Human-in-the-Loop Training Process of the DNN of a Single Observer

LOHIC FOTIO TIOTSOP
JEG-HYBRID GROUP

VQEG MEETING: DECEMBER 2023

Artificial Intelligence-Based Observers: AIOs

- AIO: neural network trained to mimic the quality perception of an individual subject

- 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

Research Challenges

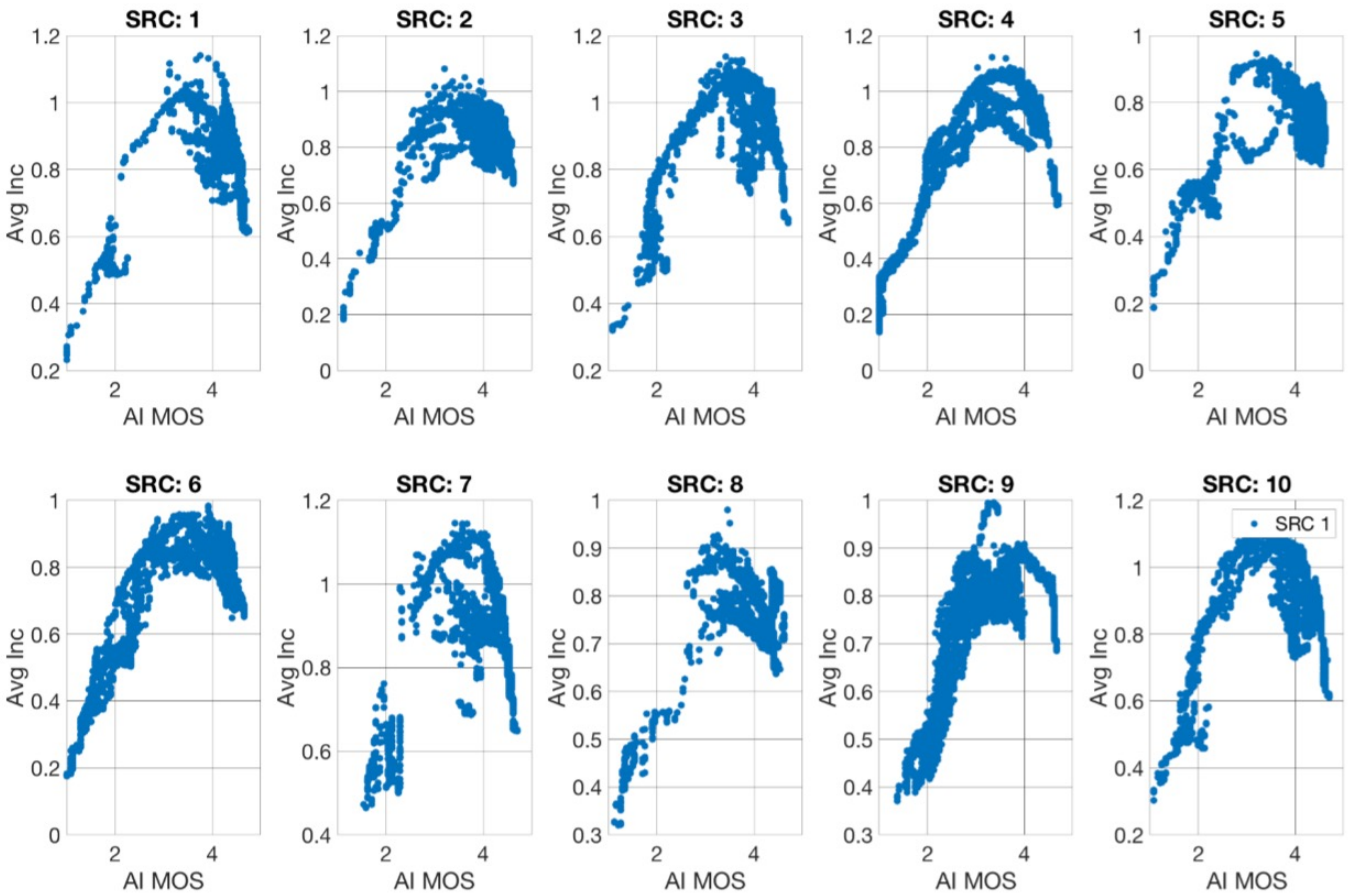
- **Learning task with noisy labels:**
  - Individual opinion scores are noisier than the MOS
  - Need for complex models and large number of training samples
- **Limited number of training samples:**
  - Difficult to gather a very large number of opinion scores from the same subject
  - This hinders the model's capacity to make inference accurately
  - Mitigation strategies:
    - Shallow NN architectures
    - Transfer learning
    - Data augmentation

