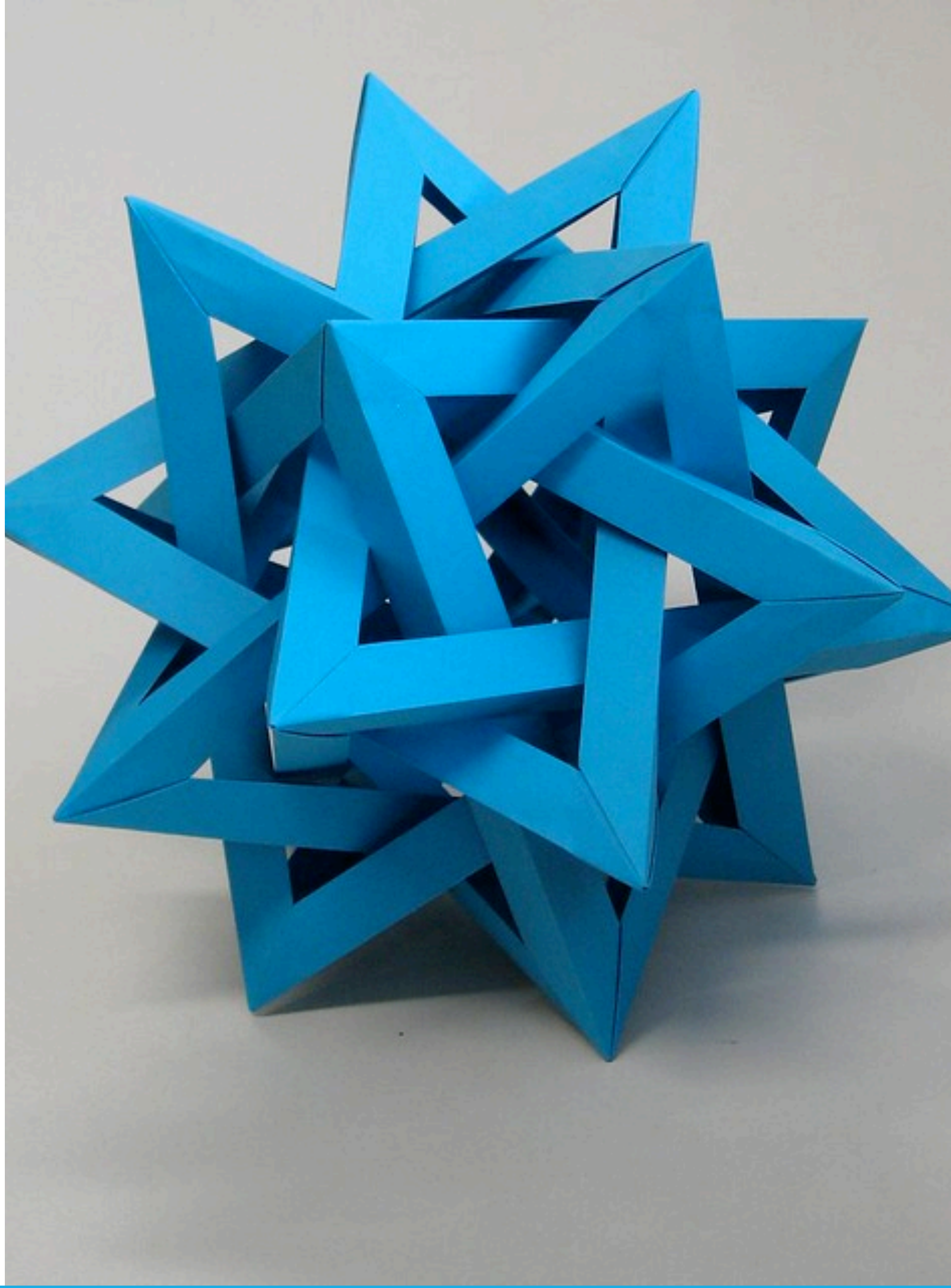
Politecnico di Torino

On the Sensitivity of Artificial Intelligence-based Observers to Input Signal Modification

LOHIC FOTIO TIOTSOP

JEG-HYBRID

VQEG MEETING, MAY 2022



Artificial Intelligence-based Observer (AIOs)

- AIO: a NN trained to mimic an individual subject in terms of quality perception:
  - Predicting individual opinion scores rather than the MOS;
  - The AIO outputs a discrete probability distribution;
  - The realizations of such a distribution mimic the subjects’ inability to repeat previous assessment.

- Challenges:
  - Dealing with a learning task with very noisy labels;
  - The lack of training samples is further emphasized.

