AIO and the MOS for the different datasets and distortions.

AIOs Mimic Subjects' Bias and Inconsistency

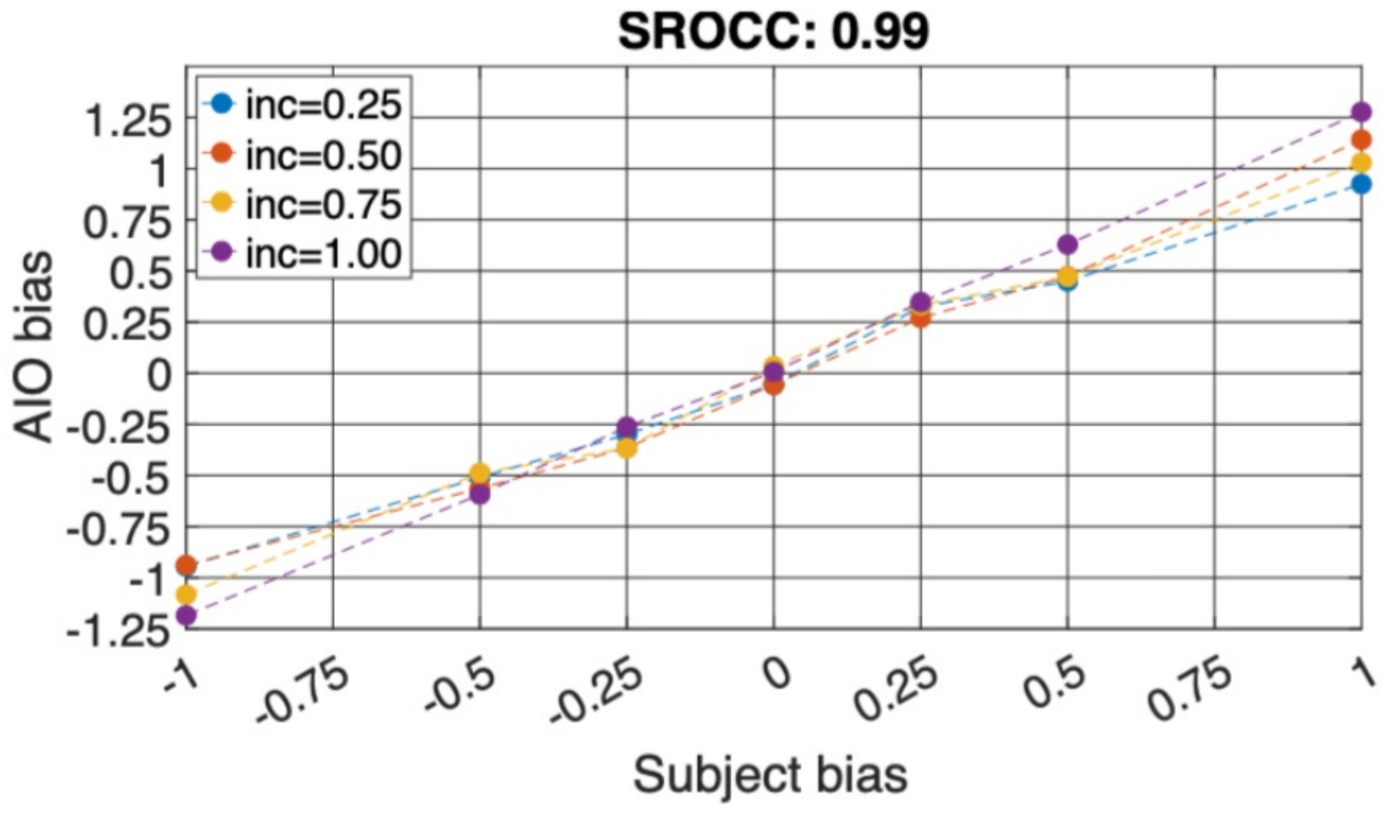
- We **simulated** the ratings of a subject with **specific bias and inconsistency**
  - We used the scoring model: $u_{ij} = \psi_j + \Delta_i + N(0, v_i)$, ***implemented in Sureal***
  - The MOS of each image was considered as an estimate of its ground truth quality
  - Values of bias were uniformly chosen in the range [-1 1]
  - Values of inconsistency were uniformly chosen in the range [0.25 1]