Shallow NN-based AIOs

- Hand-crafted features regressed to one-hot encoding of individual opinion scores
- Optimal number of hidden layers
  - Vary from subject to subject
  - Experimental evidences: 1 to 3 hidden layers, each with no more than 5 neurons
- Very simple approach, yet effective to some extent
  - Shallow NN-based AIOs mimic the SOS hypothesis

Shallow NN-based AIOs at Quality Scale Extremes

- The variance of the predicted distribution models the AIO inconsistency
- **AIOs mimic the lower inconsistency of subjects at quality scale extremes**



Fotio Tiotsop, L., Mizdos, T., Barkowsky, M., Pocta, P., Servetti, A. and Masala, E., **2022**. Mimicking individual media quality perception with neural network based artificial observers. **ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)**, 18(1), pp.1-25.

Mean Opinion Score of AIOs (AI MOS) vs their average inconsistency (Avg Inc) on the quality of 20,000 stimuli generated from 10 different sources

Limits of Hand-Crafted Features

- **Input signal approximation**
  - The AIO does not receive, as input, the signal that the subject rated
  - The AIOs accuracy strongly depends on the ability of the features to represent the actual signal shown to the subject
- **Over-generalization**
  - Bob and Alice might perceive the same distortion differently: e.g., Bob might be mainly sensitive to blurred edges while Alice may have a more uniform perception of blur
  - Algorithms for hand-crafted feature computation often focus on a single idea, like assessing Gaussian filter-induced blur
- Deep NNs solve both issues
  - Currently using **deep CNNs** and **Vision Transformers**

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 at least 50 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