AIOs Training Approaches

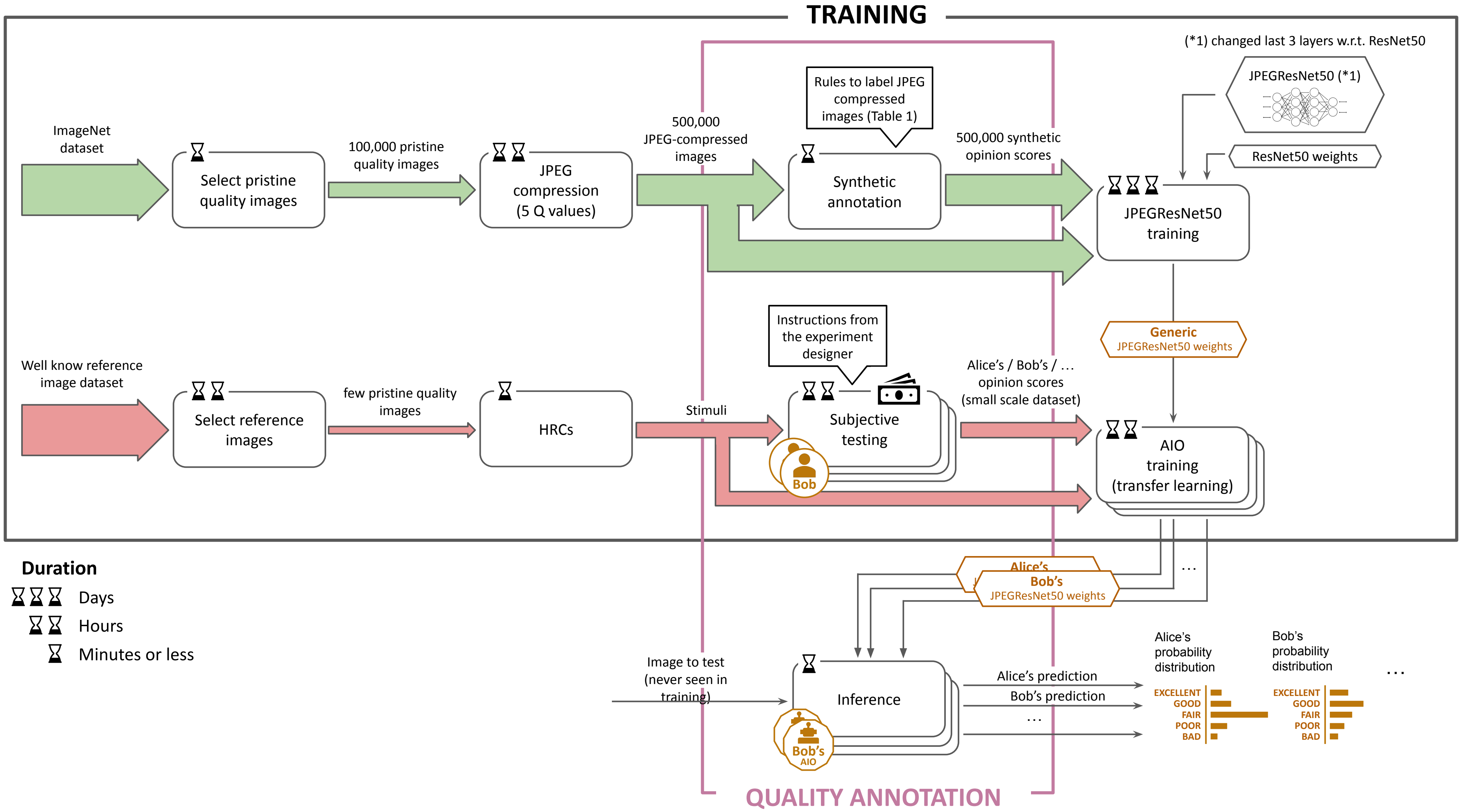
- Shallow NNs based AIOs: [1]
  - A common set of hand-crafted features;
  - Finding the best feature subset for each subject to be modelled;
  - Mapping selected features to subject ratings with a NN;
  - The NN architecture and the best feature subset change from one subject to the other.

- Deep CNNs based AIOs: [2]
  - Avoiding input signal approximation;
  - Avoiding hand-crafted features “over-generalization”;
  - Extracting “opinion-aware” individual features.

[1] L. Fotio Tiotsop, T. Mizdos, M. Barkowsky, P. Pocta, A. Servetti, E. Masala. ”Mimicking Individual Media Quality Perception with Neural Networks based Artificial Observers”. In: ACM Transactions on Multimedia   Computing, Communications and Applications

[2] L. Fotio Tiotsop, A. Servetti, T. Mizdos, M. Uhrina, P. Pocta, G.Van Wallendael, M. Barkowsky, E. Masala. ”Predicting Individual Quality Ratings of Compressed Images through Deep CNNs-based Artificial. Observers”. Submitted to Signal Processing: Image Communication.

Deep CNNs based AIOs: Training Approach



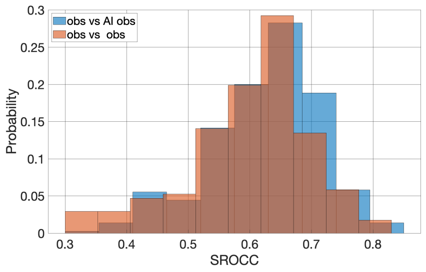
Results Discussed at Previous VQEG Meetings

- Shallow NNs based AIOs:
  - Show an accuracy that differs less than 12% from a benchmark value derived from a subjective test;
  - The AIOs' MOS correlates well with actual observers' MOS;
  - More accurate at the extreme of quality scale as humans;
  - The variance of the predicted probability distribution has the properties of a subject inconsistency measure.

