



**Politecnico
di Torino**

Guidelines to Train, Validate and Publish DNN- based Objective Measures (Ongoing Work)

**LOHIC FOTIO TIOTSOP
JEG-HYBRID GROUP**

VQEG MEETING: DEC 2022

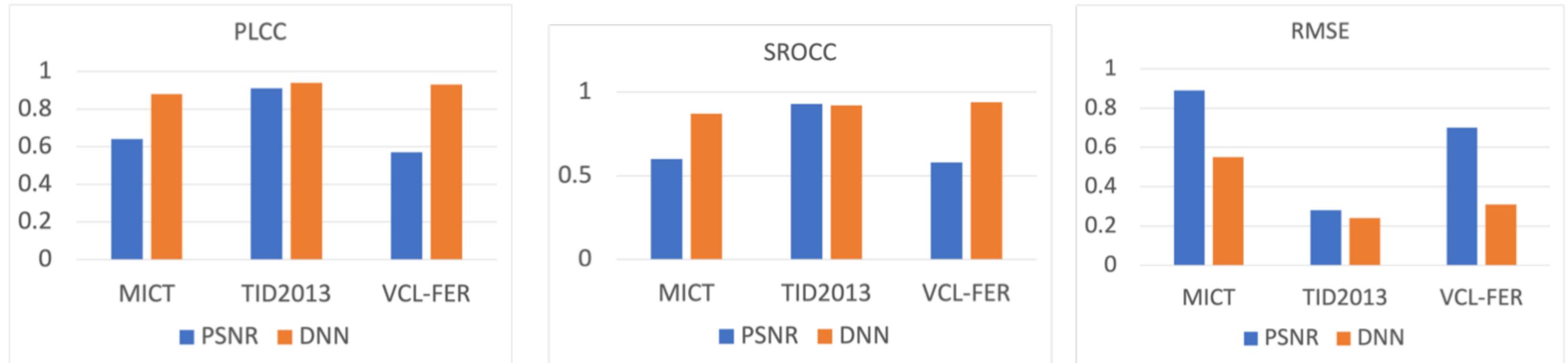
Aim

- **Short/mid term goal:** writing a journal paper with guidelines about how to train, validate and publish DNN-based objective measures
- In particular, proposing:
 - A list of typical “mistakes” to avoid
 - A few guidelines to train the model
 - A more effective objective approach to benchmark the trained model
 - A universal database to test metrics and record their performance
 - **A template to report/publish the results**
- **Long term goal:** evolve it into an international recommendation

Motivations

- Enhancing research reproducibility
 - A template to publish the results can strongly help in this aspect
- Towards a fair review process
 - The template will make it easier to compare different manuscripts
- An effective comparison of metrics
 - The correlation coefficients and RMSE are not enough
- Avoiding over-generalization and/or misleading claims
 - e.g., “our metrics is suitable for any context, since it is trained with images impaired by several distortions”
 - e.g., “Our metrics takes as input an NxM image patch” while it is not clear how create the patch from other image sizes (resize, crop, etc.)

PLCC, SROCC and RMSE are not enough



- Is the DNN-based model really a better metric than the PSNR?