Subject Accuracy vs AIO Accuracy

TABLE I

ACCURACY OF REAL OBSERVERS TO PREDICT THEIR OPINIONS

Obs 1	Obs 2	Obs3	Obs 4	Obs 5
58%	57%	73%	62%	47%

TABLE II

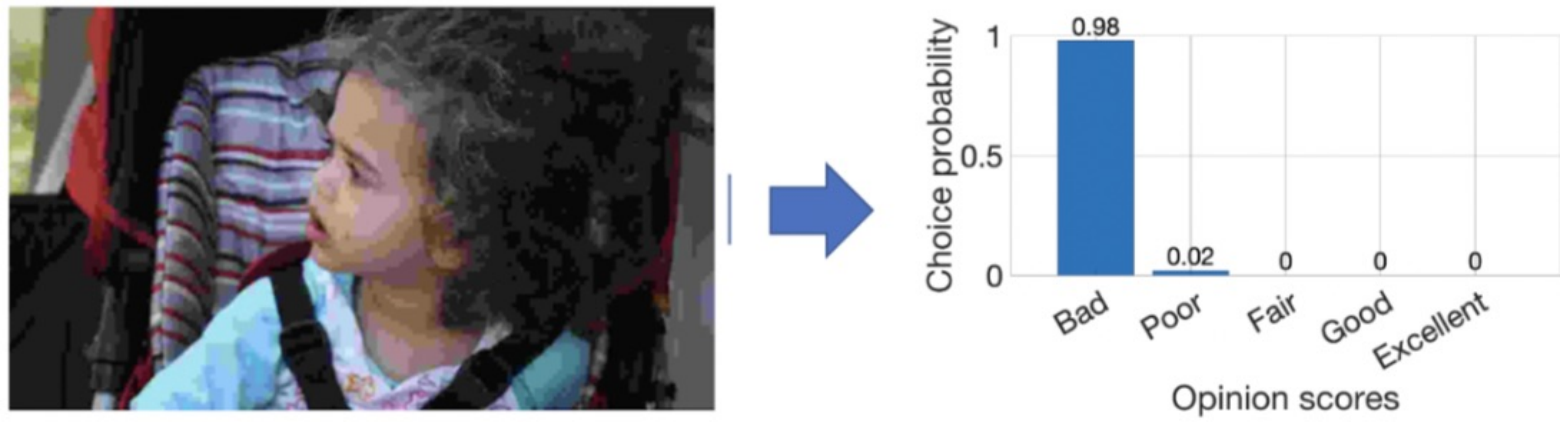
ACCURACY OF AIOS TO PREDICT THE OPINIONS OF REAL OBSERVERS

	Obs 1	Obs 2	Obs3	Obs 4	Obs 5
AIO_1	**63%**	59%	43%	39%	50%
AIO_2	59%	**67%**	39%	36%	52%
AIO_3	39%	41%	**57%**	**58%**	57%
AIO_4	45%	42%	52%	51%	59%
AIO_5	36%	35%	55%	53%	**60%**

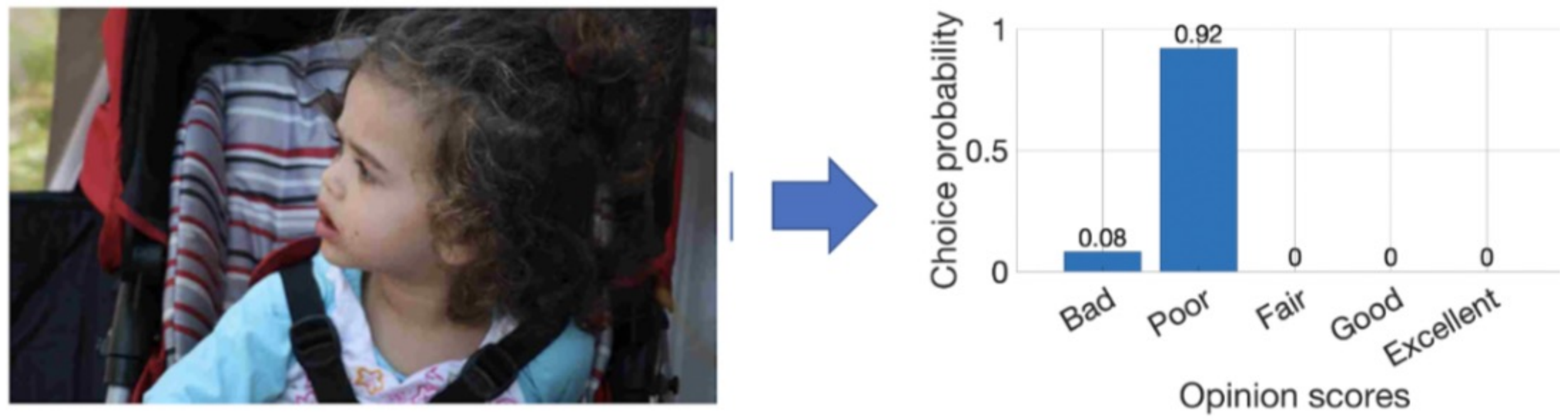
- These results were obtained by training 5 AIOs in 3 iterations of HLT
- Trained AIOs exhibit test set accuracy comparable to real observer repetition performance
- Each AIO excels at mimicking its associated real observer
- Deep CNNs and Vision Transformers showed comparable performances

P. Majer, L. Fotio Tiotsop and M. Barkowsky, "Training the DNN of a Single Observer by Conducting Individualized Subjective Experiments," 2023 15th International Conference on Quality of Multimedia Experience (**QoMEX**), Ghent, Belgium, **2023**, pp. 103-106.