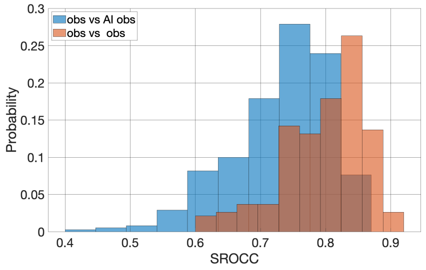
- Deep CNNs based AIOs:
  - Similar results as in the shallow NNs-based AIOs case;
  - Preliminary results on the comparison of AIOs to actual observers in terms of bias and inconsistency.

More Results on Deep CNN-based AIOs

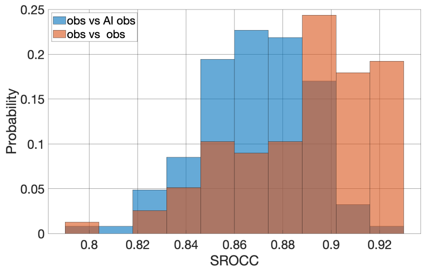
- Simulating the process of a subjective test:
  - Using the AIOs on the stimuli used in a subjective test;
  - Compute the values: corr(AIO, Observer) and corr(Observer, Observer);
  - Question of interest: can we trivially distinguish the AIOs from actual observers?



