Modern High Dynamic Range Imaging at the Time of Deep Learning

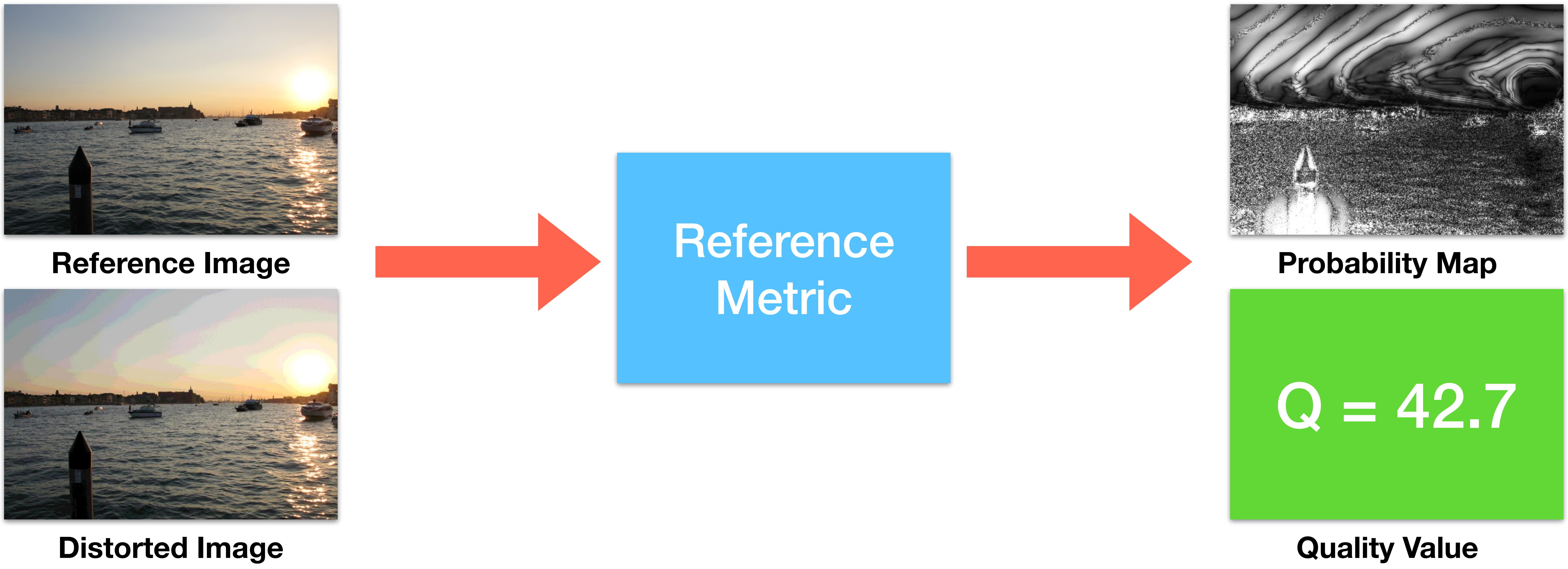
Deep HDR Metrics for Images

Francesco Banterle and Alessandro Artusi

Why Do We Need Metrics?

- In HDR/SDR Imaging, we need to determine and to understand what is happening during different steps of the pipeline:
  - **Acquisition**: we want to understand if there are artifacts due to acquisition or single image reconstruction;
  - **Compression**: we want small file size at maintaining high-quality;
  - **Tone mapping**: we want to adapt content for different display while keeping quality as it was “scene-referred”.

Reference Metrics



Reference Metrics: Current Limitations

- These models are very complex:
  - Difficult to port to GPUs with ease.
- They are computationally expensive; e.g., minutes of computations for a full HD image.
- Do we need a distortion map?
  - For most tasks we just need **a single value**!

DIQM: Deep Image Quality Metric

- A general and simple architecture meant for distilling reference-based metrics (e.g., HDR-VDP, DRIIM, etc.) into a CNN architecture.

DIQM: Datasets

	TRAINING SET	VALIDATION SET	TEST SET	TOTAL
HDR-C (HDR-VDP 2.2)	12,768	1,596	1,638	16,002
SDR-D (HDR-VDP 2.2)	11,536	1,441	1,441	14,418

DIQM: SDR-D Dataset

REFERENCE SDR IMAGE

BLUR DISTORTION

NOISE DISTORTION

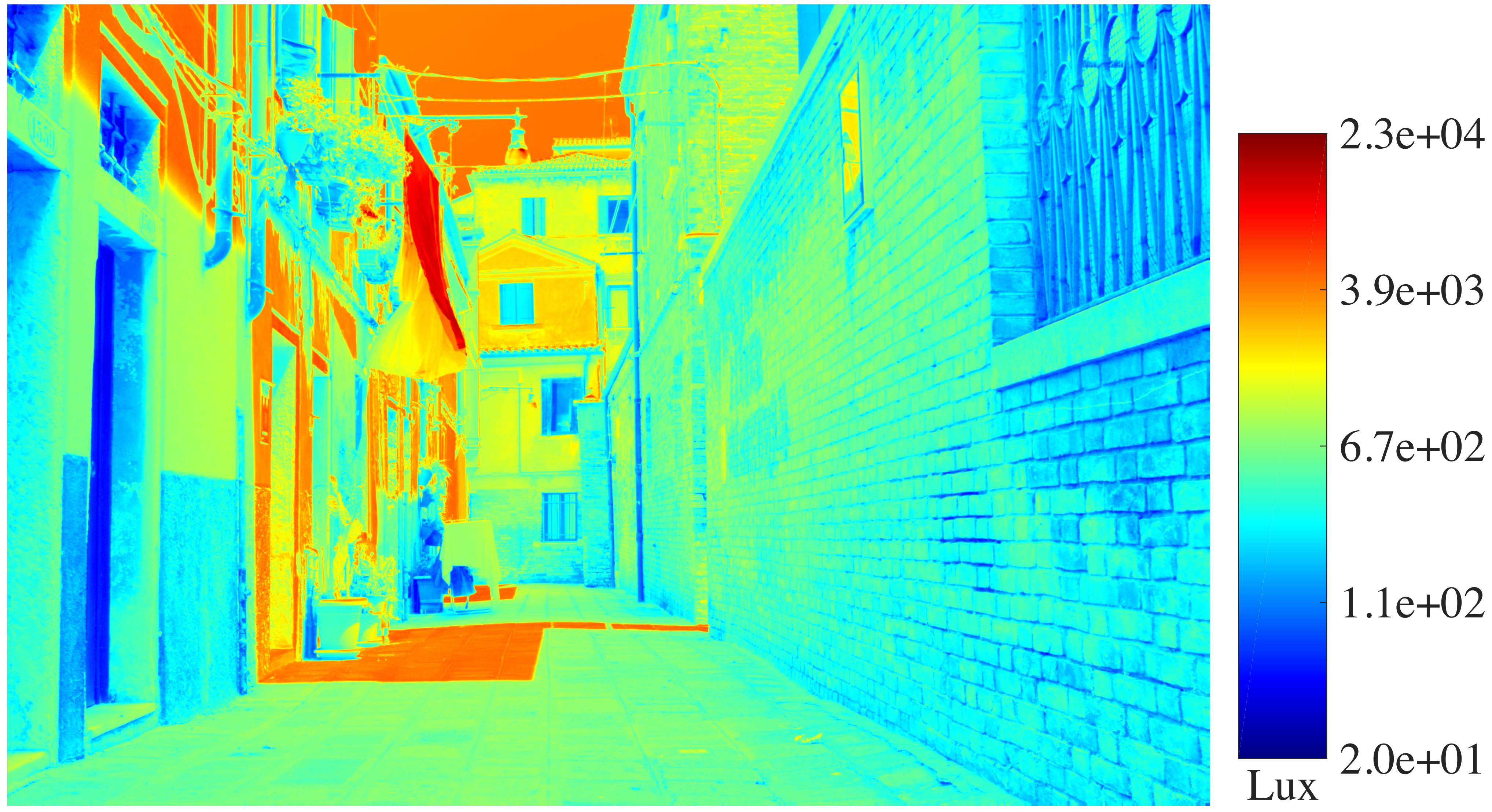
DIQM: SDR-D Dataset

REFERENCE SDR IMAGE

QUANTIZATION DISTORTION

SIN GRATE DISTORTION

DIQM: HDR-C Dataset



HDR Image

JPEG-XT:

