On the Training of AIOs for a Wider Range of Applications

LOHIC FOTIO TIOTSOP
JEG-HYBRID GROUP

VQEG MEETING: DEC 2022

AIOs

- AIO: deep neural network (DNN) trained to mimic the quality perception of an individual subject;

- The AIOs output a five-class probability distribution on the ACR scale;

- Aim and Scope: designing media processing systems that account for the characteristics of the targeted audience.

Training DNNs in Media Quality Assessment

- Current subjectively annotated datasets do not allow an effective training of DNNs for **MOS prediction** [1];

- Available datasets for the **training of AIOs** are even more limited in size;

- The classical **transfer learning concept is not** a satisfactory solution.

[1] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang and A. C. Bovik, "Deep Convolutional Neural Models for Picture-Quality Prediction: Challenges and Solutions to Data-Driven Image Quality Assessment," in *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 130-141, Nov. 2017, doi: 10.1109/MSP.2017.2736018.

Two-steps Transfer Learning

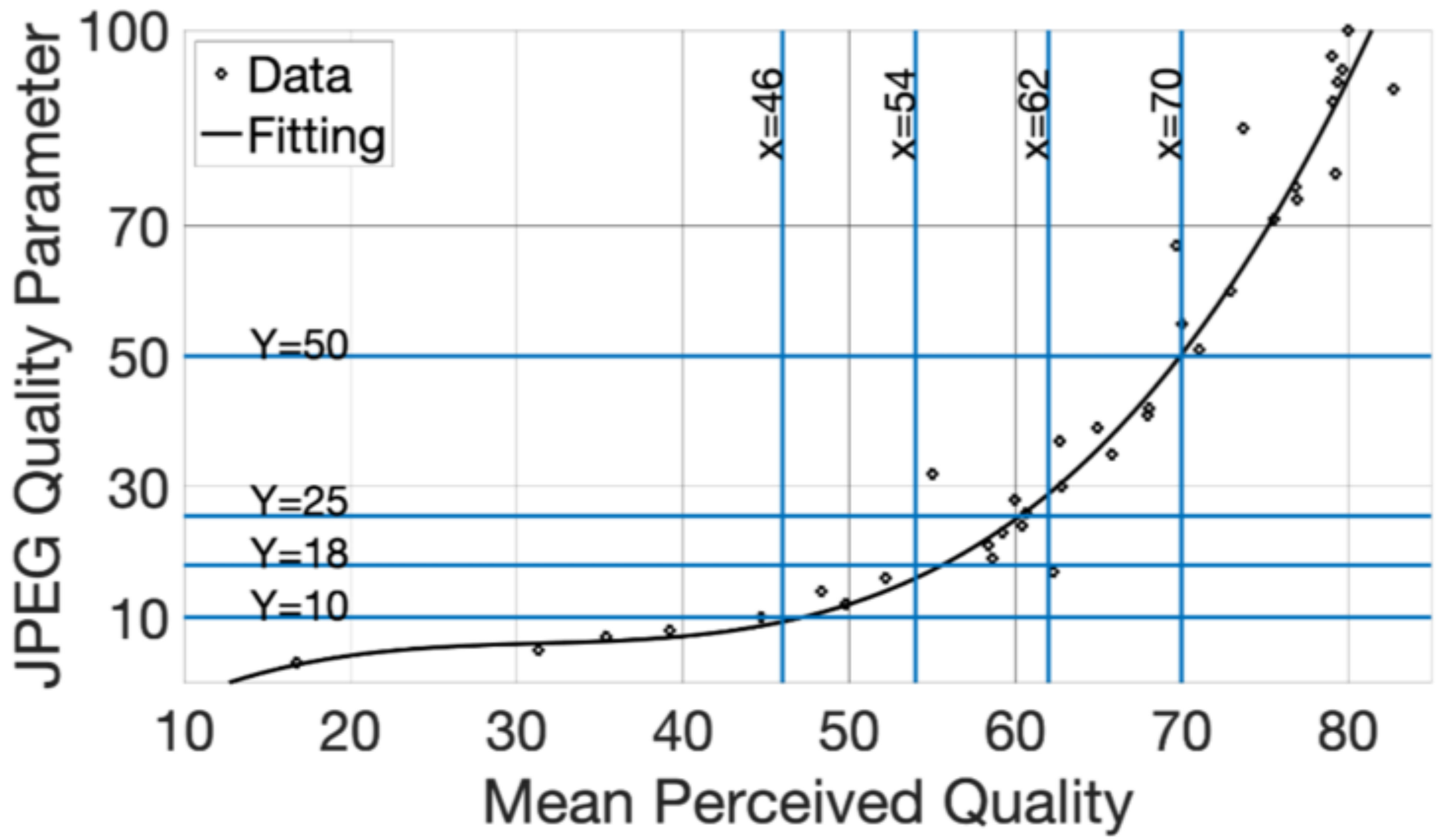
- An idea already explored in media quality assessment by some authors;

- **First step:** create a large scale synthetically annotated dataset, and use it to train a DNN to extract generic perceptual features;
- **Second step**: Refine the weights of the trained DNN by performing an additional training on a subjectively annotated dataset.

- How to obtain reasonable synthetic labels?

- Previous authors used full reference metrics; we proceed differently.

JPEG Compression Only Case (JPEGResNet50)



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

- The analysis was made on the first release of the LIVE-IQA dataset.

Effectiveness of the Two-Steps Transfer Learning Approach

Range of correlation coefficients between the ratings of the trained AIOs and those of actual observers:

Datasets	TLR	FW-TLR	Our approach
LIVE-MD-ph1 (T)	[0.09, 0.70]	[0.00, 0.69]	**[0.34, 0.84]**
LIVE-IQA-r1-ph1	[-0.19, 0.27]	[-0.28 0.32]	**[0.42, 0.87]**
LIVE-IQA-r1-ph2	[-0.37 0.60]	[-0.33, 0.46]	**[0.79, 0.92]**
LIVE-MD-ph2	[0.04, 0.67]	[-0.05, 0.63]	**[0.21, 0.64]**
MICT	[-0.23 0.36]	[-0.28, 0.50]	**[0.30 0.76]**

- **TLR**: single Transfer Learning using the ResNet50
- **FW-TLR**: Freeze 50% of the Weights of the Resnet50 and perform single Transfer learning
- **Our approach**: 2-step transfer learning

- However, by considering JPEG compression only, the trained AIOs are of limited use for real applications.