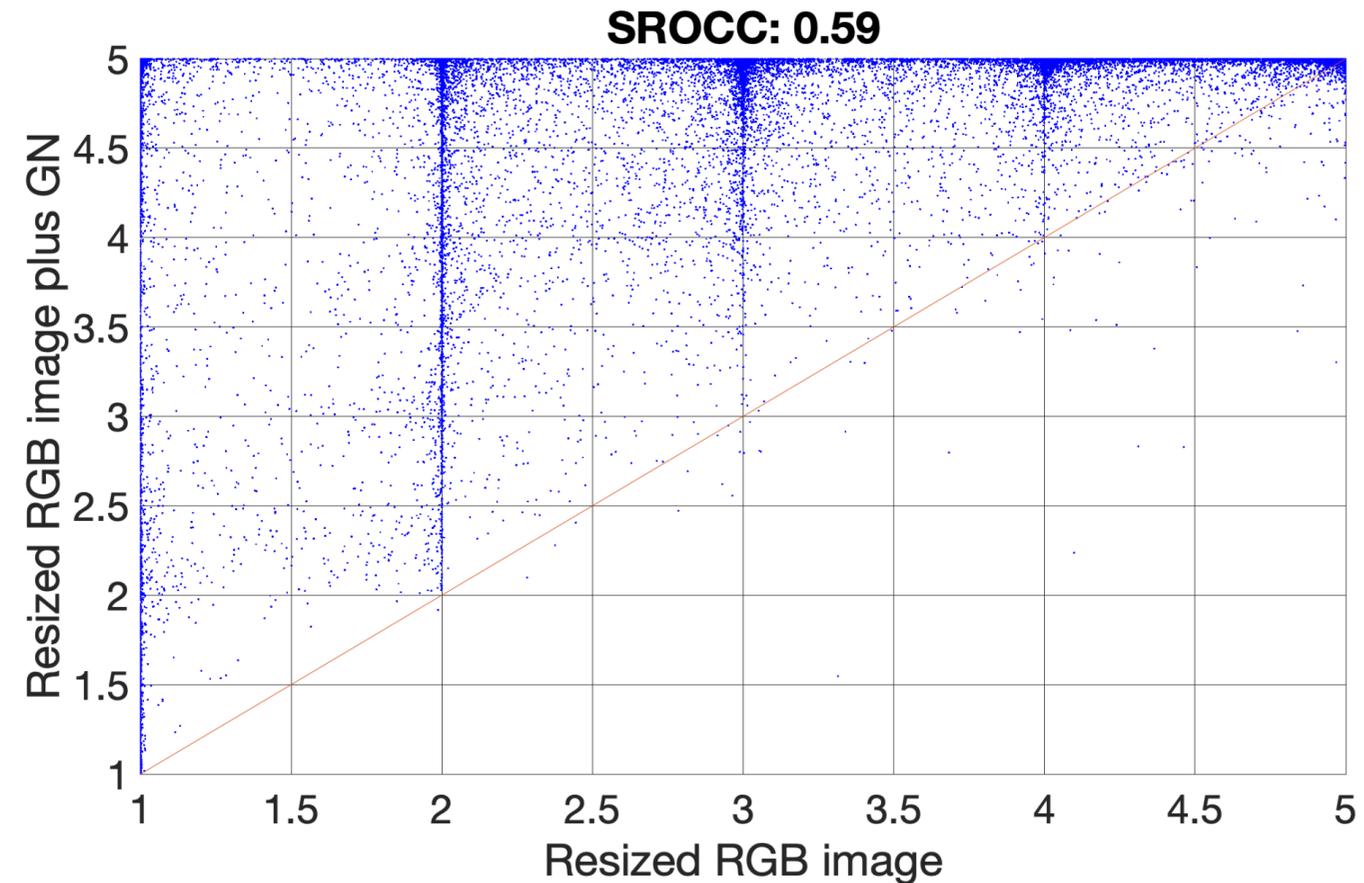
PLCC, SROCC and RMSE are not enough



RGB

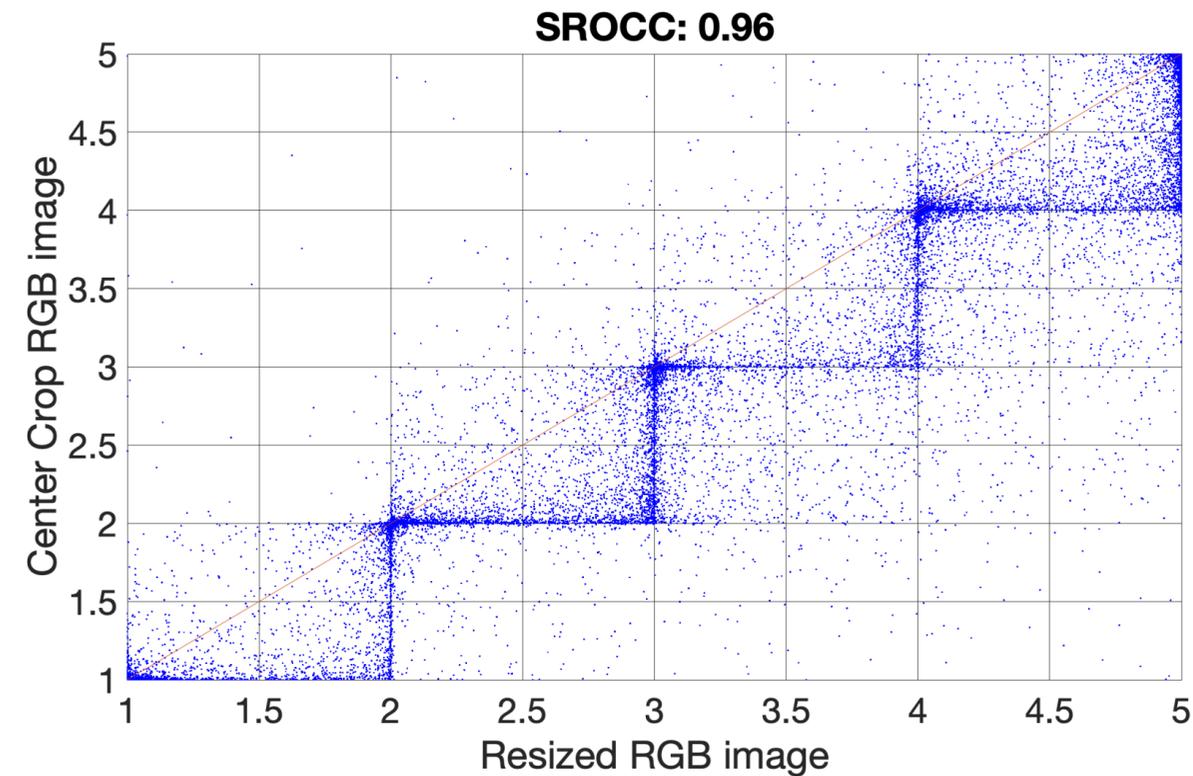


RGB+GN



- Unfortunately, the predictions of the DNN-based model may drastically change even when a very low amount of Gaussian Noise (GN) ($\sigma = 0.001$) is added to the input image

The Way the Image Patch is Obtained Matters



DNN-based Metric 1



DNN-based Metric 2

- Center cropping a small image or resizing it yields, in some cases, patches with different predicted perceptual quality

Issues with Over-Generalization: Example

DATASET	DISTOR-TION	BRISQUE	PIQUE	NIQE	DNN	PSNR	SSIM
CSIQ	JPEG	0.85	0.85	0.90	0.71	0.90	0.93
MICT	JPEG	0.92	0.69	0.81	0.35	0.60	0.66
SDIVL	JPEG	0.54	0.59	0.54	0.44	0.76	0.82
TID2013	JPEG	0.83	0.79	0.90	0.73	0.93	0.90
VCL-FER	JPEG	0.79	0.68	0.82	0.52	0.58	0.82
LIVE-IQA-r1	JPEG	0.92	0.87	0.89	0.81	0.93	0.94
LIVE-IQA-r2	JPEG	0.97 (T)	0.84	0.84	0.80	0.94	0.95

- A DNN-based model trained with “real-world distorted pictures” struggles to predict the quality of JPEG compressed images

Work Ongoing

- Currently working on different aspects of the template
 - What to report?
 - How to report it?
- Any person that can contribute is welcome
 - The updates are discussed during JEG-Hybrid meetings
 - Meeting schedule: <https://vqegjeg.github.io/jeg-hybrid/meetings>