- Random Profile
- Random Residual Compressione

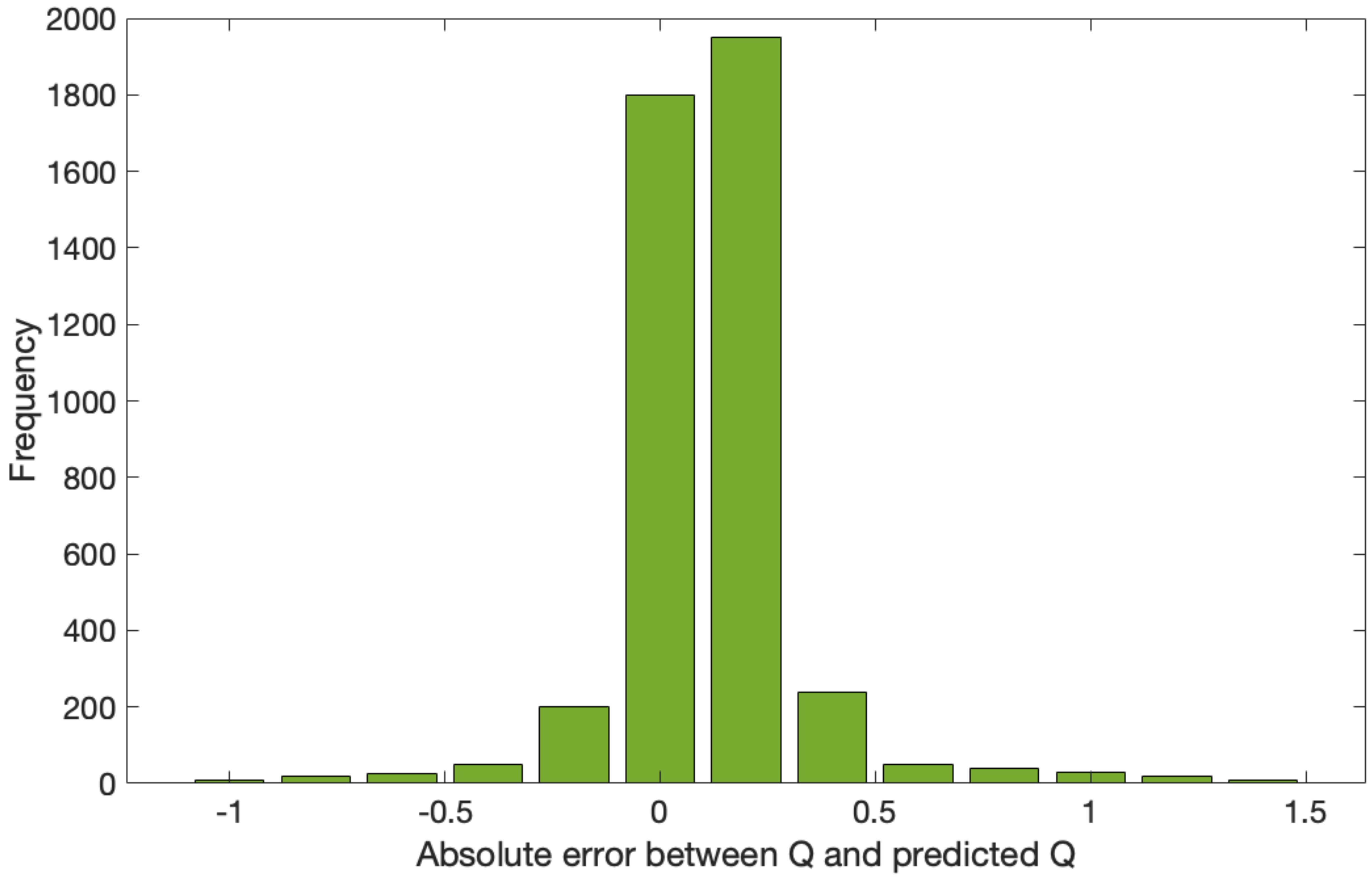
8-bit Layer

METADATA

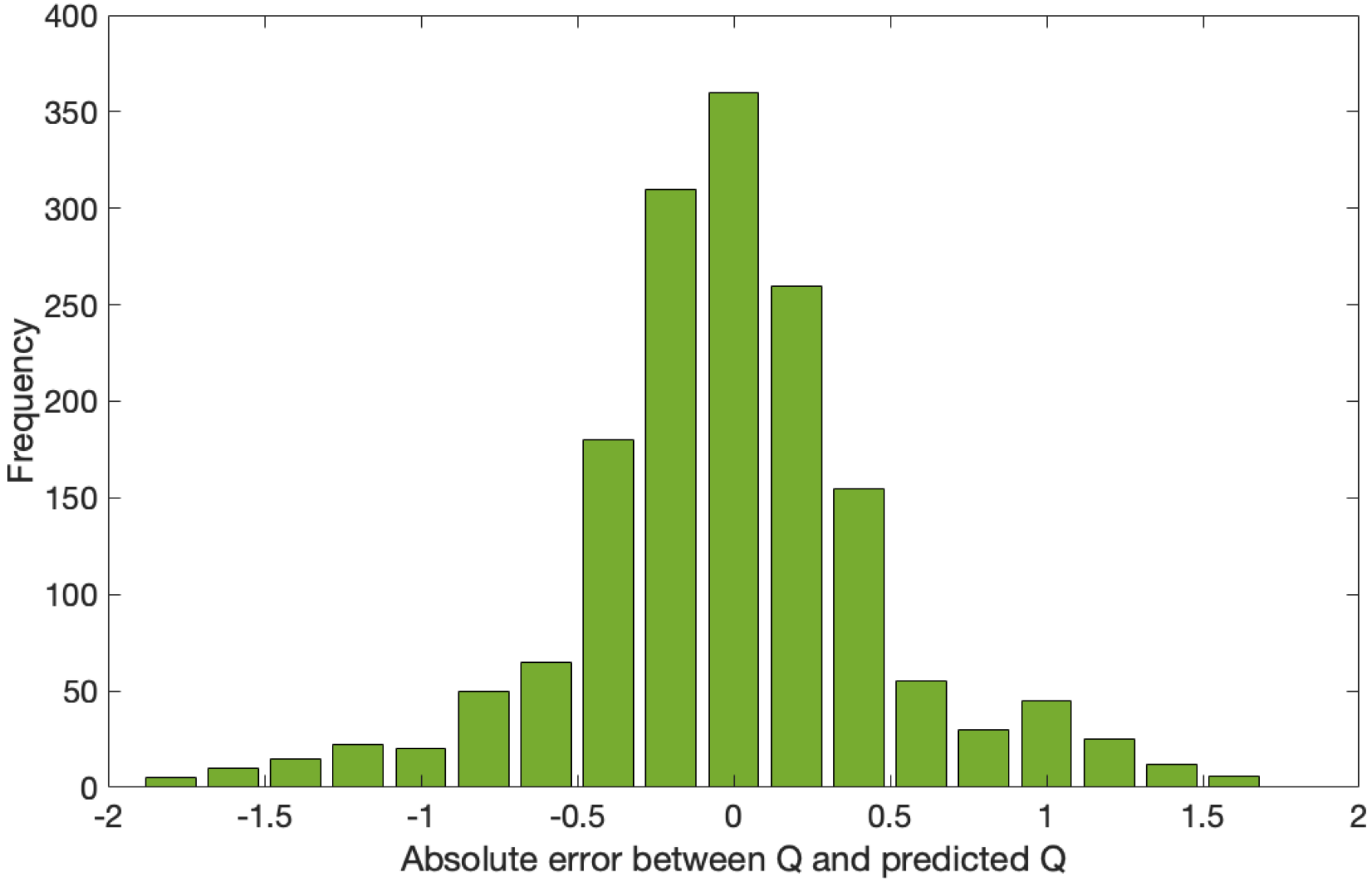
DIQM: Loss and Encoding

- Loss is a classic MSE; it works well for predicting quantitative values.
- Encoding:
  - SDR Images: linear scaling to fit the range $[0,1]$
  - HDR Images: $\log_{10}(x+1)$

DIQM: Results Test Set

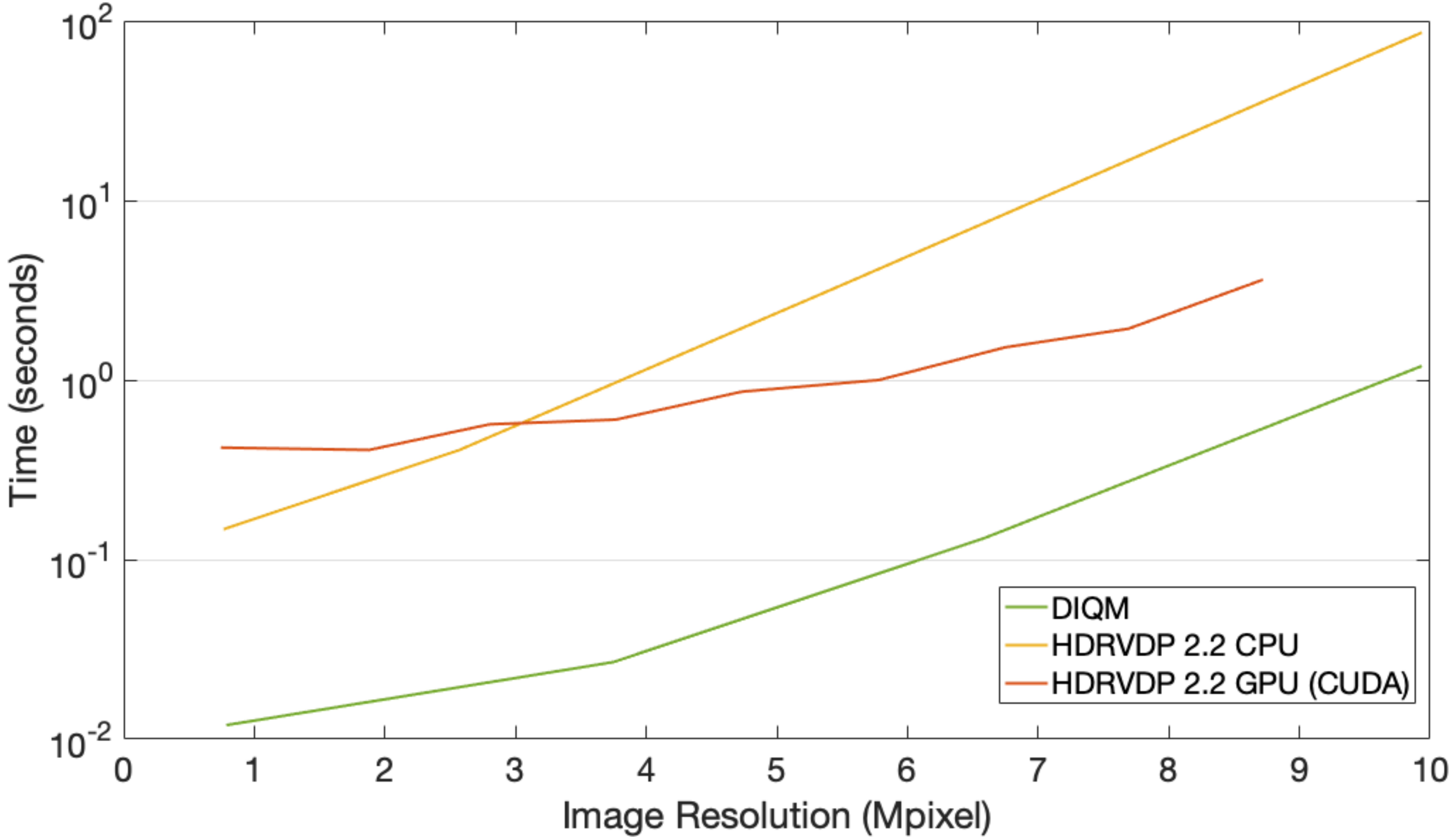


HDR-C



SDR-D

DIQM: Timings Results



DIQM: Conclusions

- There two main results:
  - We can distill metrics into a CNN with reasonable quality;
  - The CNN can be simple; no need of overly complex models:
    - The CNN runs real-time at inference time;
    - Small weights.

Visibility Distortion Maps CNN-based

- Several applications (imaging and computer graphics) are requiring a visual difference map
  - Traditional objective metrics can not be used, e.g., single numeric value
  - Existing visibility metrics produce a visual difference map, but they are inaccurate
    - Lack of large image collections with good coverage of possible distortion
    - A large dataset of image pairs (ground truth, distorted) is collected, e.g., user marking indicate wether the distortion is visible
    - A CNN is used and trained on this large dataset