Proposed Generic Algorithm to Compute Parameter Intervals

- Generic algorithm suitable for any distortion type controlled by a single parameter

Algorithm 2 Proposed Algorithm

Inptut: $\{d_i\}, \quad \{MOS_i\}$

while $d \in \{d_i\}$ **do**

$\quad I_d \leftarrow \{i \parallel \ d_i = d\}$

$\quad MOS_{avg}(d) \leftarrow avg\{MOS_i \ \parallel \ i \in I_d\}$

end while

$g(.) \leftarrow \arg\min_{g(.)} \left(\sum_i (g(d_i) - MOS_{avg}(d_i))^2\right)$

$q_{avg}^{max} \leftarrow Percentile_{0.975}\{MOS_{avg}(d_i)\}$

$q_{avg}^{min} \leftarrow Percentile_{0.025}\{MOS_{avg}(d_i)\}$

$[I_q^1, I_q^2, \ldots, I_q^5] = Split_5(q_{avg}^{min}, q_{avg}^{max})$

$I_d^k = g^{-1}(I_q^k) \quad k = 1, 2, \ldots, 5$

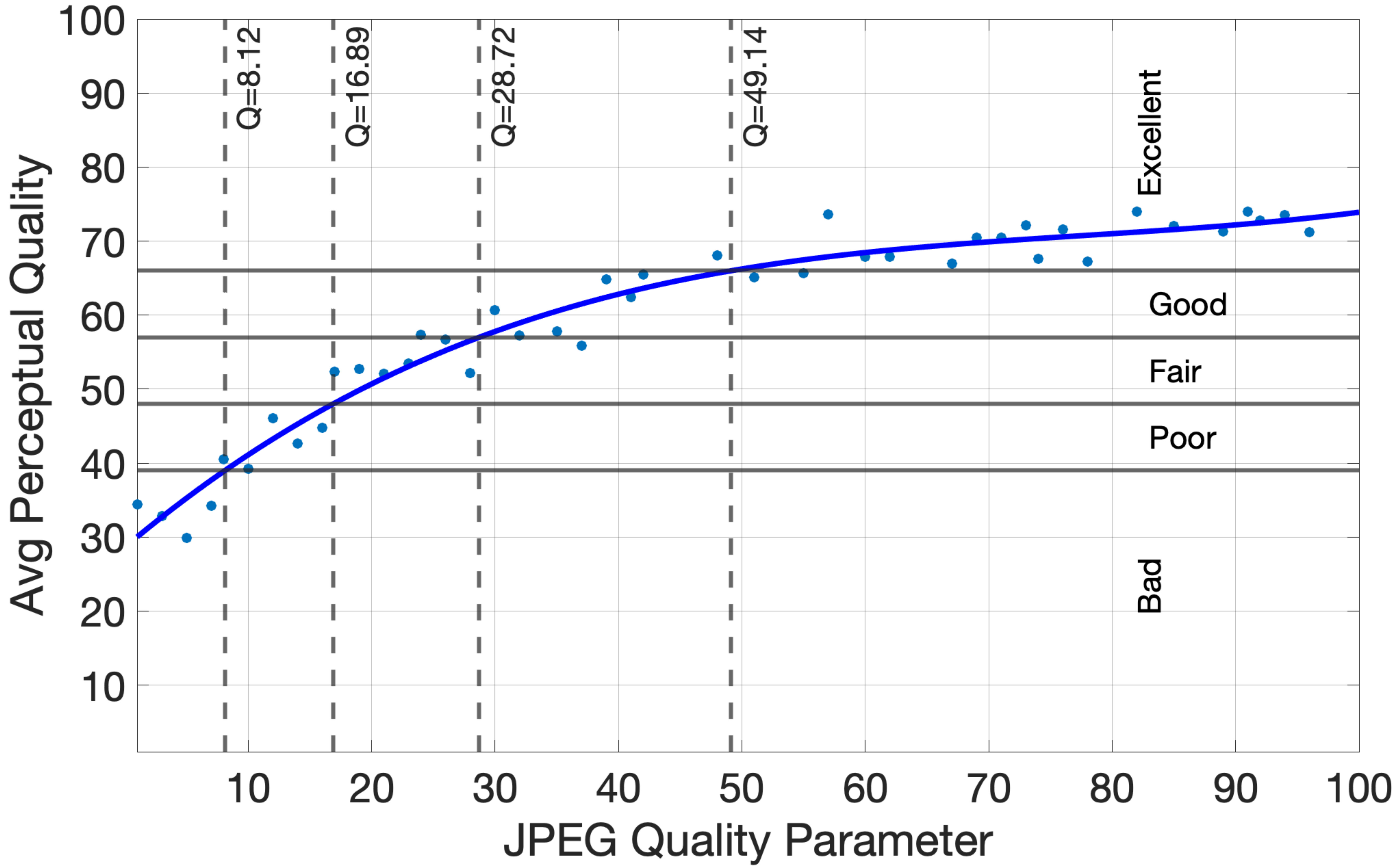
Ouptut: $[I_d^1, I_d^2, \ldots, I_d^5]$

Using the Algorithm vs Full Reference VQMs Labelling

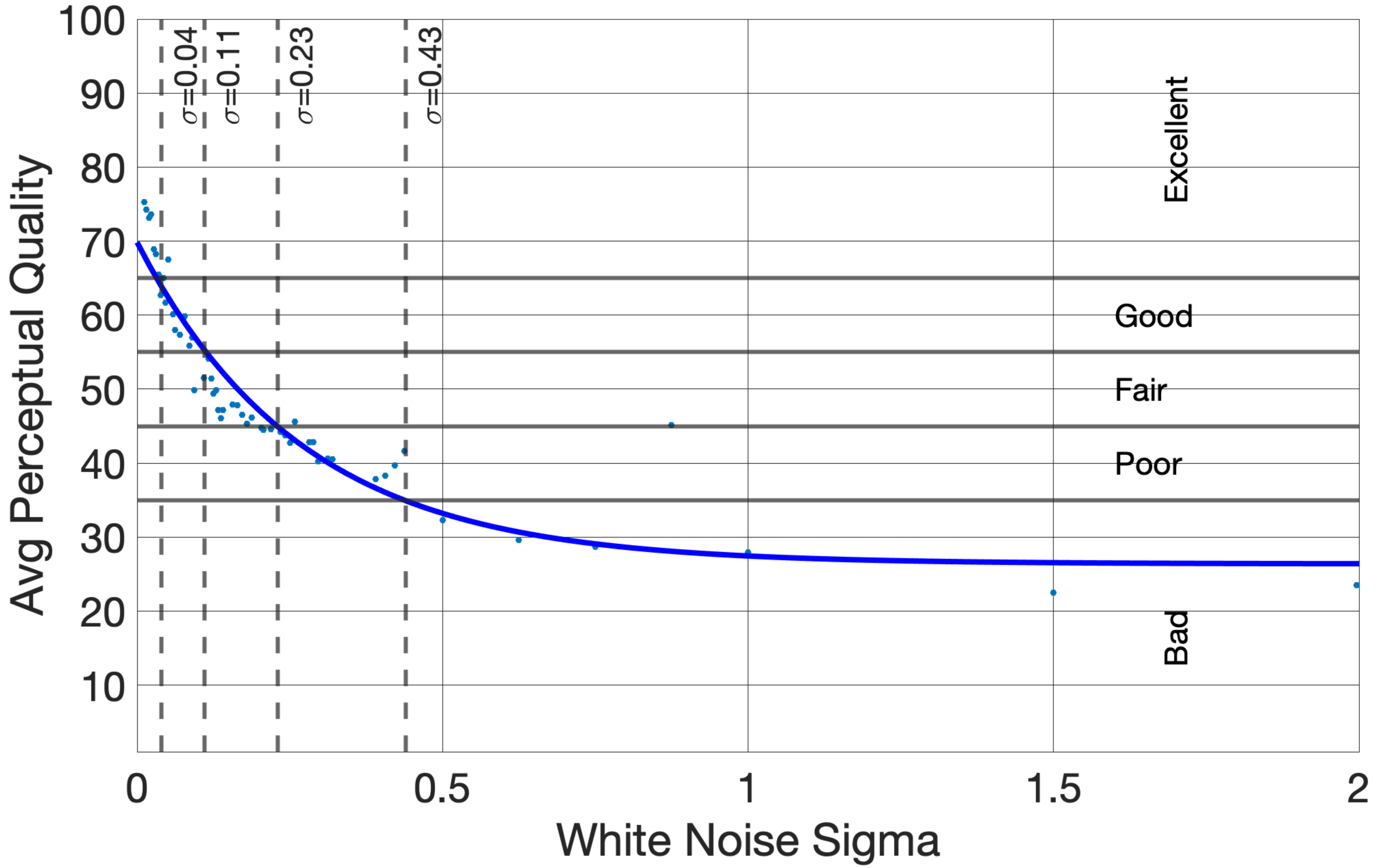
- **Advantage:** Not upper-bounding the performance of the network trained at the first stage by that of a specific full reference metric;

- **Disadvantage**: Full reference metrics probably better account for the characteristics of the stimuli.