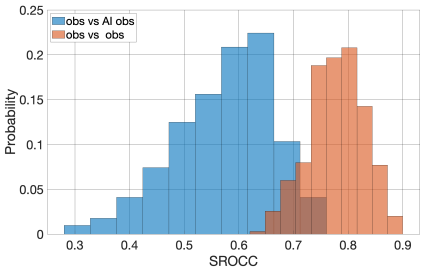
Training set



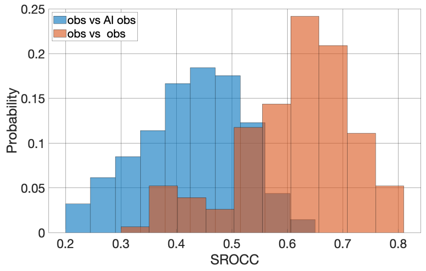
LIVE IQA PH1



LIVE IQA PH2

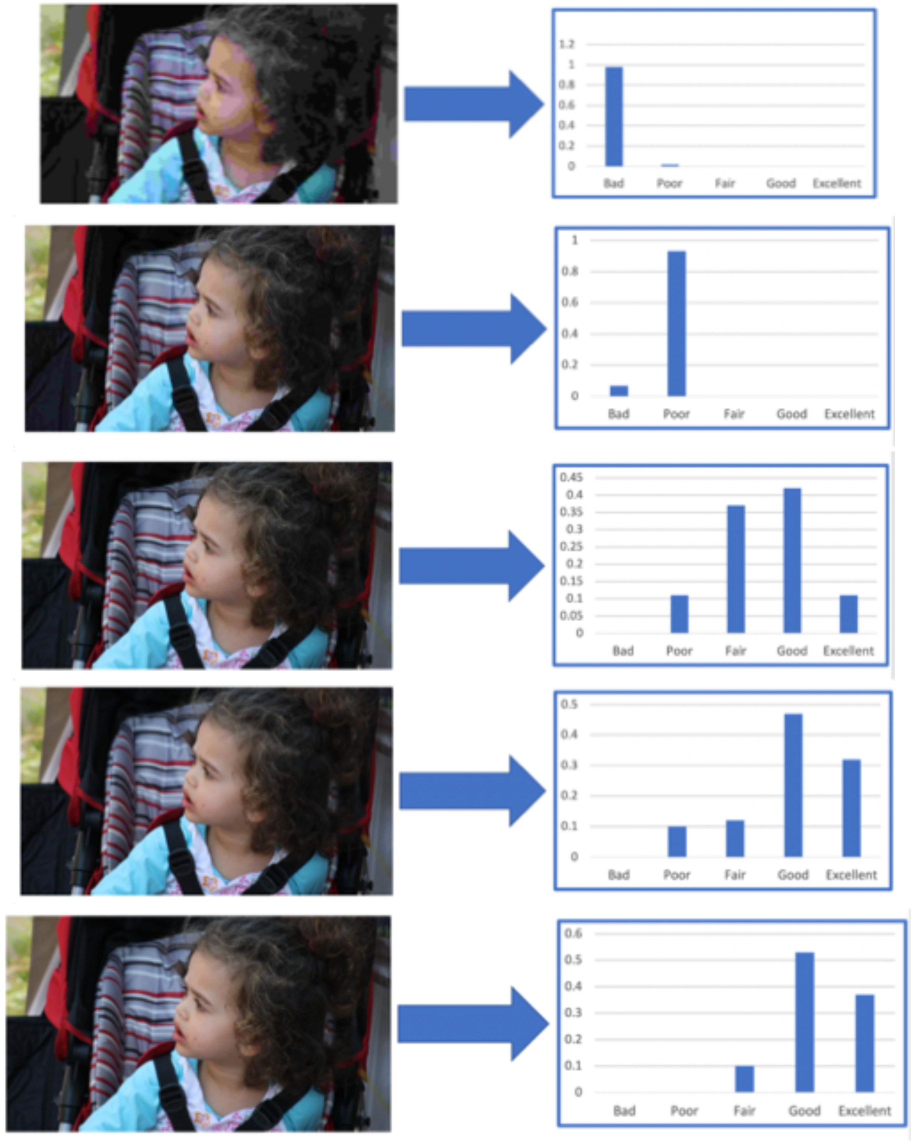


MICT



LIVE MD PH2

Some Results on Deep CNNs based AIOs



AIOs output as a function of the JPEG compression level

AIOs Robustness

- Shallow NNs based AIOs:
  - What if some noise is added to the input features?

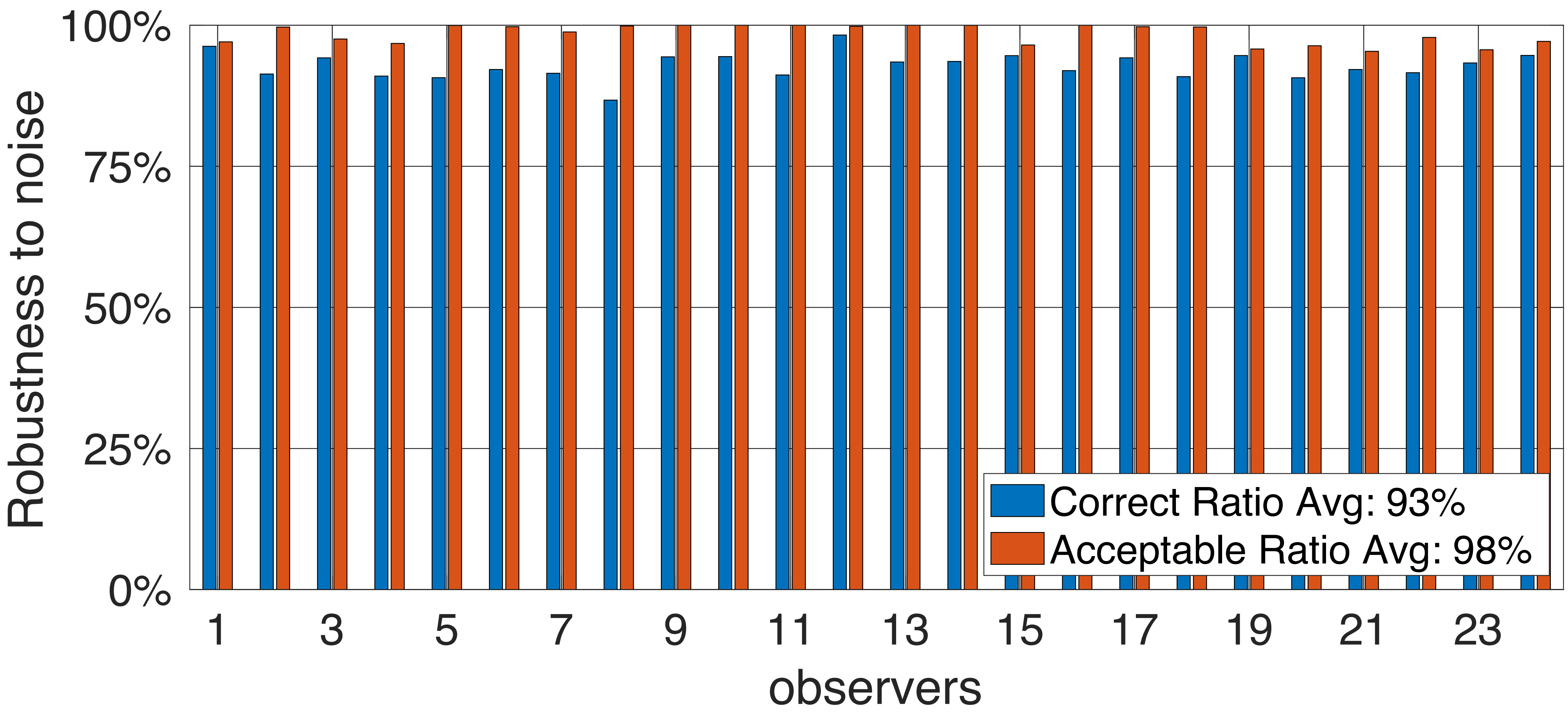
- Deep CNNs based AIOs:
  - What if the input image undergoes a transformation after which a human subject is “expected” not to change his opinion score?
  - To which extent the approach yielding the input image patch matters?

Shallow NNs based AIOs Robustness

- Adding a uniformly distributed noise to the input features:
  - noise range: between -1% and +1% of the feature “practical” variation range;
  - 10 000 noise realizations are considered for each AIO.