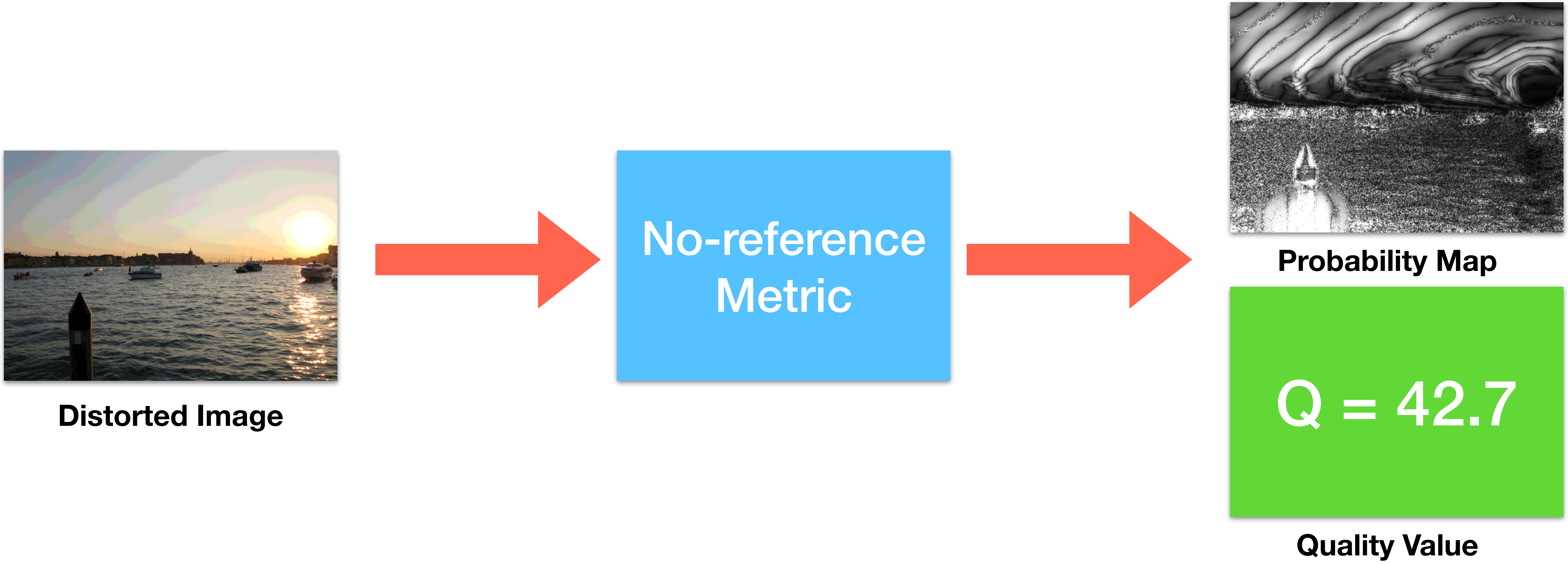
Visibility Distortion Maps CNN-based

Visibility Distortion Map: Conclusions

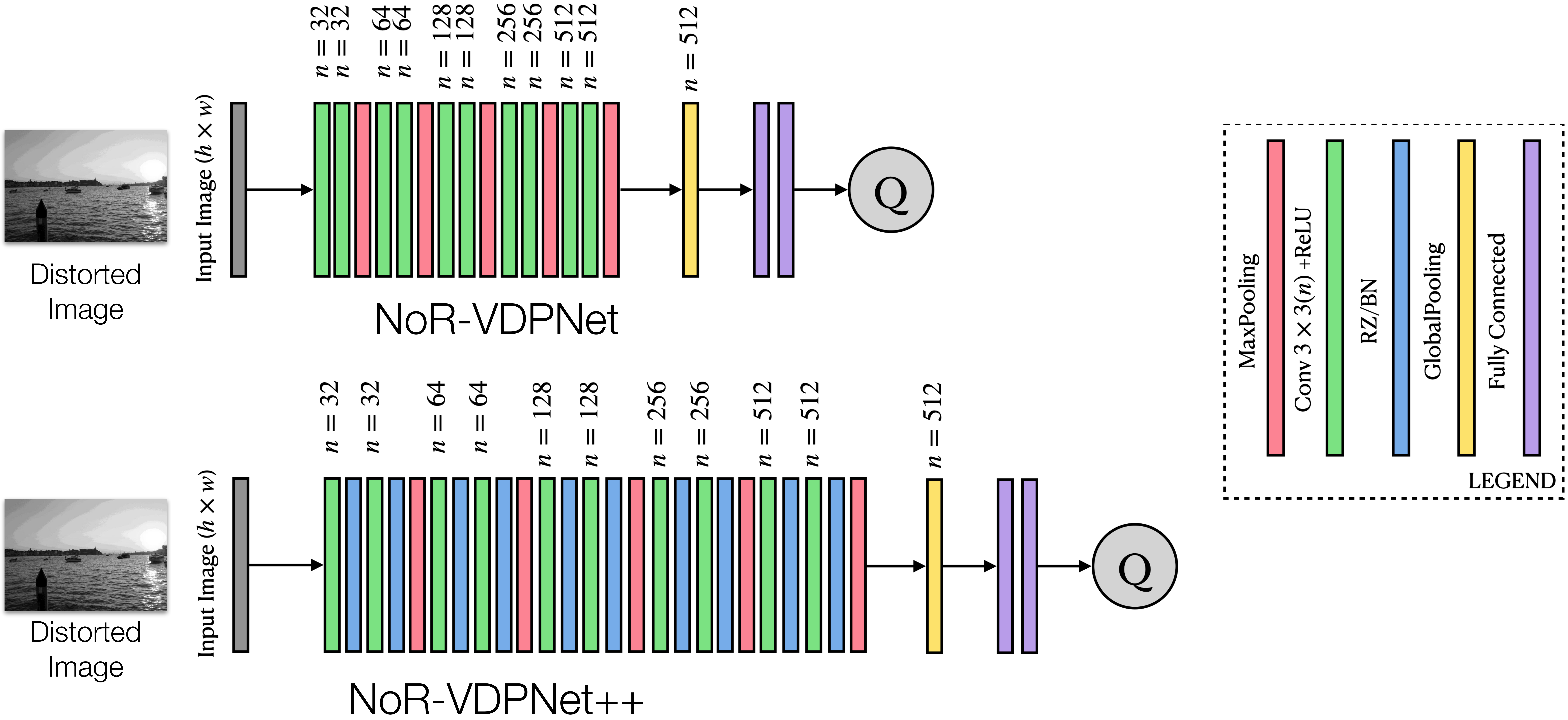
- There main results:
  - A statistical model has been proposed to fit the large data collected and used as loss function
  - Existing visibility metrics can be improved through the usage of a CNN based method, which it is trained using the collected dataset and using as loss function the proposed statical model