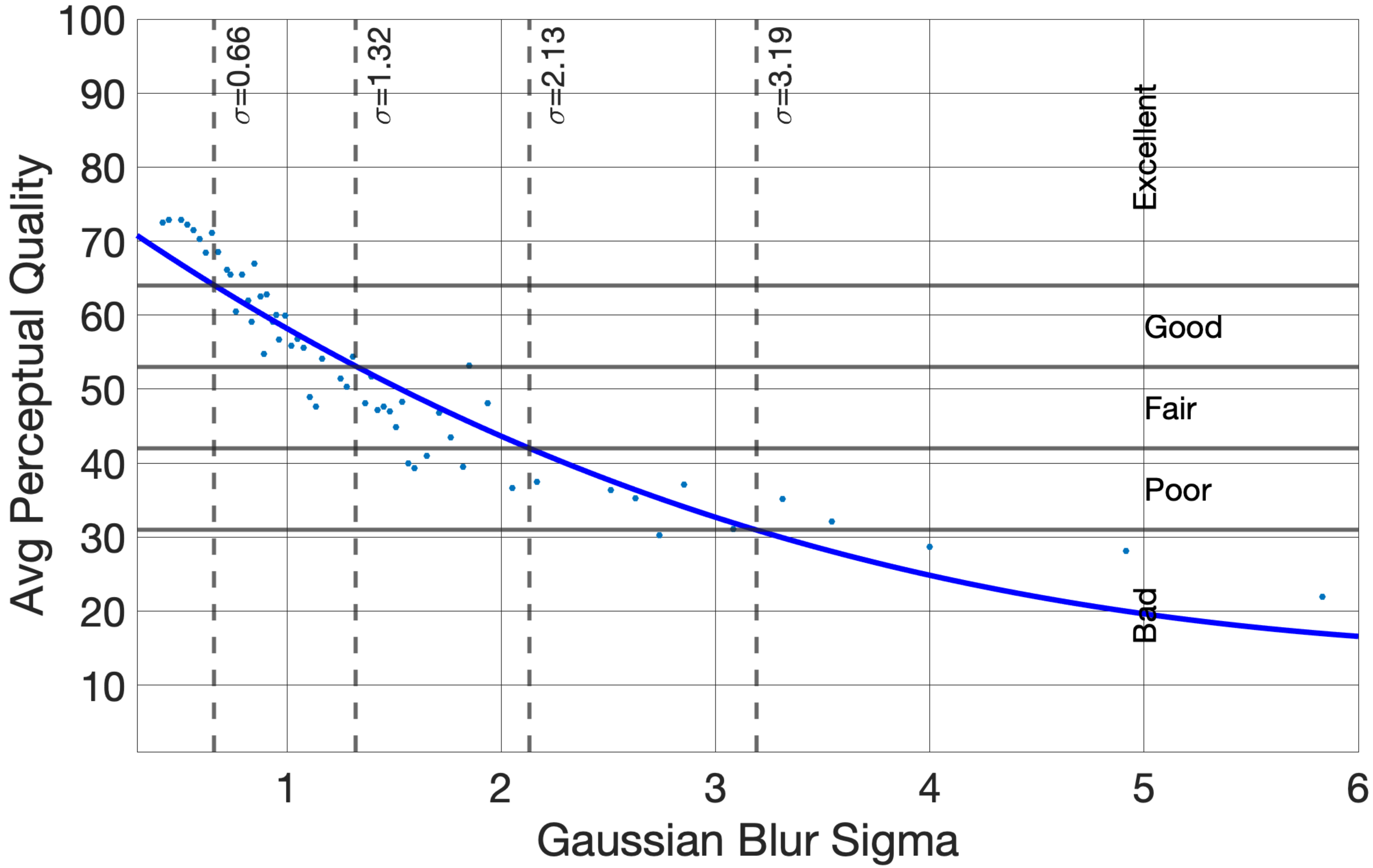
JPEG to the Average Perceived Quality



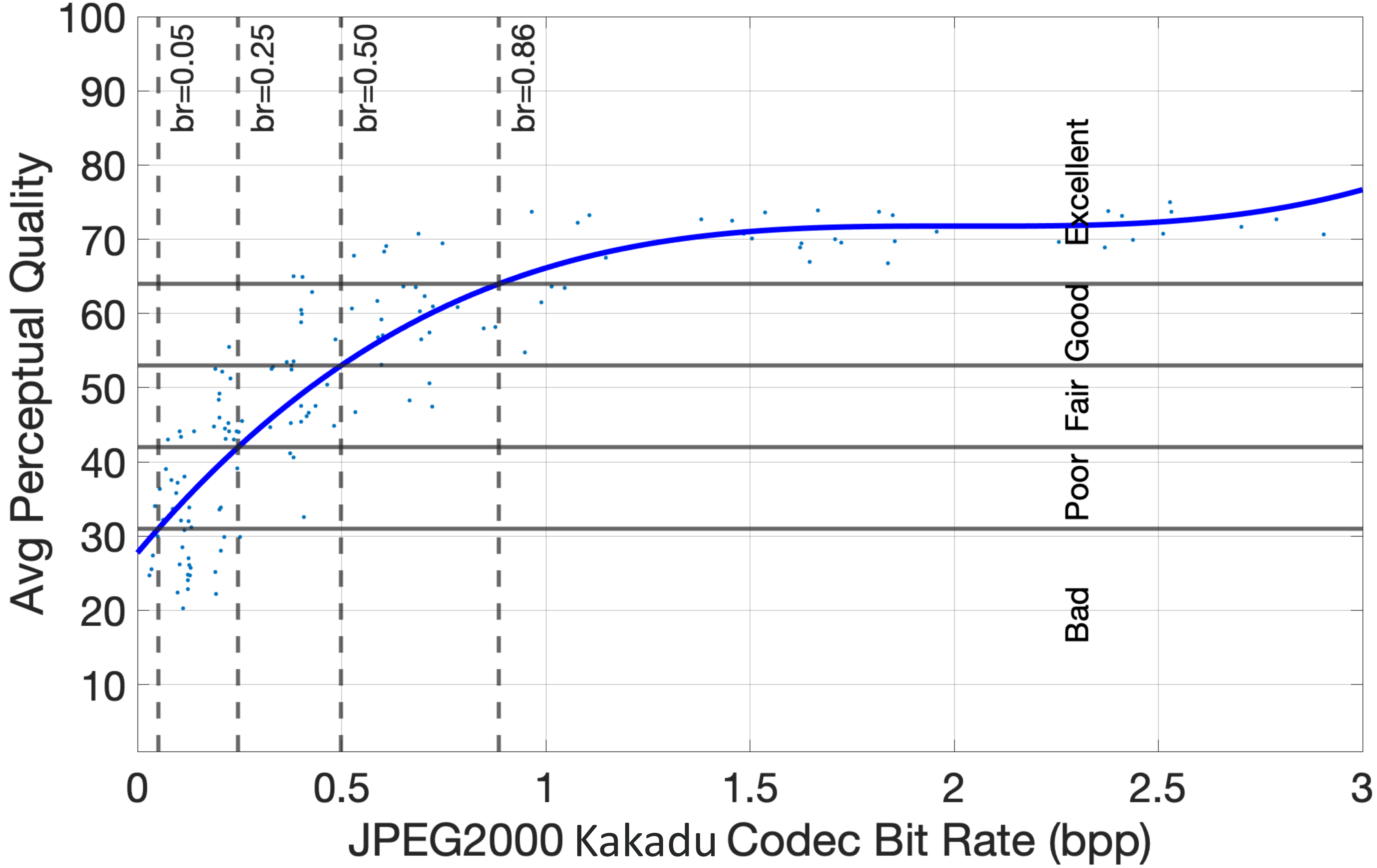
White Noise to the Average Perceived Quality



Gaussian Blur to the Average Perceived Quality



JPEG2K to the Average Perceived Quality



The Created Synthetically Annotated Datasets

- We started from 100 000 images from the ImageNet dataset;