- The **AIO of a subject with a specific bias and inconsistency was then trained**
  - The AIO bias is measured as the average deviation of its opinion scores from the MOS of the AIOs
  - The AIO inconsistency is measured as the average of the variances of its predictions (variance of the softmax layer)

AIOs Bias vs Ground Truth Bias



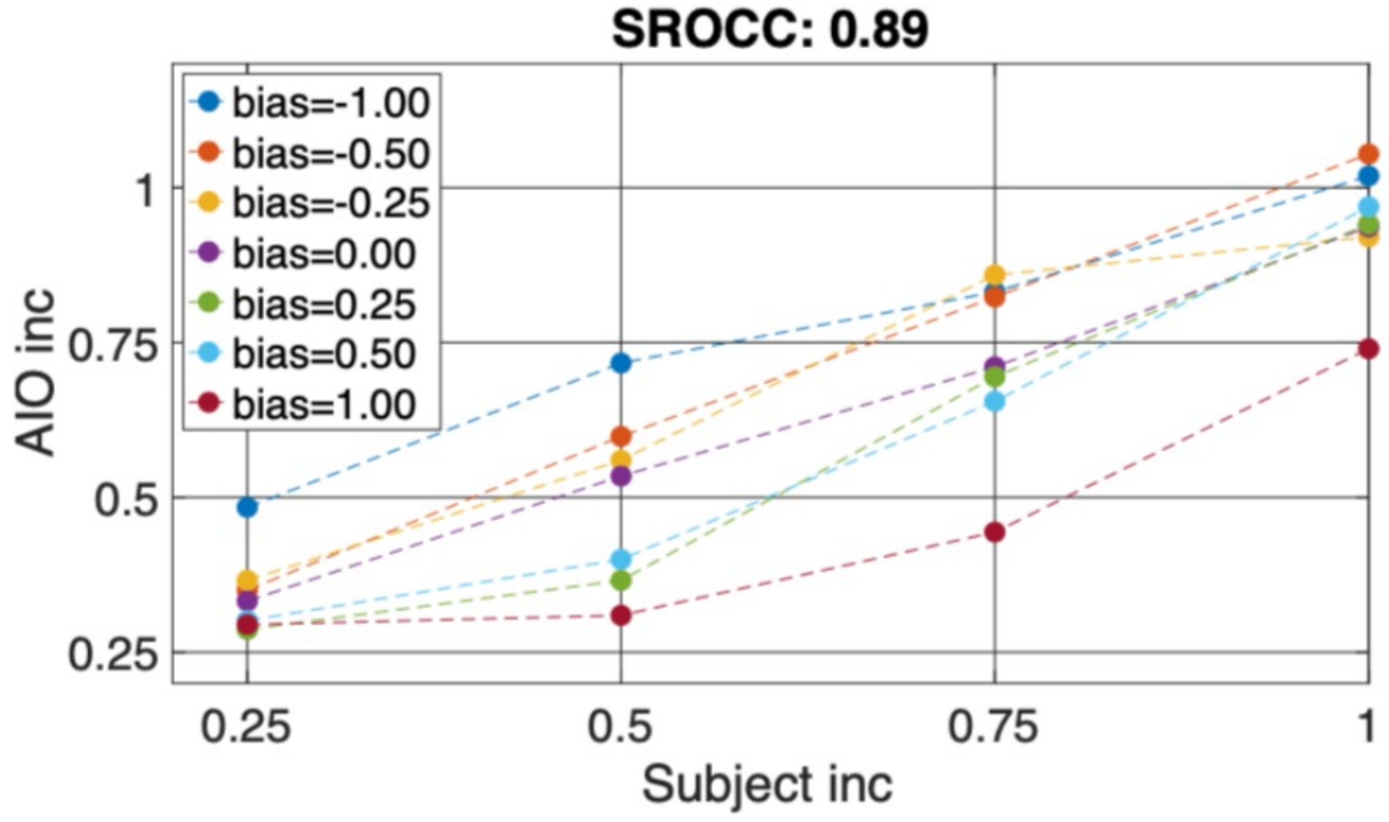
(a) Bias, LIVE IQA



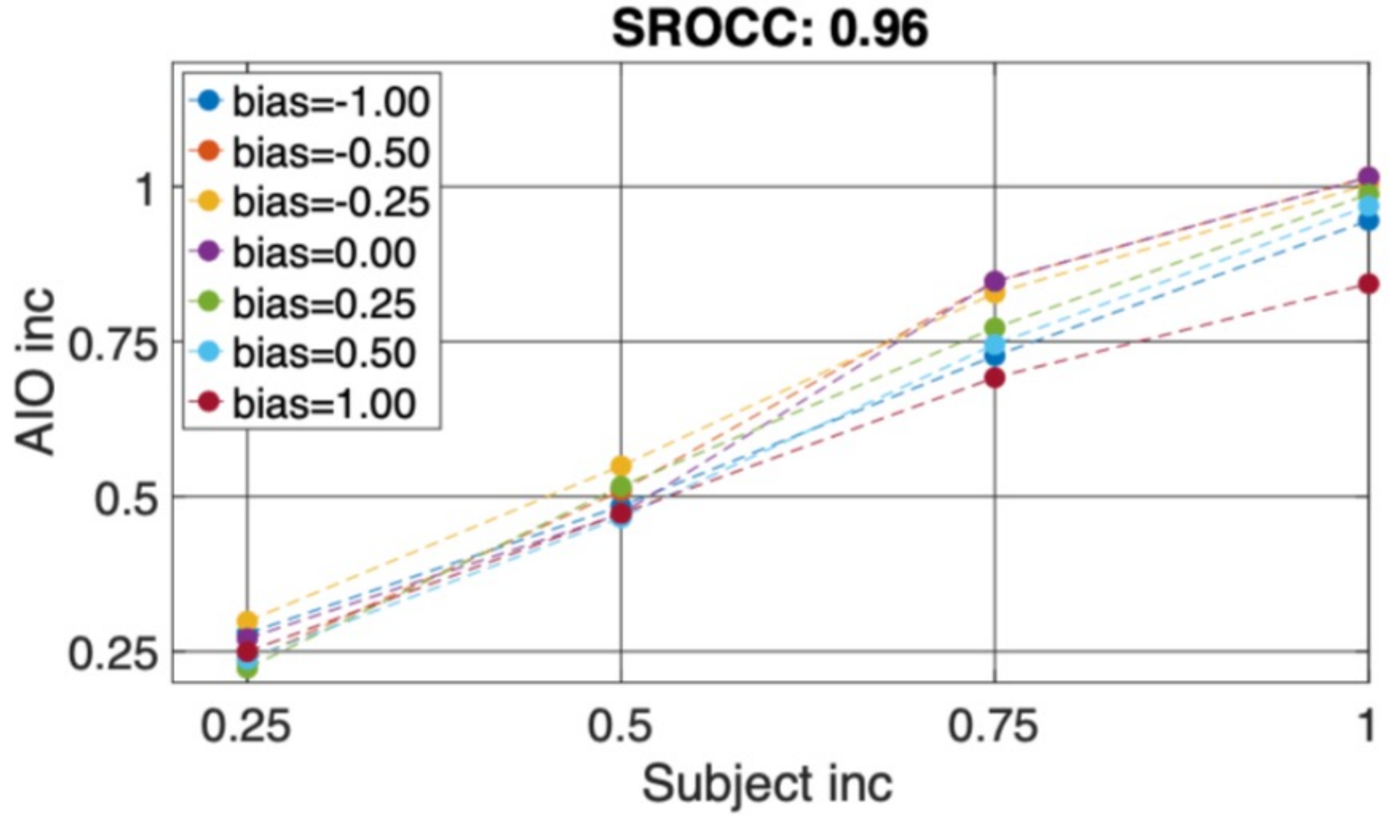
(c) Bias, KonIQ-10k

- Whatever is a subject's inconsistency, their AIO mimics their expectation in terms of quality (measured in terms of bias).

AIOs Inconsistency vs Ground Truth Inconsistency



(b) Inconsistency, LIVE IQA



(d) Inconsistency, KonIQ-10k

- The average variance of the softmax layer of an AIO is correlated with the subject's inconsistency
- Thus, the model's confidence is correlated to the subject scoring accuracy, this does not happen in most of the computer vision tasks

Do AIOs Mimic More than Bias and Inconsistency?

Next steps:

- Evaluating through simulation and real data whether AIOs mimic content preference
  - e.g. putting bias on specific ratings, rather than adding a permanent bias
  - running subjective texts that include amongst stimuli, pictures on activities that are of high interest for each participant

- Evaluating whether an AIO mimics the perception of aesthetics of the related subject
  - e.g. does my AIO see blur as an aesthetic artifact, when I do ?