Going No-Reference

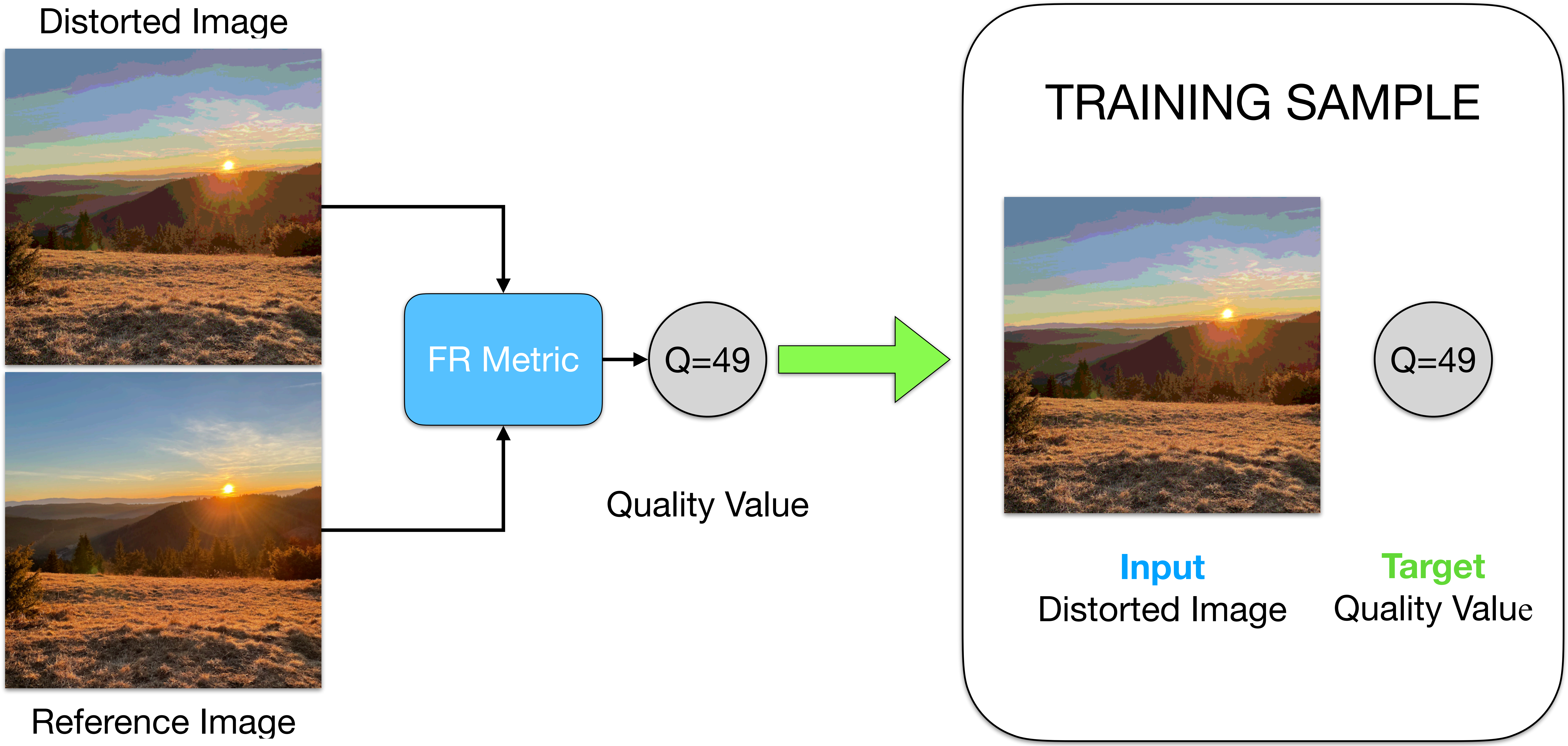
No-Reference Metrics



NoR-VDPNet(++): Architecture



Training Set

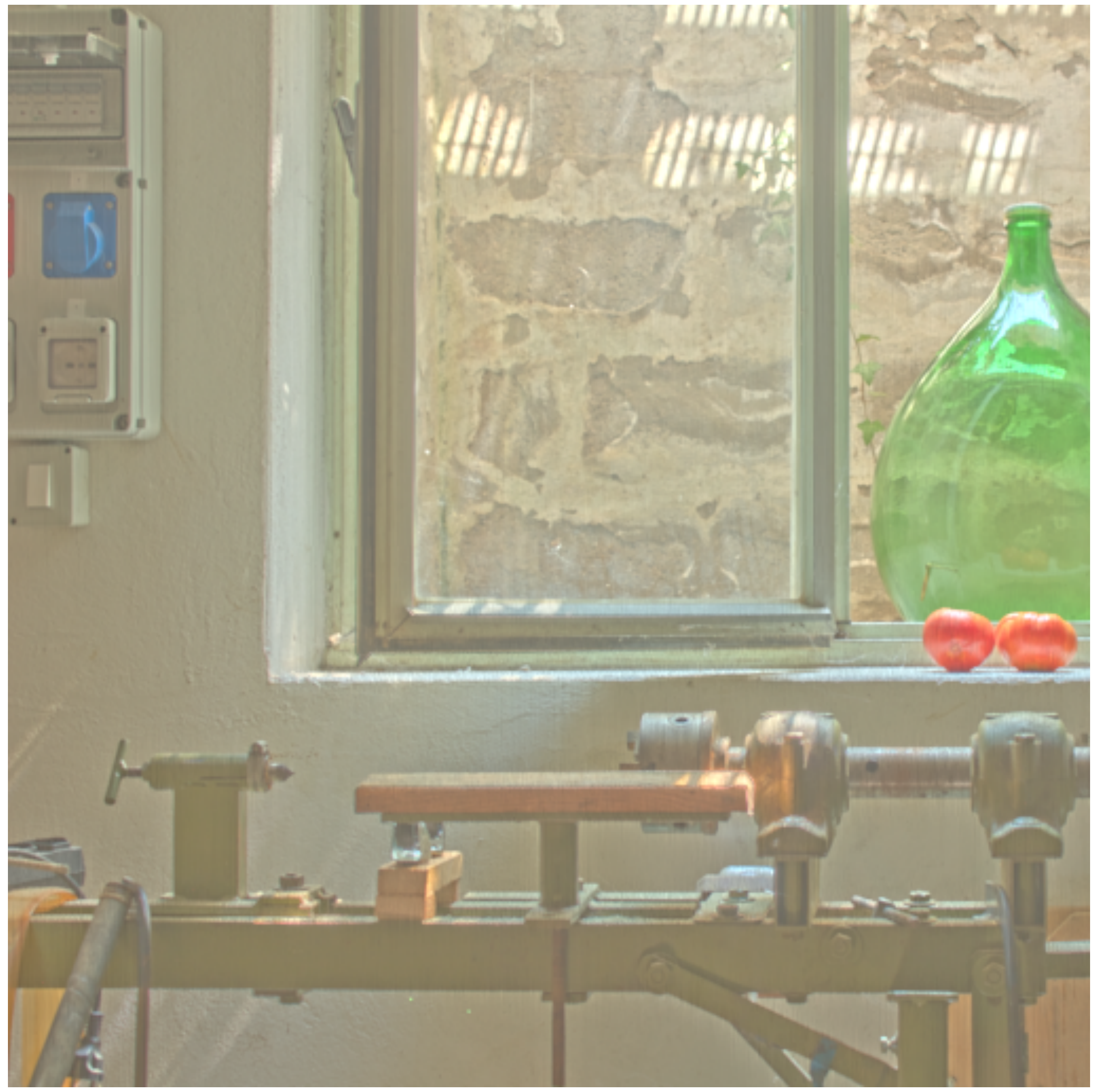


NoRVDPNet(++): HDR-VDP2.2/TMQI Datasets

	TRAINING SET	VALIDATION SET	TEST SET	TOTAL
HDR-C (HDR-VDP2.2)	49.602	6.216	6.216	62.034
SDR-D (HDR-VDP2.2)	80.244	10.025	10.044	100.313
TMO (TMQI)	106.290	13.320	13.320	132.930
ITMO (HDR-VDP2.2)	106.290	13.320	13.320	132.930

NoRVDPNet(++): TMO Dataset


Drago et al. 2003


Durand and Dorsey 2002


Reinhard et al. 2002

18 tone mapping operators from the HDR-Toolbox: https://github.com/banterle/HDR_Toolbox/

NoRVDPNet(++): ITMO Dataset

Input SDR Image

Eilertsen et al. 2017
(tonemapped)