- 75 000 were dedicated to the training set, 12 500 to the validation set and 12 500 to the test set;
- We obtained 75 000*5*4 = 1 500 000 training samples, 250 000 for the validation and 250 000 for the test;
- **Note: there is no intersection between the three sets.**

Training setup of the MDResNet50

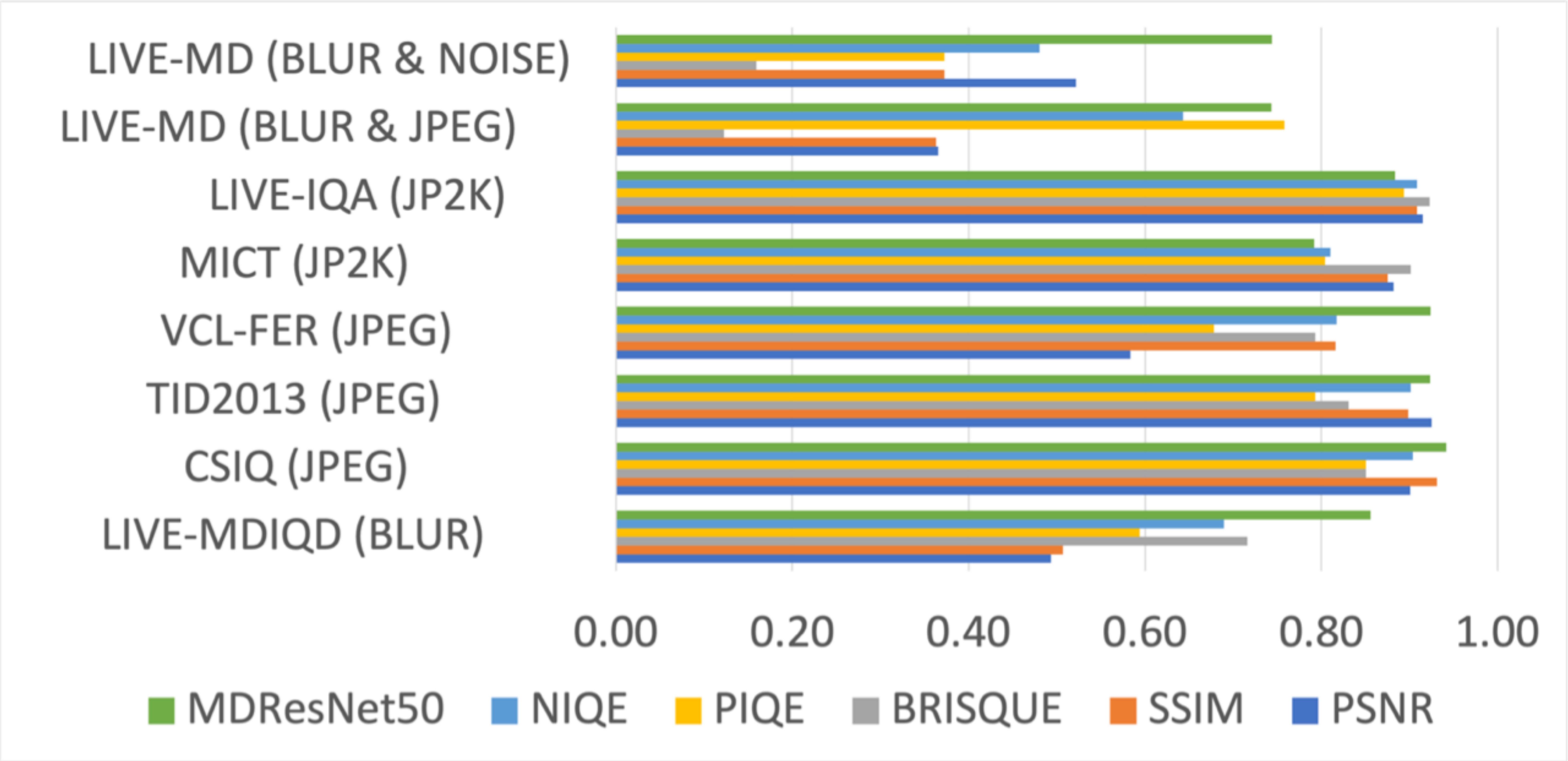
- Done with 1.5 M training samples and 250 000 samples in the validation set;