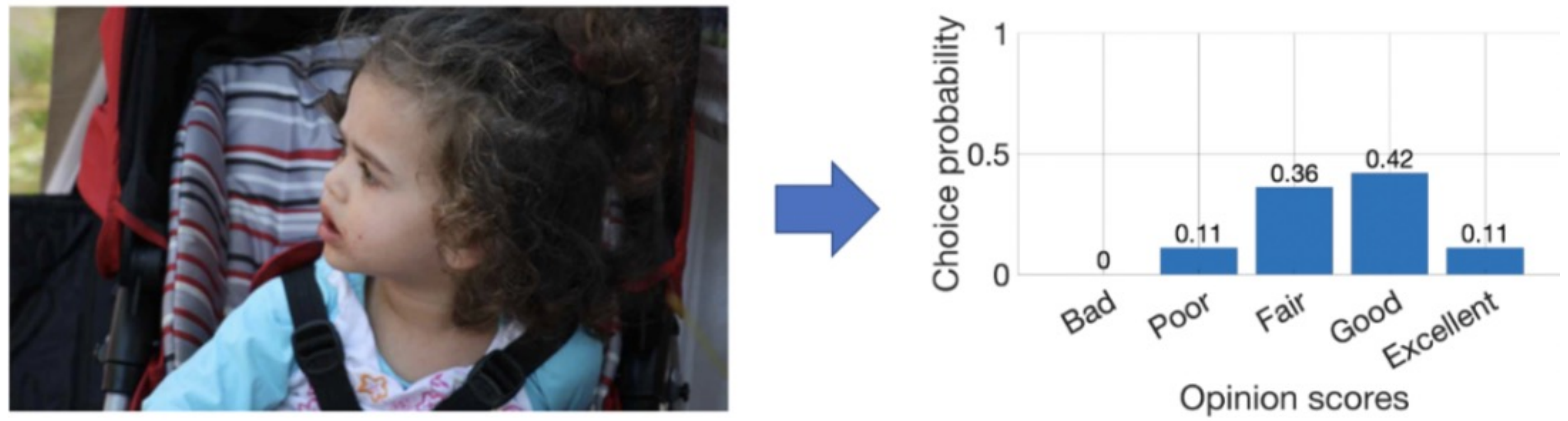
AIOs Distribution of Opinion Scores



(a) JPEG Quality Parameter equal to 5



(b) JPEG Quality Parameter equal to 15



(c) JPEG Quality Parameter equal to 35