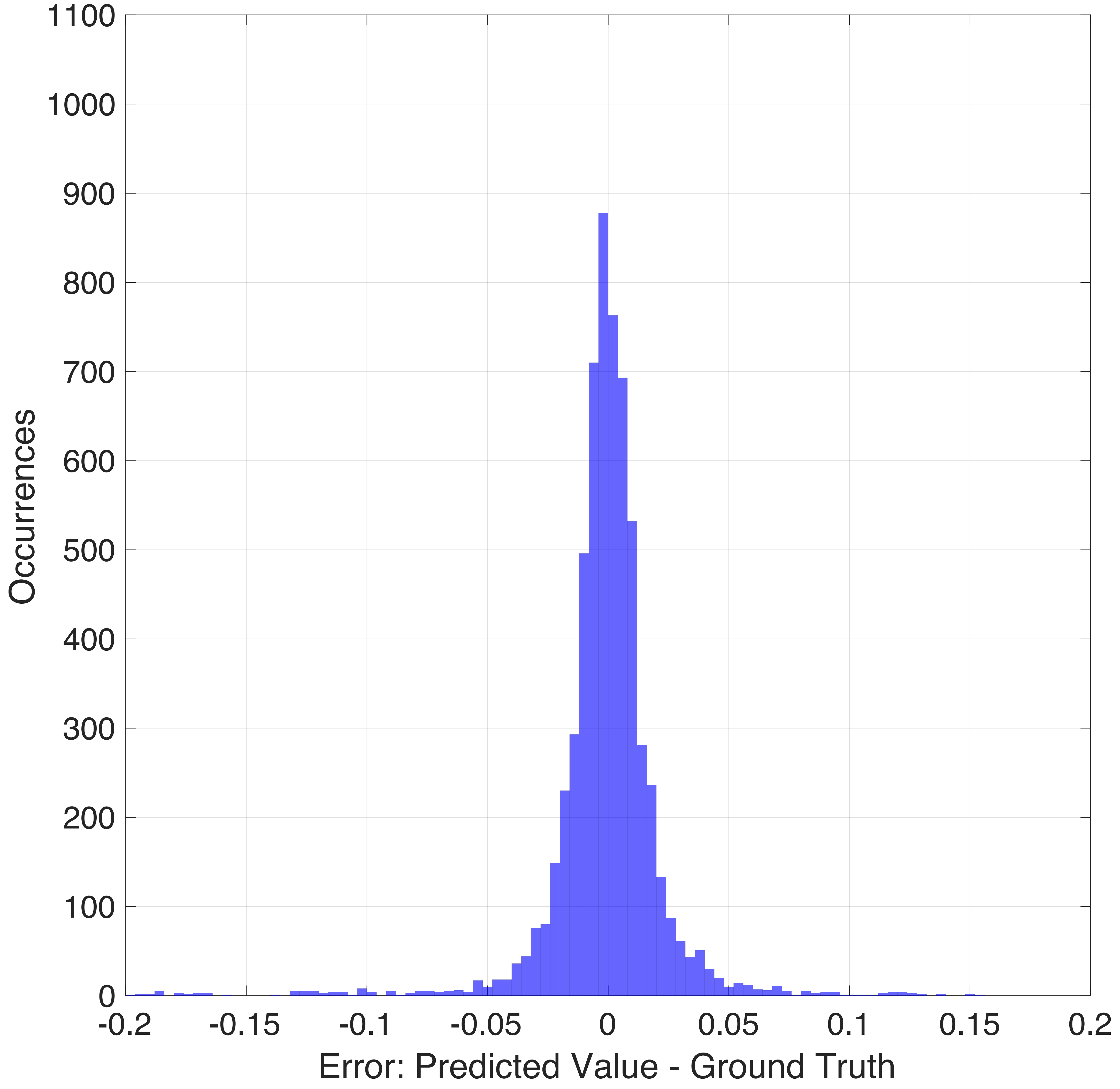
Santos et al. 20202
(tonemapped)

6 inverse tone mapping operators 4 available in the HDR-Toolbox: https://github.com/banterle/HDR_Toolbox/

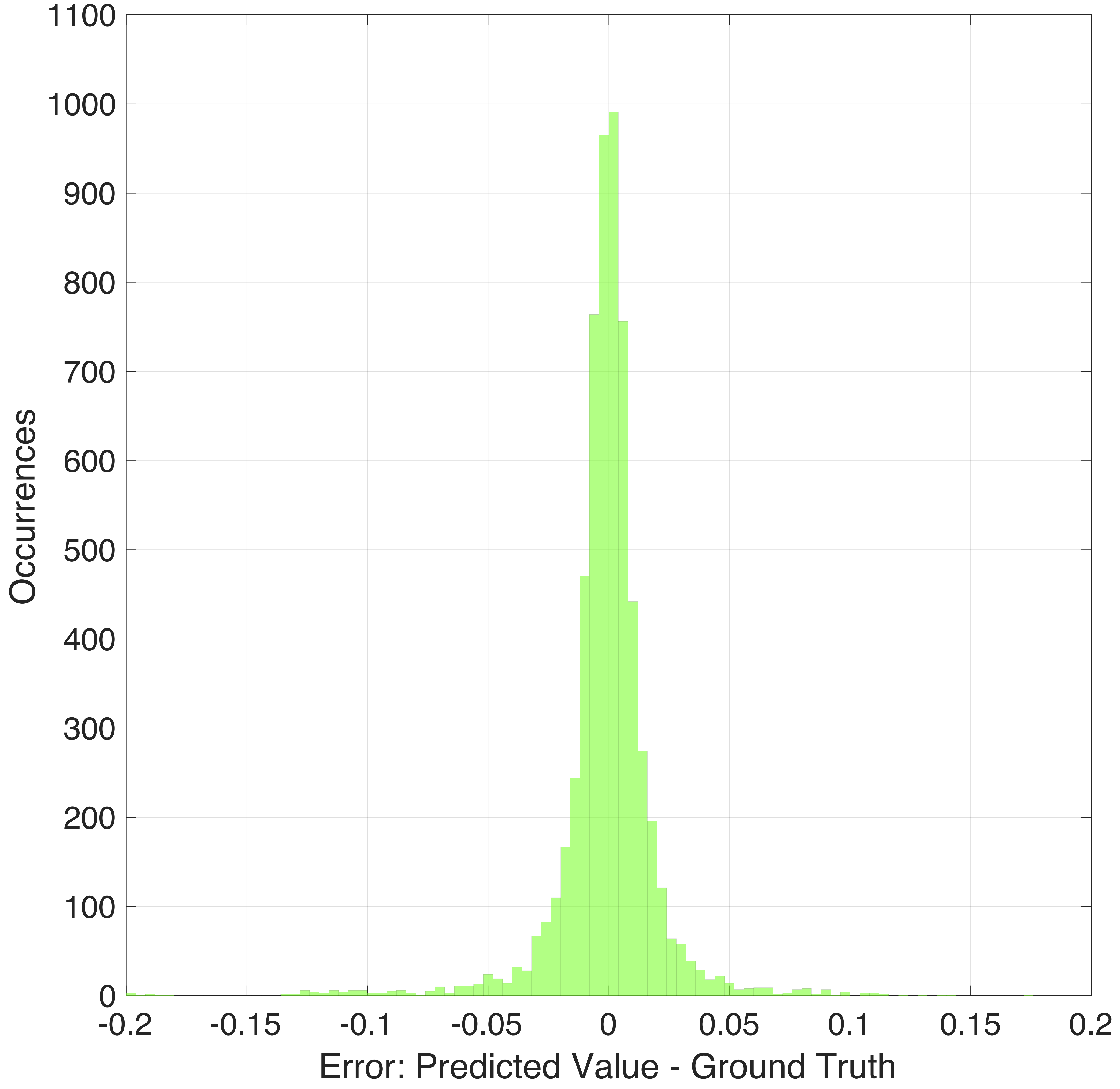
NoR-VDPNet(++): Loss and Encoding

- Loss is a classic MSE; it works well for predicting quantitative values:
- Encoding:
  - SDR Images: linear scaling to fit the range $[0,1]$
  - HDR Images: $\log_{10}(x+1)$