- Evaluation metrics:
  - **Correct ratio:** probability that the AIOs prediction will not change;
  - **Acceptable ratio:** probability that the AIOs prediction will move by at most one class on the ACR scale after adding the noise.

Shallow NNs based AIOs Robustness



Deep NNs based AIOs Robustness

- Input image modification:
  - From RGB to gray scale;
  - From RGB to RGB plus a not “perceptible” gaussian noise.

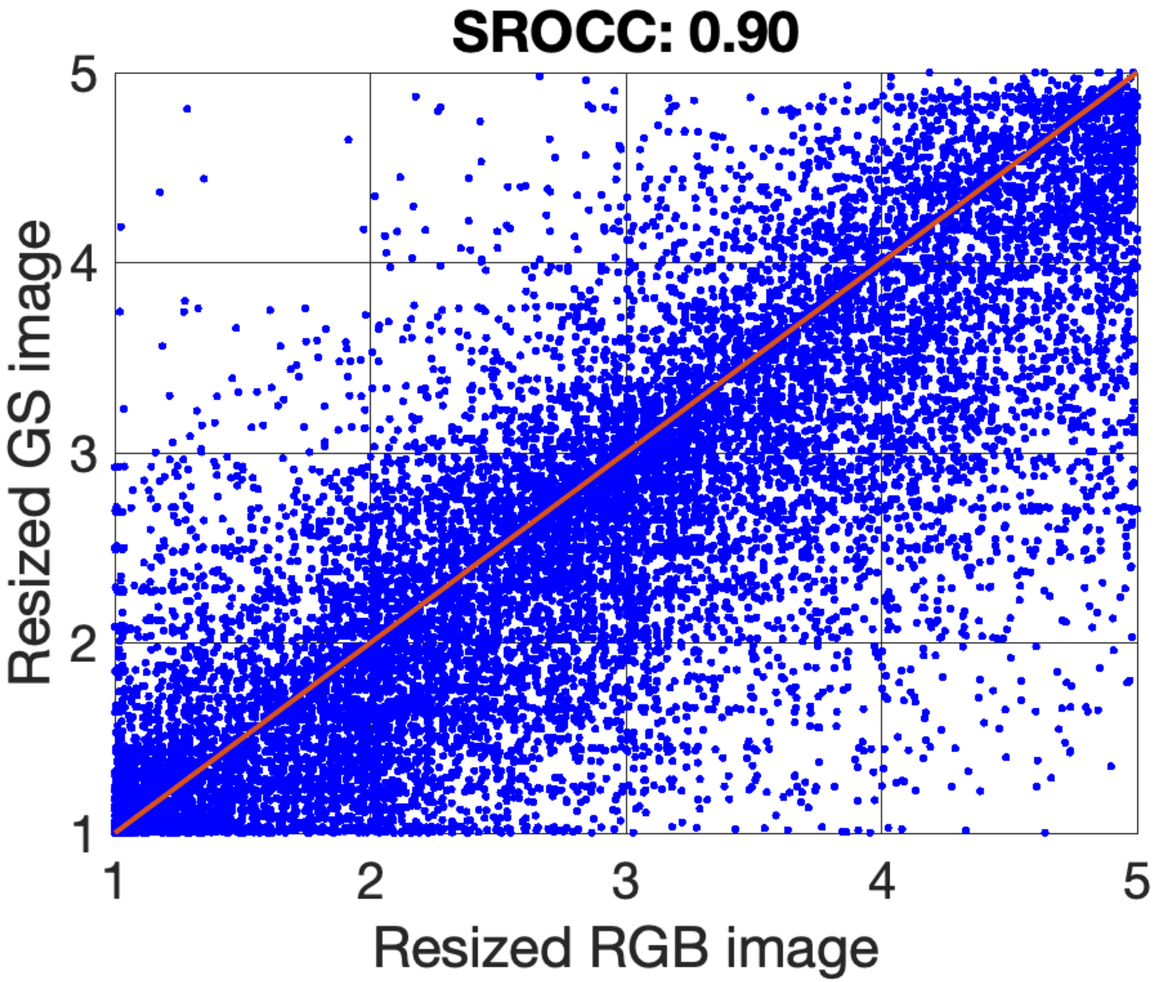
- Approach yielding the 224*224 input patch:
  - Resizing the input image (used during the training);
  - Center crop.
- Dataset:
  - 20 000 images: 4000 images*5 JPEG compression levels;
  - The 4000 original images were selected from the ImageNet competition dataset.