- **Available computational resources:**
  - **GPU NVIDIA GeForce RTX 3090 24 GB ram**
  - **CPU Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz with 64 GB ram**

- The training lasted 7 days;

- Training Progress comparable to that of the JPEGResNet50

MDResNet50 vs Existing Metrics



SROCC between the MOS and the VQMs

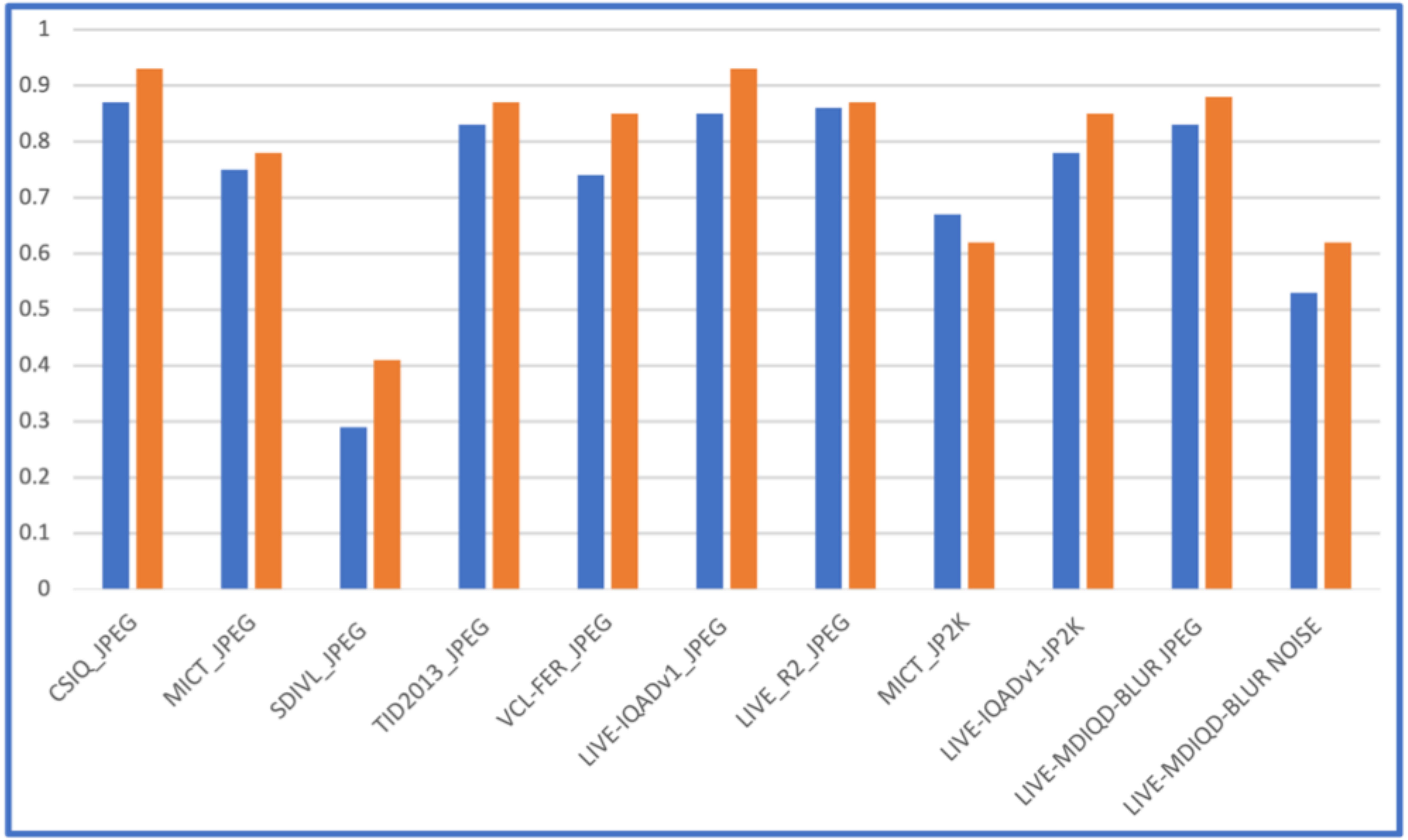
The MD-AIOs

- Previous result suggests that the MDResNet50 is a suitable starting point for transfer learning;
- An additional learning step was performed to get the AIOs;
- The training set (LIVE-MD) contains 15 ref + 225 distorted images (blur only, JPEG only and blur + JPEG);
- **19 subjects participated in the test, yielding 19 MD-AIOs.**

MD-AIOs vs Previous AIOs

MD-AIOs

AIOs



SROCC between the actual MOS and the mean of the ratings of the AIOs and the MD-AIOs

On The MD-AIOs Sensitivity to Input Modification

- Adding a few noise to the input image;
- Converting the RGB input image into a gray scale one.