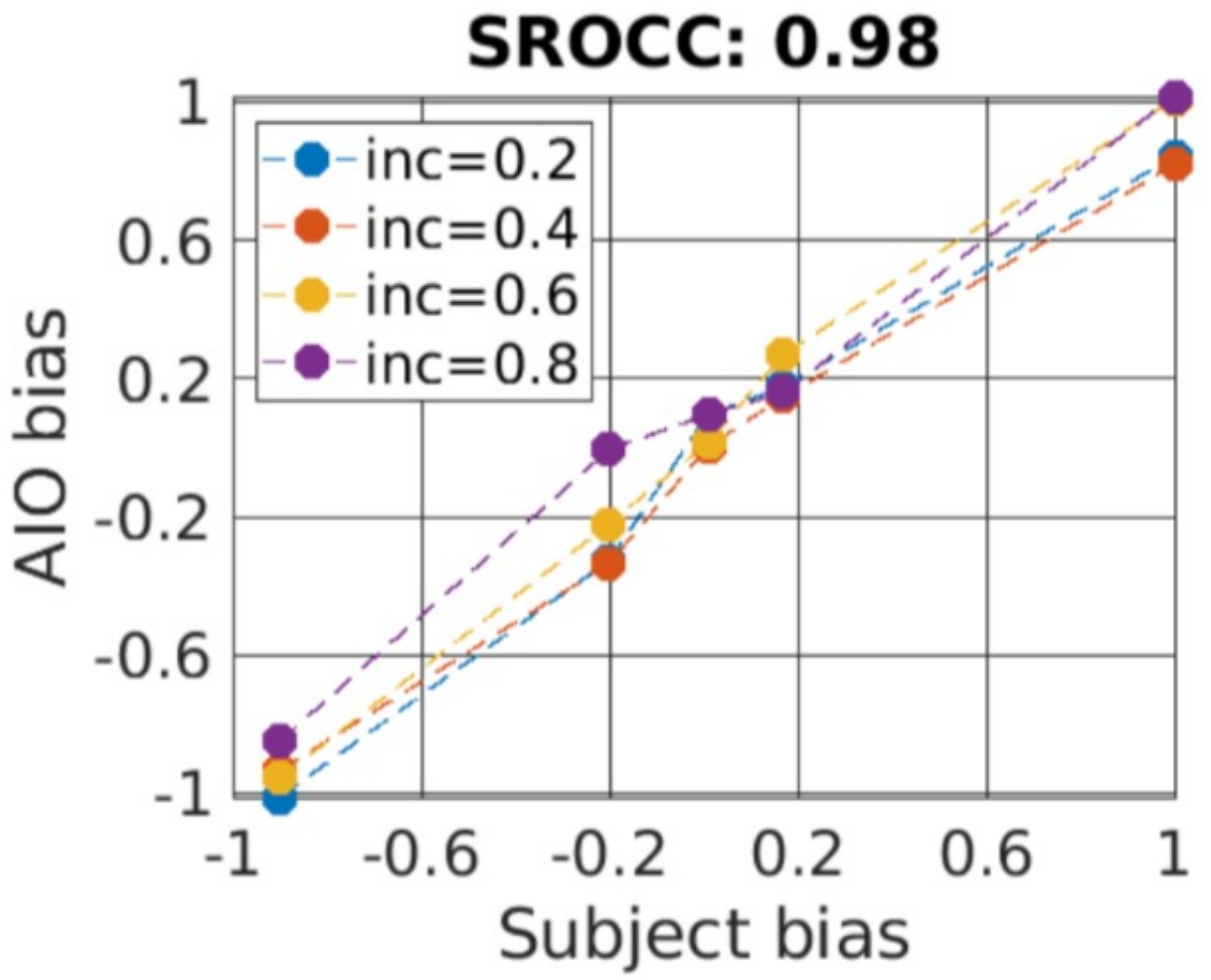
- This result was obtained by training 19 AIOs in 1 iteration of HLT
- The AIOs replicate the diversity of opinion scores as expected in a real subjective test