RGB vs Corresponding Gray Scale (GS) Resized Image

- Comparing the Mean Opinion Scores of the AIOs

RGB

GS

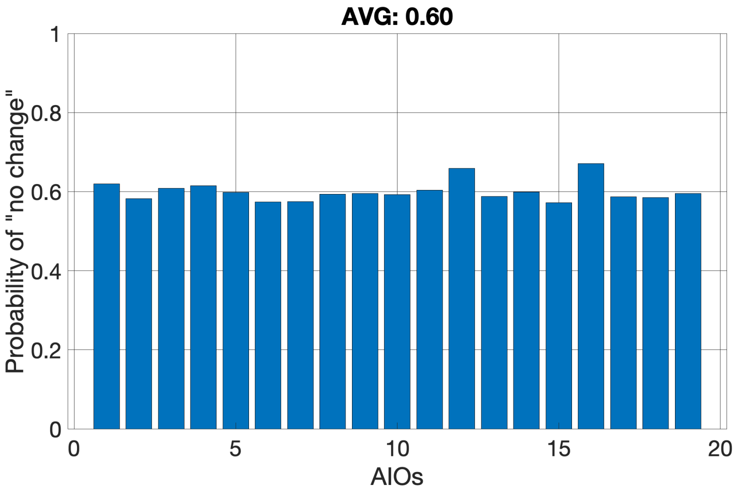


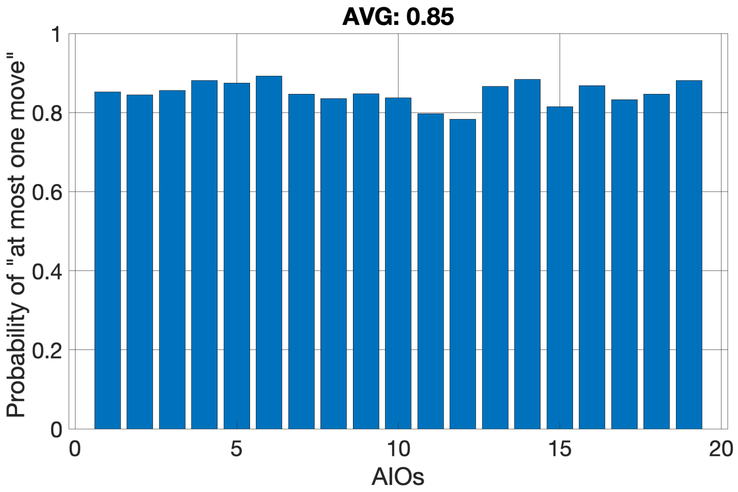
RGB vs Corresponding Gray Scale (GS) Resized Image

- What happens at the level of single subjects?

RGB

GS



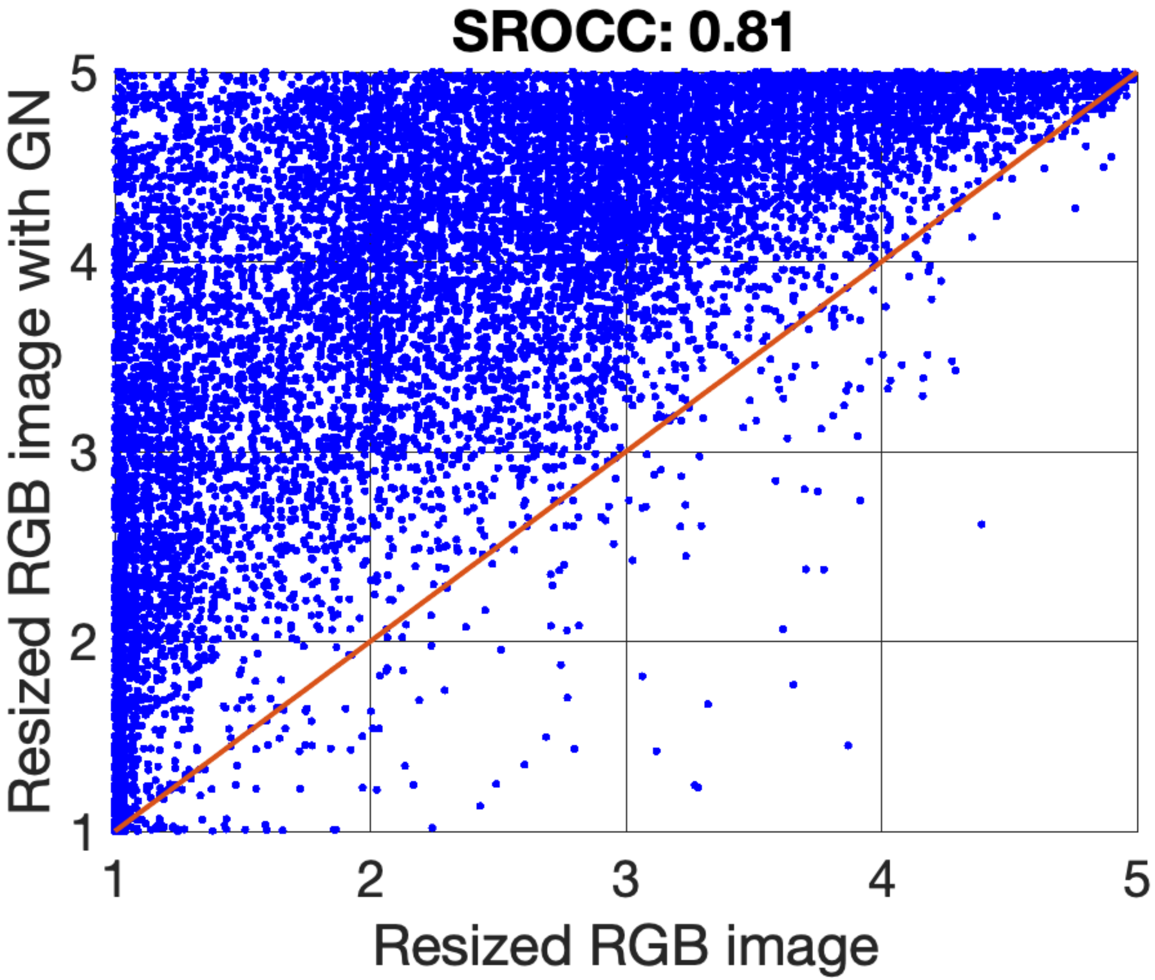


Adding a ”Not Perceptible” Gaussian Noise (GN)