Adding a few Gaussian Noise (GN)

RGB

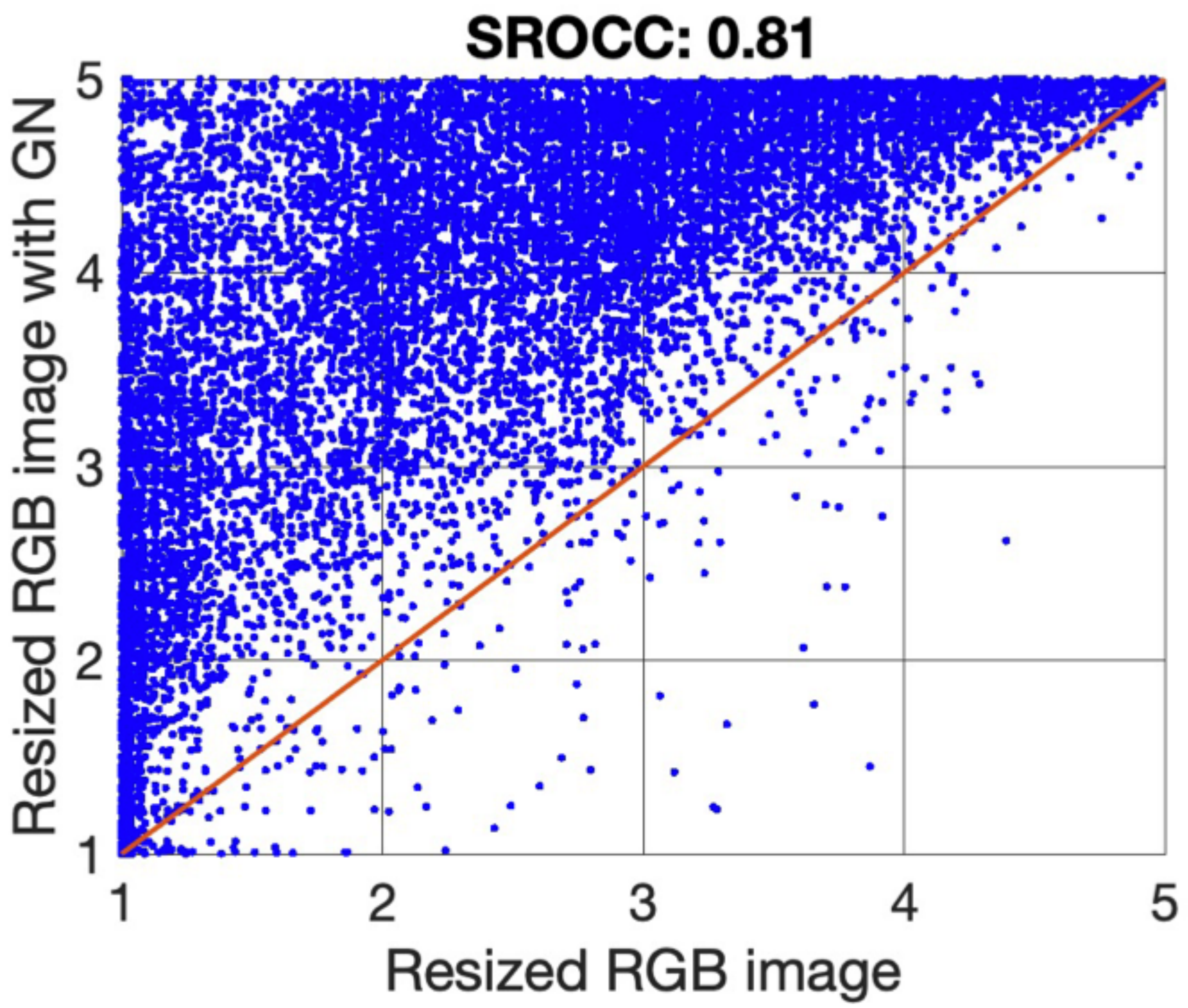
RGB+GN

From RGB to Gray Scale (GS)