AIOs Sensitivity to Colour Saturation

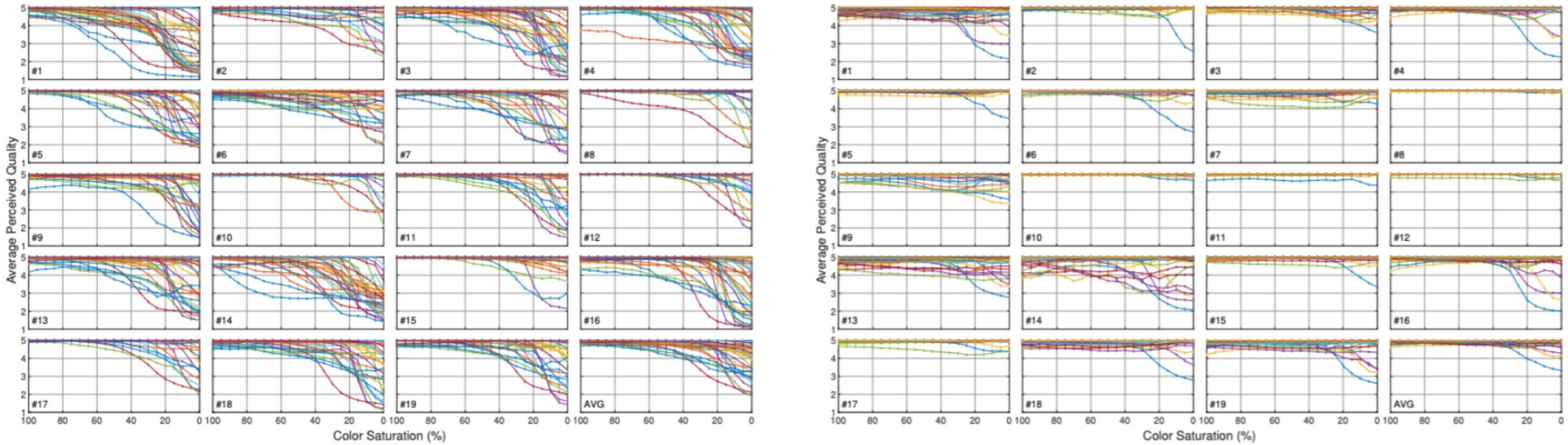
- **Reasonable assumption:** an accurate subject asked to assess the disturbance of artifacts caused by noise, blur, JPEG, and JPEG2K compression would assign **very similar opinion scores** to images (a) to (e) in a single stimulus test.

(a) perc=80% (b) perc=60% (c) perc=40% (d) perc=20% (e) perc=0%

Fig. 9. Visual effect of the color saturation reduction applied to the images in our sensitivity analysis. The modification is measured in terms of the percentage of the remaining color saturation with respect to the original image. We expect that a subject tasked to rate distortion caused by noise, blur, JPEG and JPEG 2000 compression does not perceive a significant difference in the image quality when moving from (a) to (e).

AIOs Sensitivity to Colour Saturation

- The integration of the **aforementioned assumption**, through **data augmentation,** yielded AIOs with **a sensitivity to colour saturation that more resembles that of a human observer**



(a) AIOs published in [53]

(b) Proposed AIOs

Fig. 11. The average perceived quality by the state-of-the-art AIOs published in [53] (left) and the proposed AIOs (right) as a function of the percentage of the color saturation of the input image. Each plot (excluding the one labeled as "AVG") corresponds to one AIO and each curve in each plot shows the trend of the perceived quality by the AIO as the image color saturation decreases. The plot labeled as "AVG" shows the average of the trends exhibited by the 19 AIOs.