- Comparing the Mean Opinion Scores of the AIOs

RGB

RGB+GN

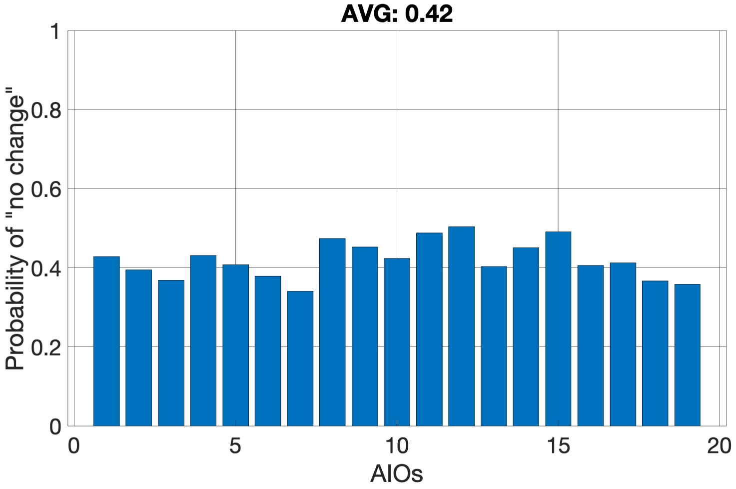


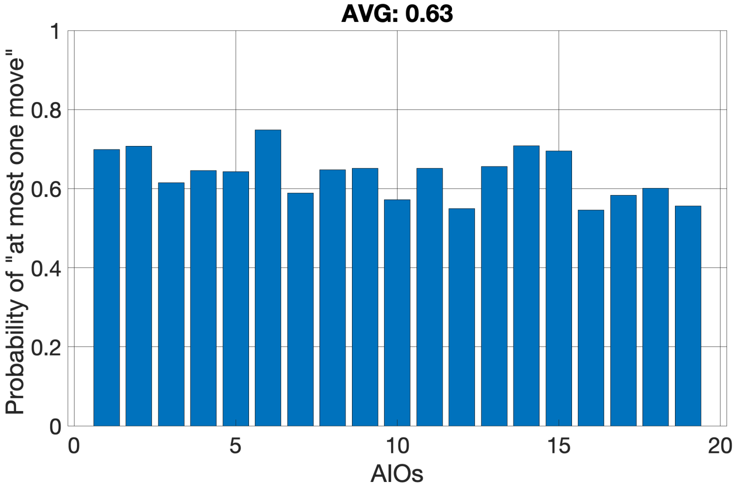
Adding a “Not Perceptible” Gaussian Noise (GN)

- What happens at the single subject level?