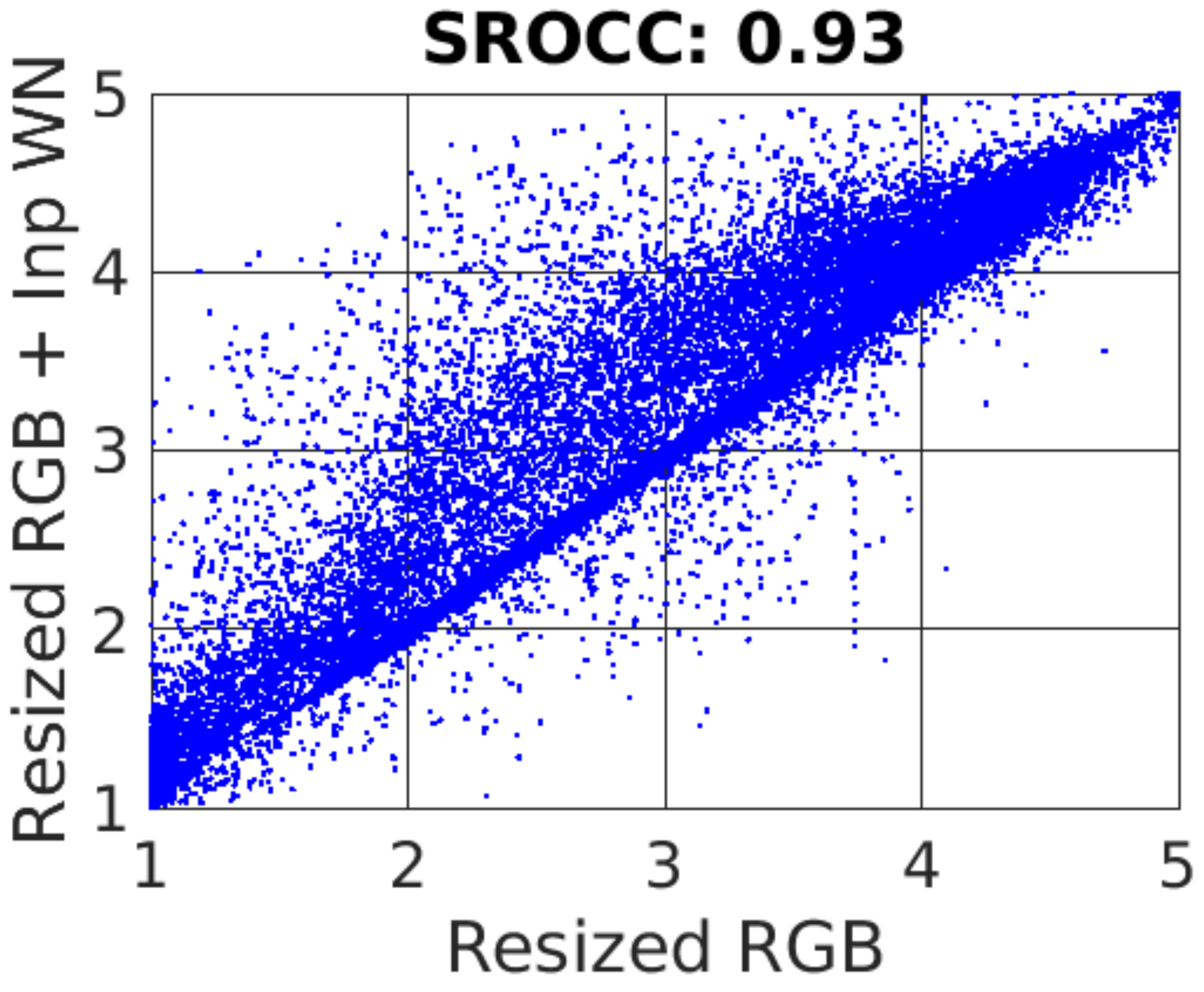
RGB

GS

Adding "Not Perceptible" Gaussian Noise

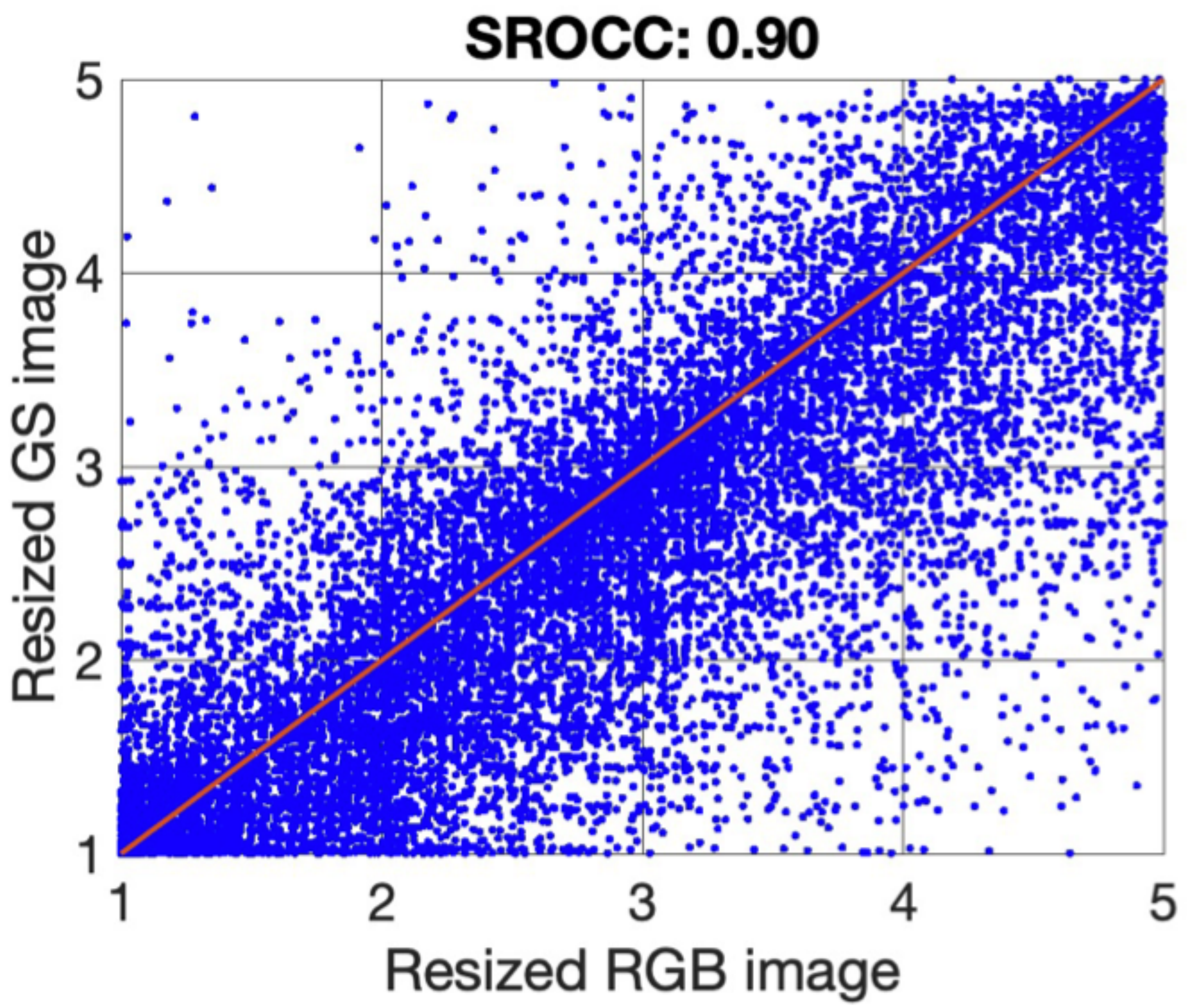


AIOs

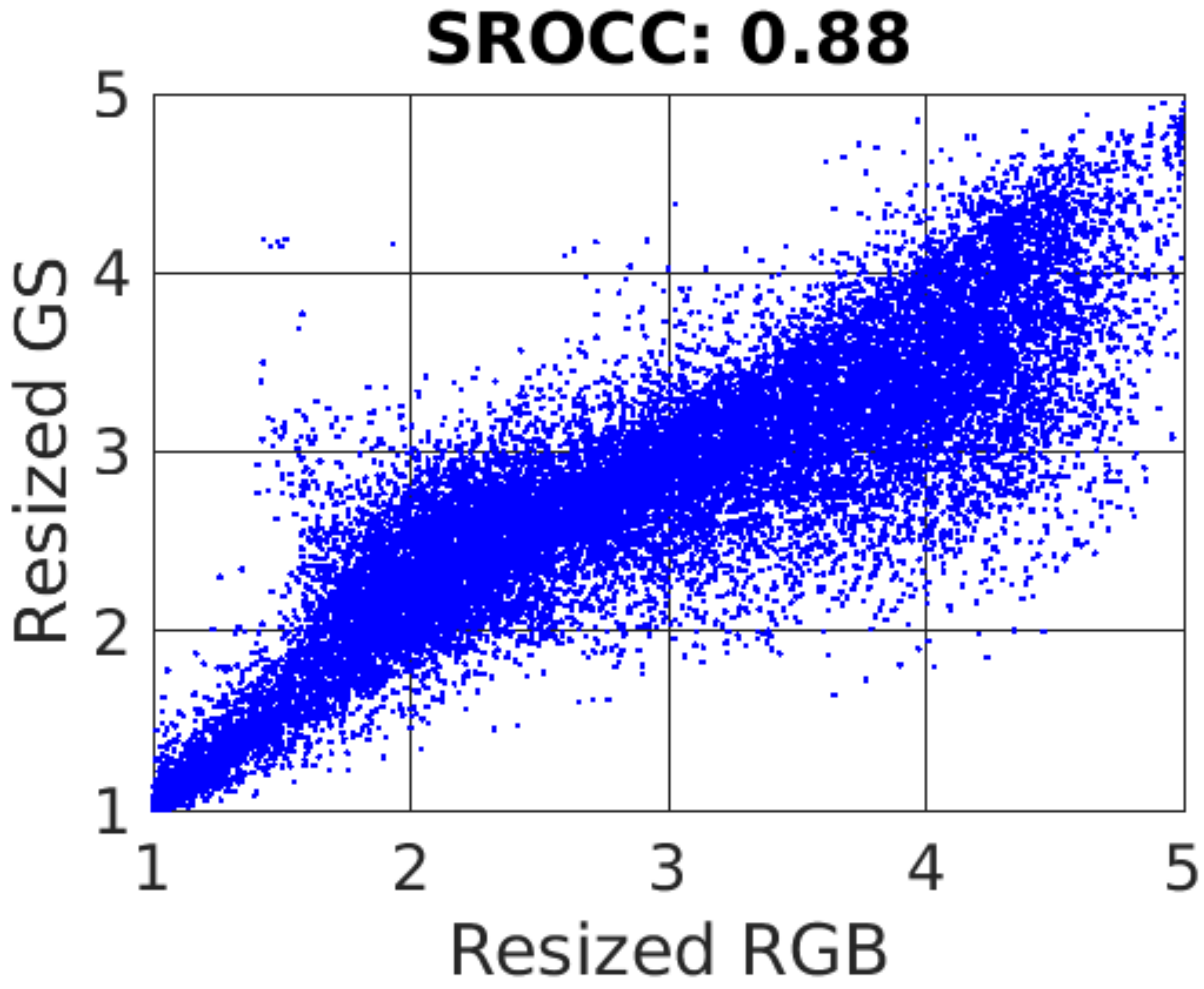


MD-AIOs

From RGB to Gray Scale



AIOs



MD-AIOs

Mimicking the Subjects' Inconsistency

- We defined the variance of the predicted discrete probability distribution as a measure of inconsistency [2];
  - [2] L. Fotio et al., "Mimicking Individual Media Quality Perception with Neural Network Based Artificial Observers", ACM TOMM 2022

- We only showed that such a measure satisfies some properties expected on the subject inconsistency;

- Here, we exploit the scoring model [3]: $r = q + b + N(0, \sigma)$ (the same used in the Sureal software) for more investigation