Results: HDR-C Test Set

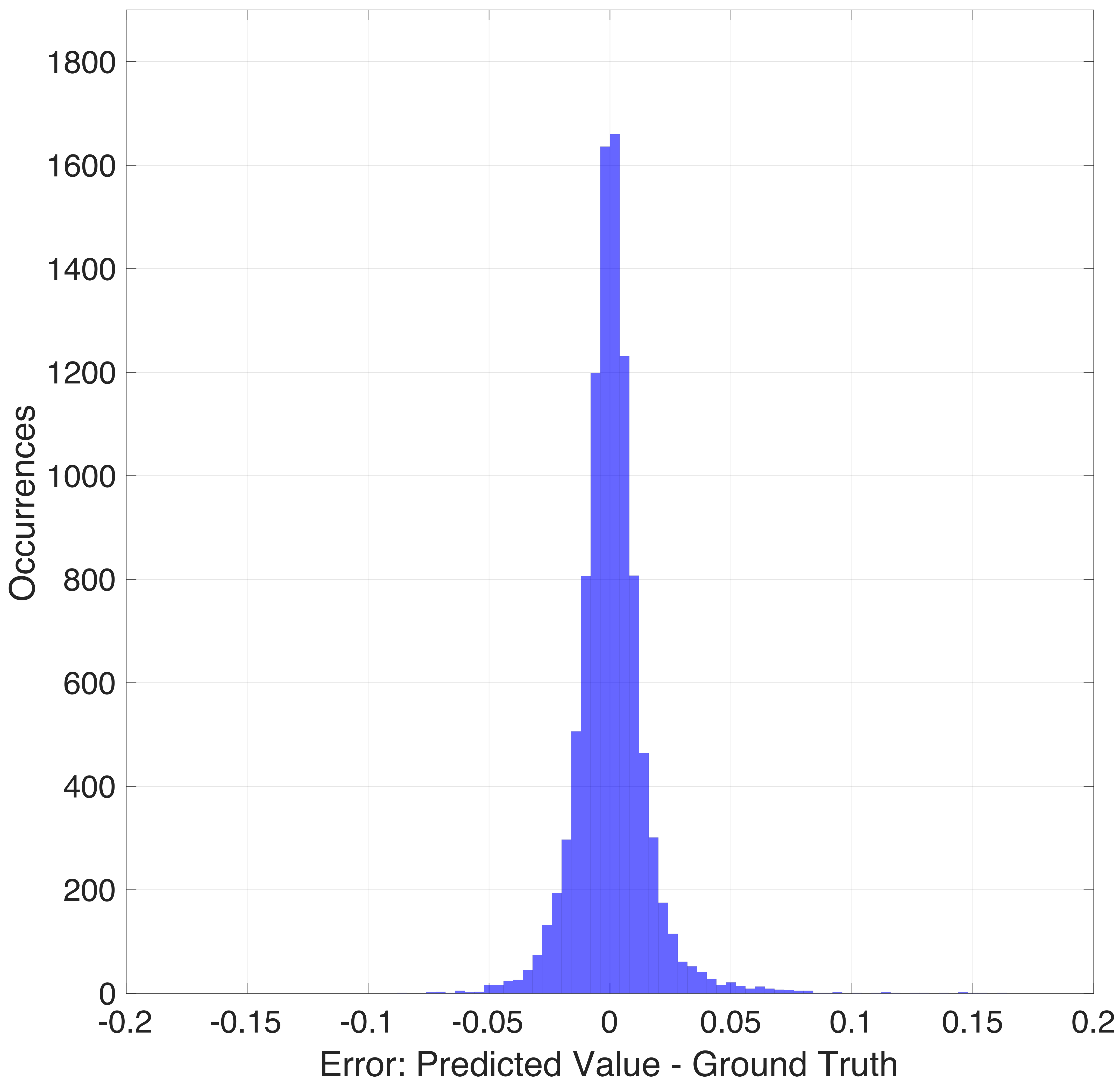


NoRVDPNet

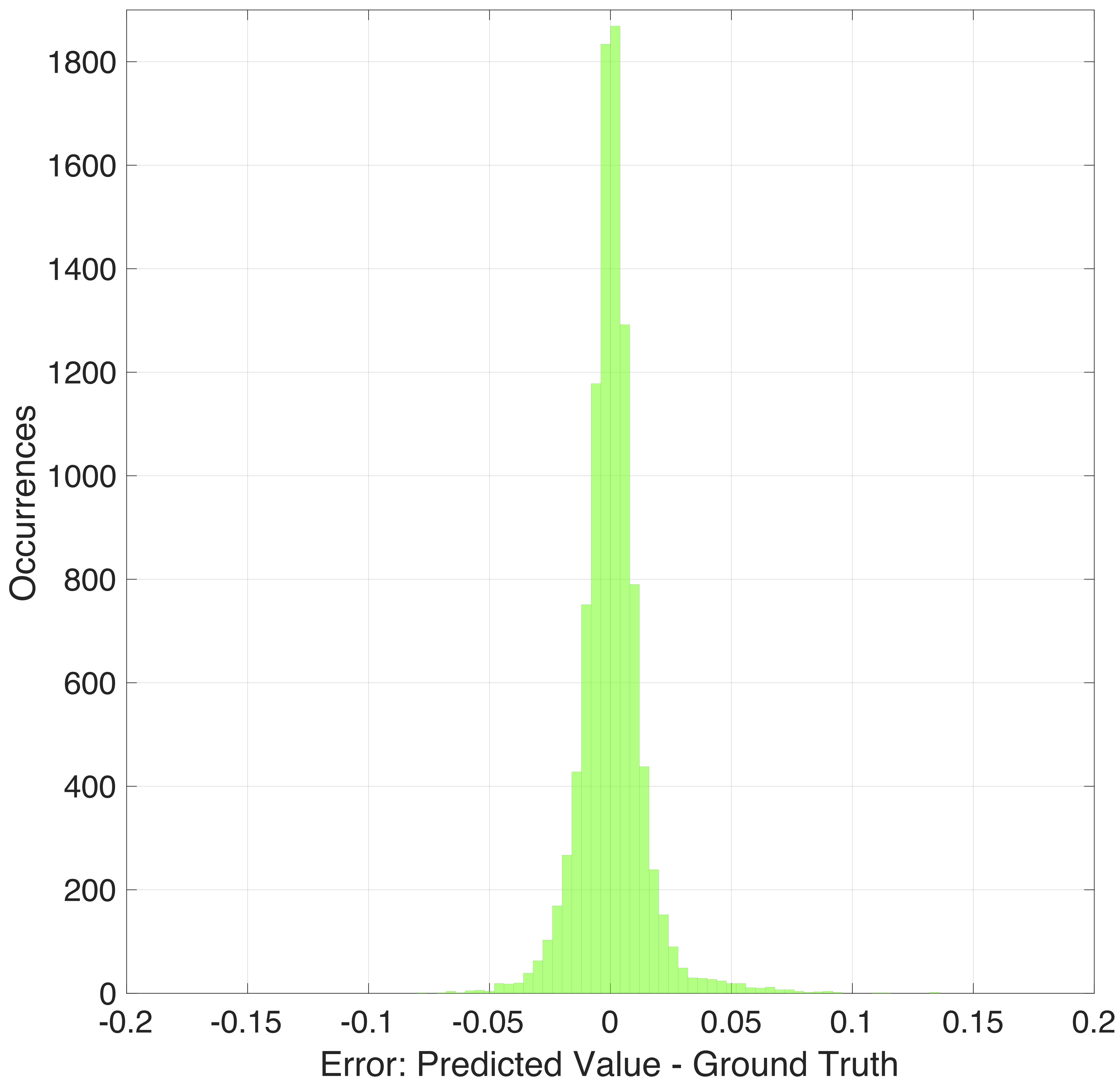


NoRVDPNet++