RGB

RGB+GN



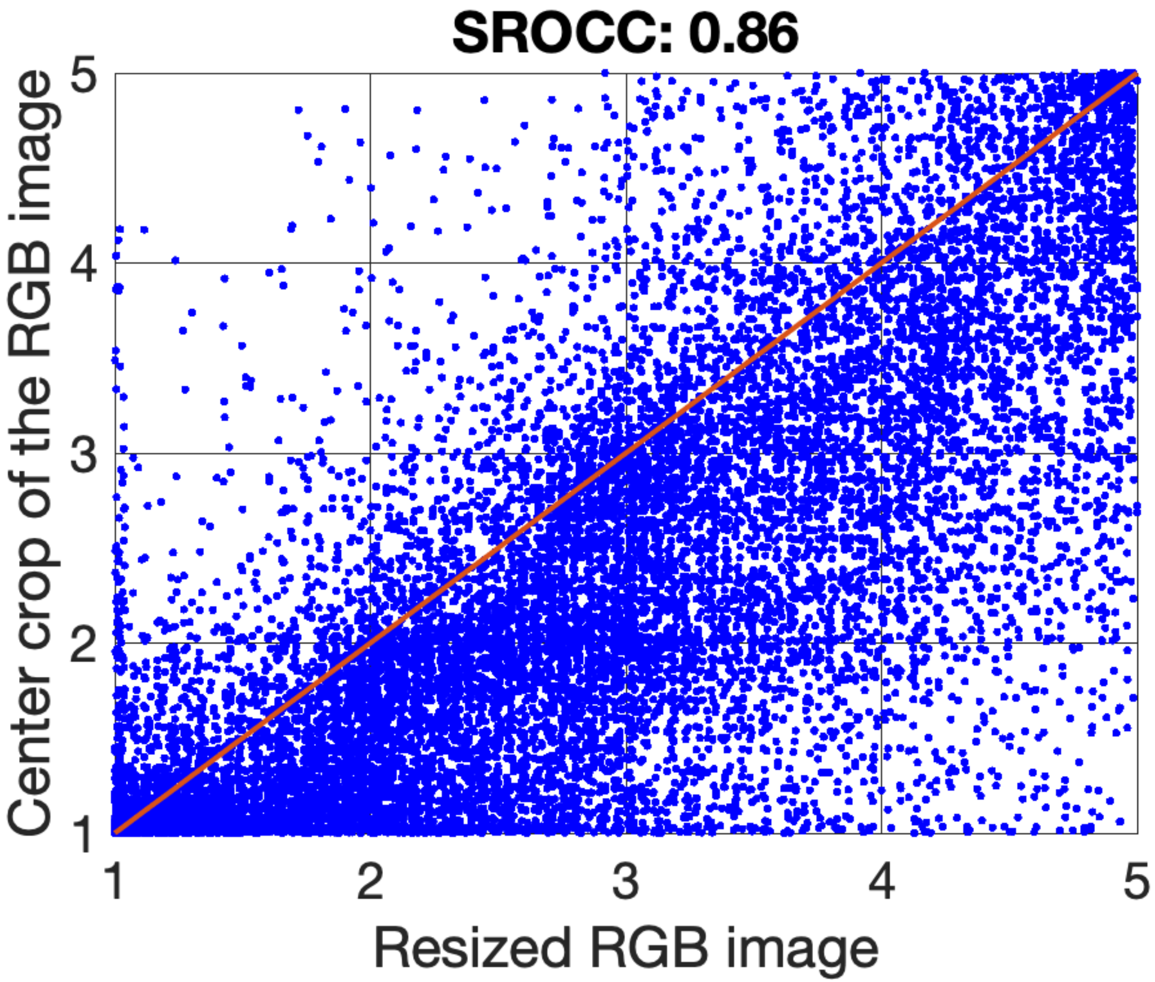


The Approach Yielding the Input Image Patch Matters

- Comparing the Mean Opinion Scores of the AIOs

Resized

Center Crop

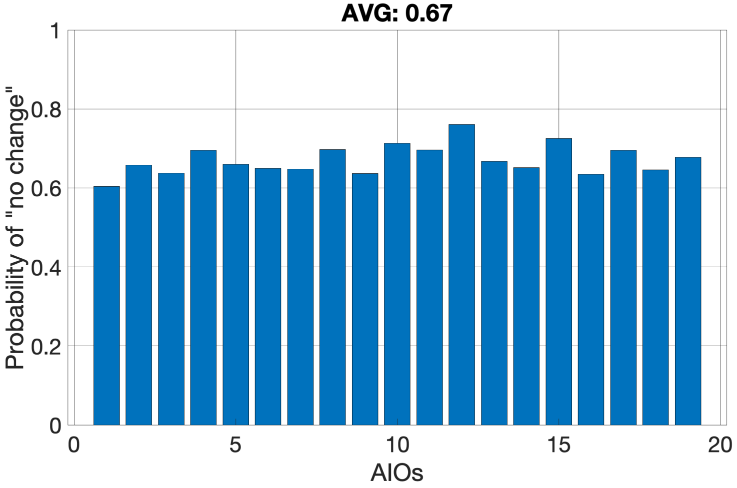


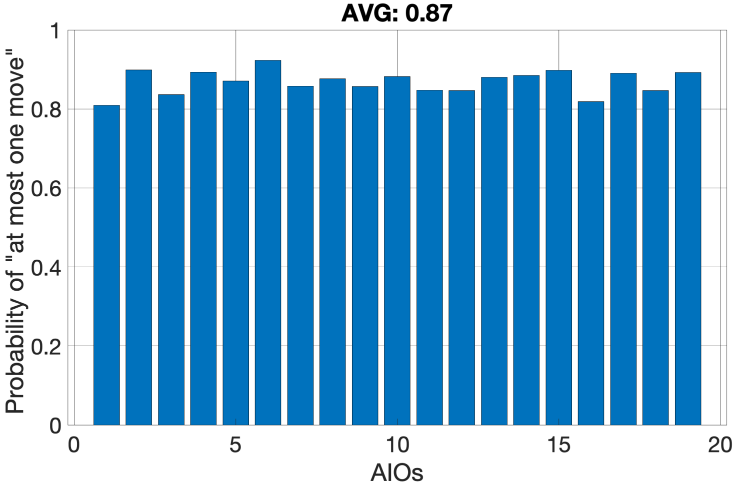
The Approach Yielding the Input Image Patch Matters

- What happens at the single subject level?

Resized

Center Crop





Next Steps

- Refining Deep CNNs-based AIOs:
  - Embedding data augmentation approaches to enhance the AIOs robustness
  - Should we crop rather than resizing?

- On the Robustness of Deep Learning-based Quality Measures:
  - Running the same experiments on other deep CNNs based metrics;
  - Is the issue task-related?