  - [3] L. Janowski, M. Pinson, "The Accuracy of Subjects in a Quality Experiment: A Theoretical Subject Model", IEEE TMM 2015

Experiment Setup

- The experiment is done with the release 2 of the LIVE-IQA;
- A total of 808 stimuli were considered;
- Each stimulus i with the related subjective quality q_i;
- We chose some ground truth bias and inconsistency values

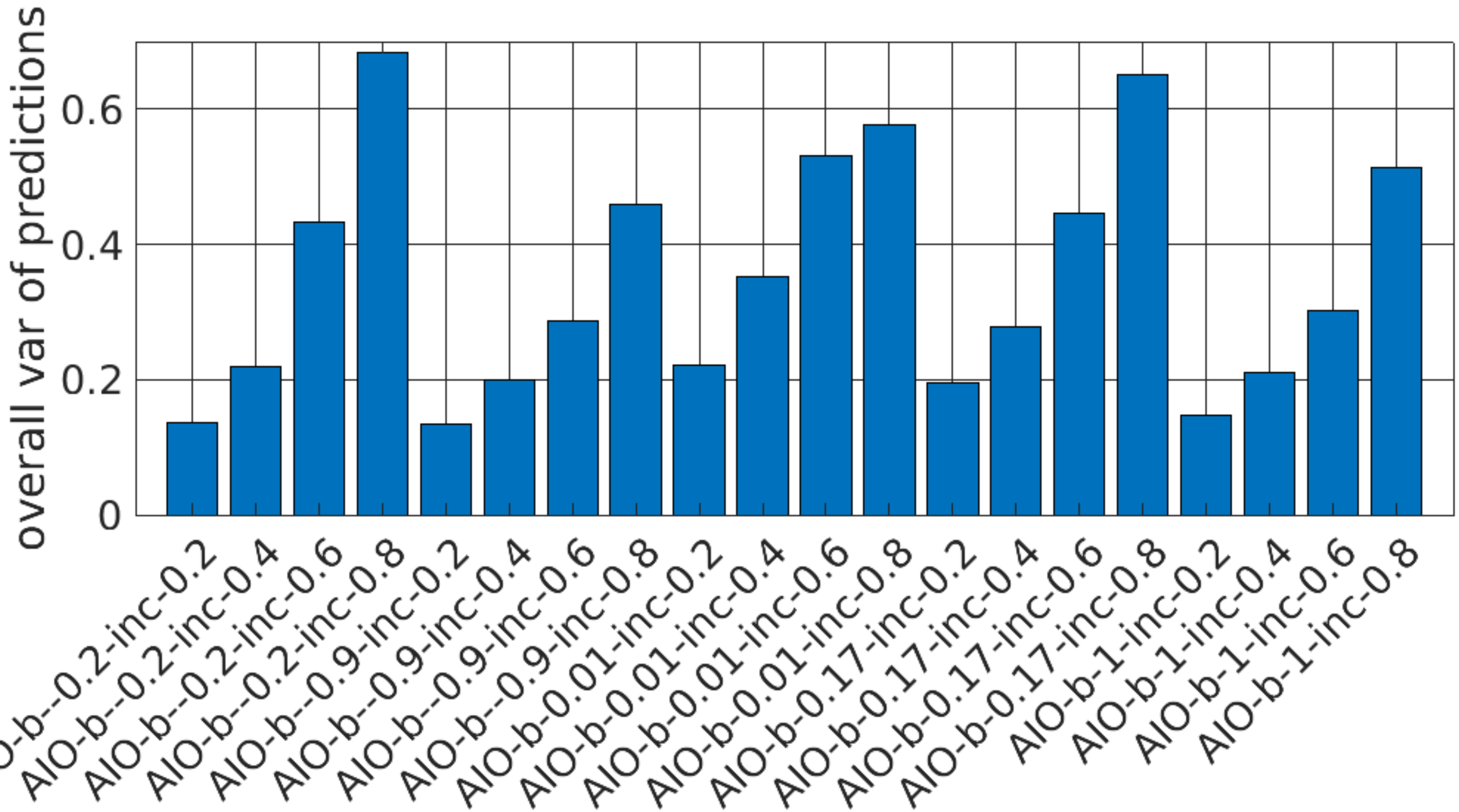
$$\mathbf{b} = [-0.9 \ \ -0.2 \ \ 0.01 \ \ 0.17 \ \ 1] \quad \boldsymbol{\sigma} = [0.2 \ \ 0.4 \ \ 0.6 \ \ 0.8];$$

- We Simulated the ratings of 20 subjects on each stimulus i as:

$$r_i = q_i + b + N(0, \sigma) \ \ b \in \mathbf{b} \text{ and } \sigma \in \boldsymbol{\sigma}$$

- We then trained 20 AIOs to mimic these simulated subjects.

Ground Truth Inc vs Avg variance of the prediction



Ground Truth Bias vs Avg Choice Probabilities