AIOs Sensitivity to Noise

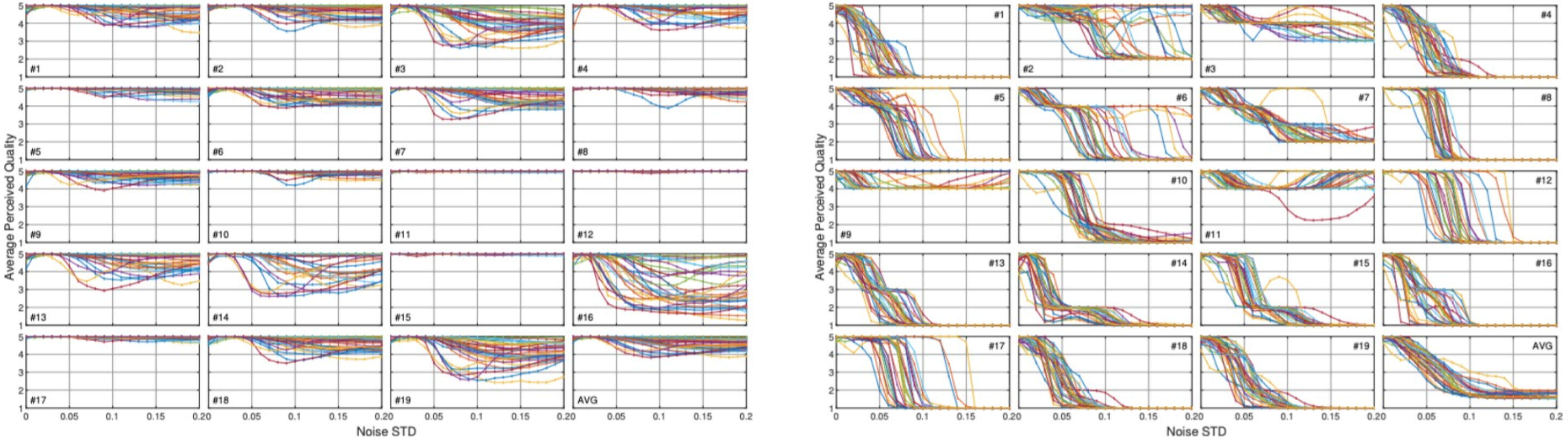
- **Reasonable assumption:** an accurate subject asked to assess the disturbance of artifacts caused by noise, blur, JPEG, and JPEG2K compression would assign **decreasing opinion scores** to images from (a) to (e) in a single stimulus test**.**

(a) std=0.05 (b) std=0.15 (c) std=0.25 (d) std=0.35 (e) std=0.45

Fig. 8. Visual effect of the Gaussian noise applied to the images in our sensitivity analysis. The level of degradation is controlled by the standard deviation (std) of the Gaussian noise. We expect that a subject tasked to rate distortion caused by noise, blur, JPEG and JPEG2000 compression perceives a decrease in the image quality when moving from (a) to (e).

AIOs Sensitivity to Colour Saturation

- Considering several distortions and adding a "not perceptible" noise for **data augmentation** at the first step of HLT enabled the training of AIOs whose sensitivity to noise more resembles that of a human observer



(a) AIOs published in [53]

(b) Proposed AIOs

Fig. 10. The average perceived quality by the state-of-the-art AIOs published in [53] (left) and the proposed AIOs (right) as a function of the standard deviation of the Gaussian noise added to the input image. Each plot (excluding the one labeled as "AVG") corresponds to one AIO and each curve in each plot shows the trend of the perceived quality by the AIO as the standard deviation of the added noise increases. The plot labeled as "AVG" shows the average of the trends exhibited by the 19 AIOs.

Summary of Sensitivity Analysis

- The size of current subjectively annotated datasets with individual opinion scores is not large enough to expect that AIOs goes beyond prediction, mimicking aspects they are not trained for

- Researchers, however, have some consolidated knowledge about the subjects' scoring behaviour

- To obtain AIOs that mimic rather than predicting, we should consider expressing such a knowledge as assumptions

- These assumptions can then be integrated into the training process through data augmentation strategies or constraints

Ongoing Research

- **Objective:** to train AIOs that directly receive high-resolution images (HD, FHD, and potentially higher resolutions) as input
- **Key challenges:**
  - Need for a new large-scale subjectively annotated dataset:
    - Current **dataset**: 22,000 pristine quality images collected from professional photography websites
    - Requirement: need for subjective annotations gathered through HLT
  - Hypothetical Reference Circuits (HRCs):
    - Selection: determine which HRCs to apply
    - Resolution: decide the appropriate resolution for HRC application
    - Presentation: establish the resolution at which images should be displayed to subjects and the DNNs
  - Challenges in DNN Training with High-Resolution Images:
    - Currently implementing and testing bilinear and spatial pyramid pooling layers for AIOs training

Thank you for your attention