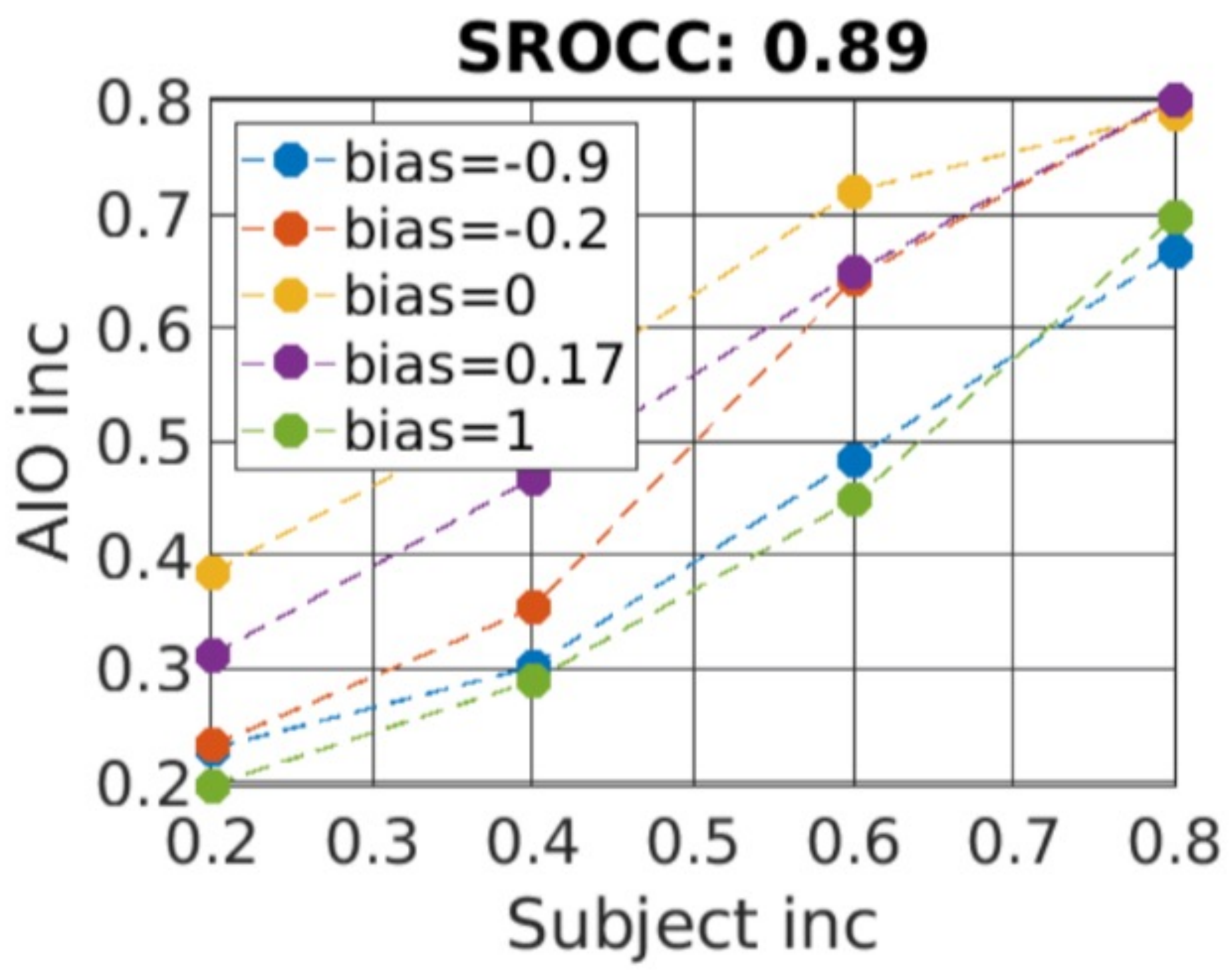
Fotio Tiotsop, L.., Servetti, A., Barkowsky, M., Pocta, P., Mizdos, T., Van Wallendael, G. and Masala, E., **2023**. Predicting individual quality ratings of compressed images through deep CNNs-based artificial observers. **Signal Processing: Image Communication**, 112, p.116917.

AIOs Mimic Bias and Inconsistency

- We simulated the opinion of 20 subjects on 1000 images
  - We used the scoring model of slide 9: $u_{ij} = \psi_j + \Delta_i + N(0, v_i)$
  - The ground truth qualities of the 1000 chosen images were known
  - Five bias and four inconsistency values were chosen to simulate the ratings of 20 different subjects
  - The AIO of a subject with a specific bias and inconsistency was then trained
- We then evaluated if the AIO replicates the bias and inconsistency of the associated subject on a test set
  - The AIO bias is measured as the average deviation of its opinion scores from the MOS-AI
  - The AIO inconsistency is measured as the average of the variances of its predictions

AIOs Mimic Bias and Inconsistency





The AIO characteristics correlate well to those of the corresponding subject

Future Work

1. Does HLT converge? How to define a suitable convergence criteria?
2. Can HLT be looked at as a standard way to collect opinion scores for AIOs training?
    1. Avoiding the collection of useless data
    2. Mitigating the effect of subjects' fatigue
3. Shall we use Deep CNNs or Vision Transformers?

Note: **we are currently looking for people willing to enter into the loop. Volunteers will be asked to rate new sets of stimuli until the training loop of their AIO converges.**