Results: SDR-D Test Set

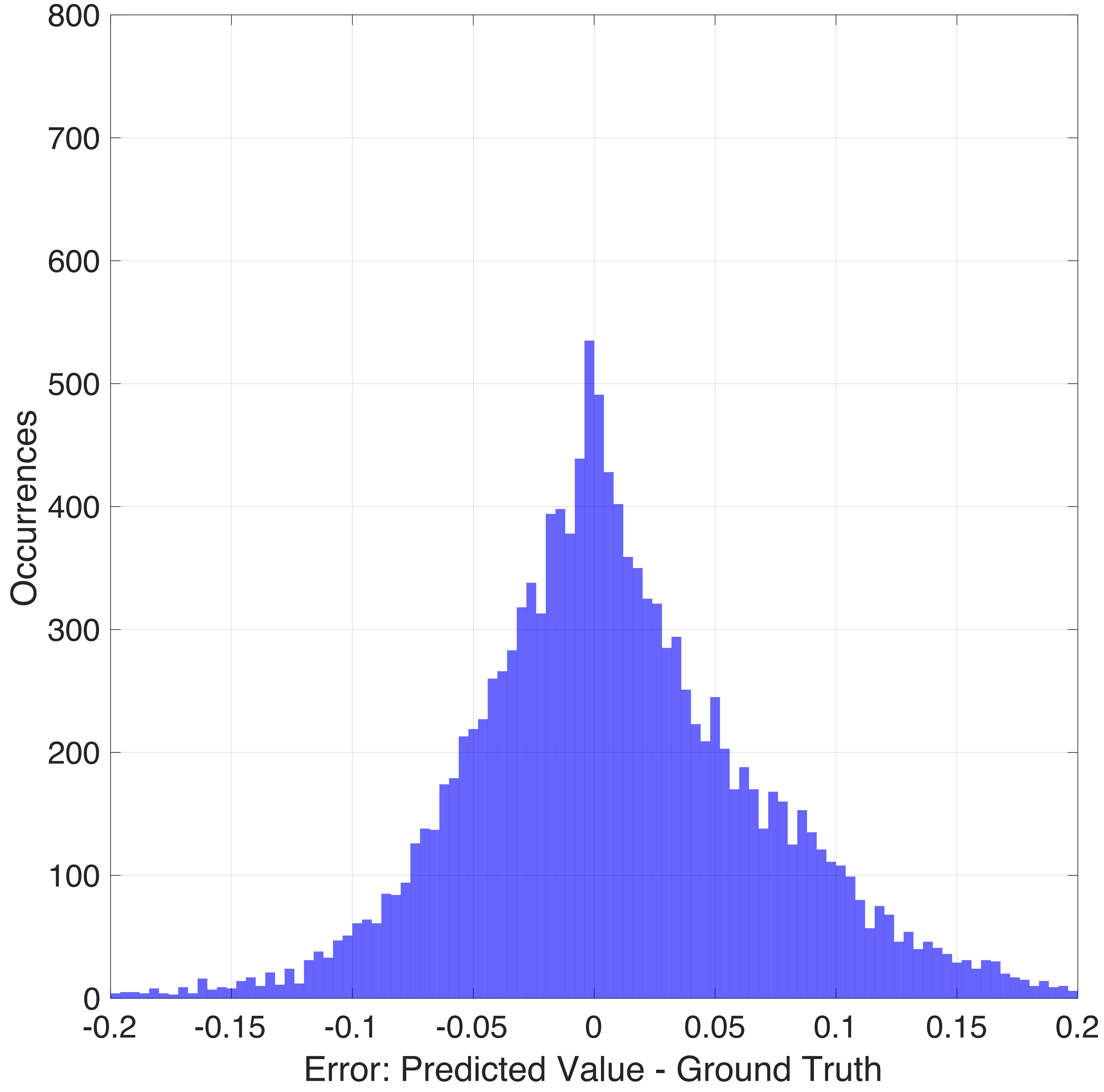


NoRVDPNet

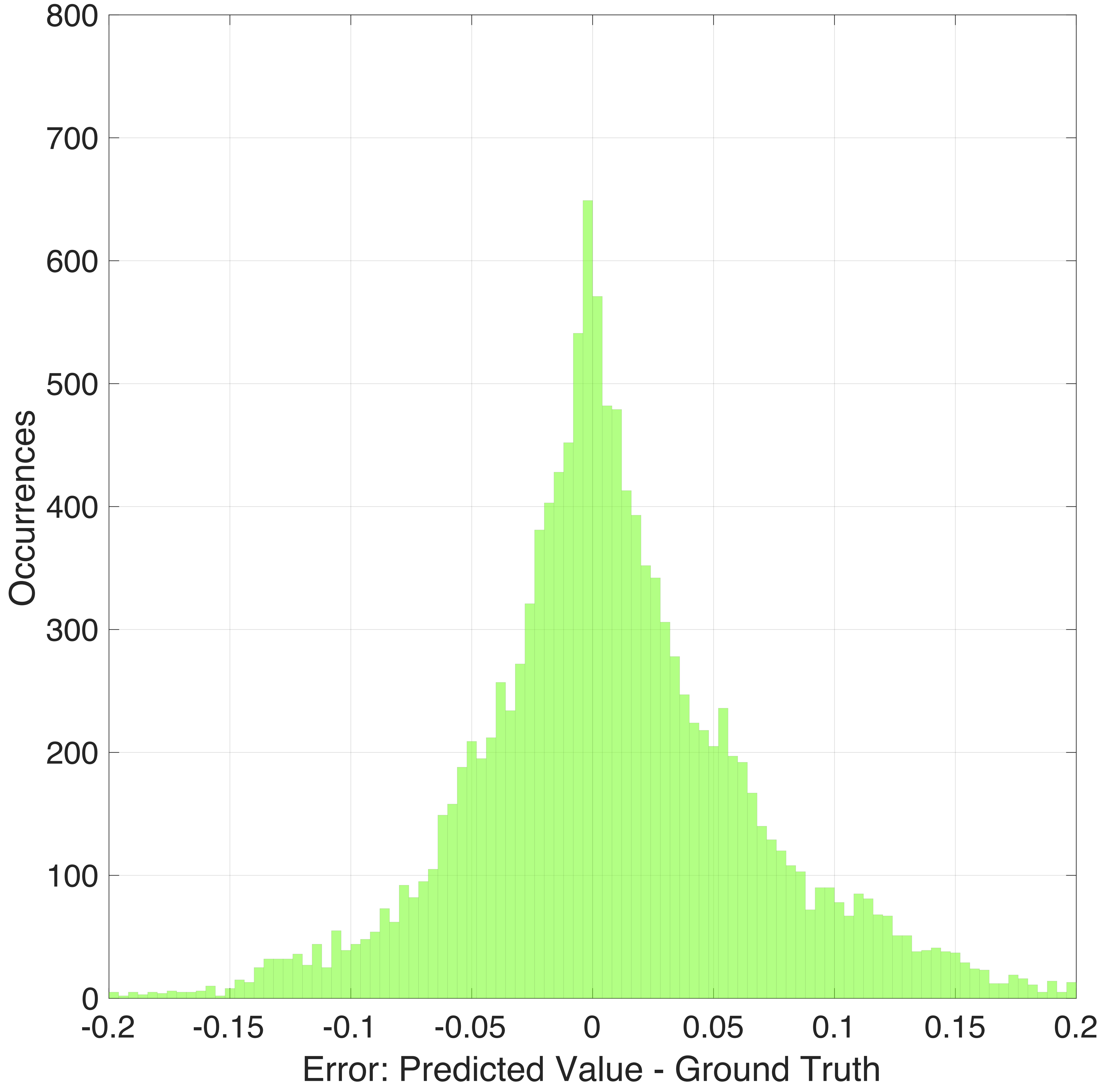


NoRVDPNet++

Results: ITMOS Test Set

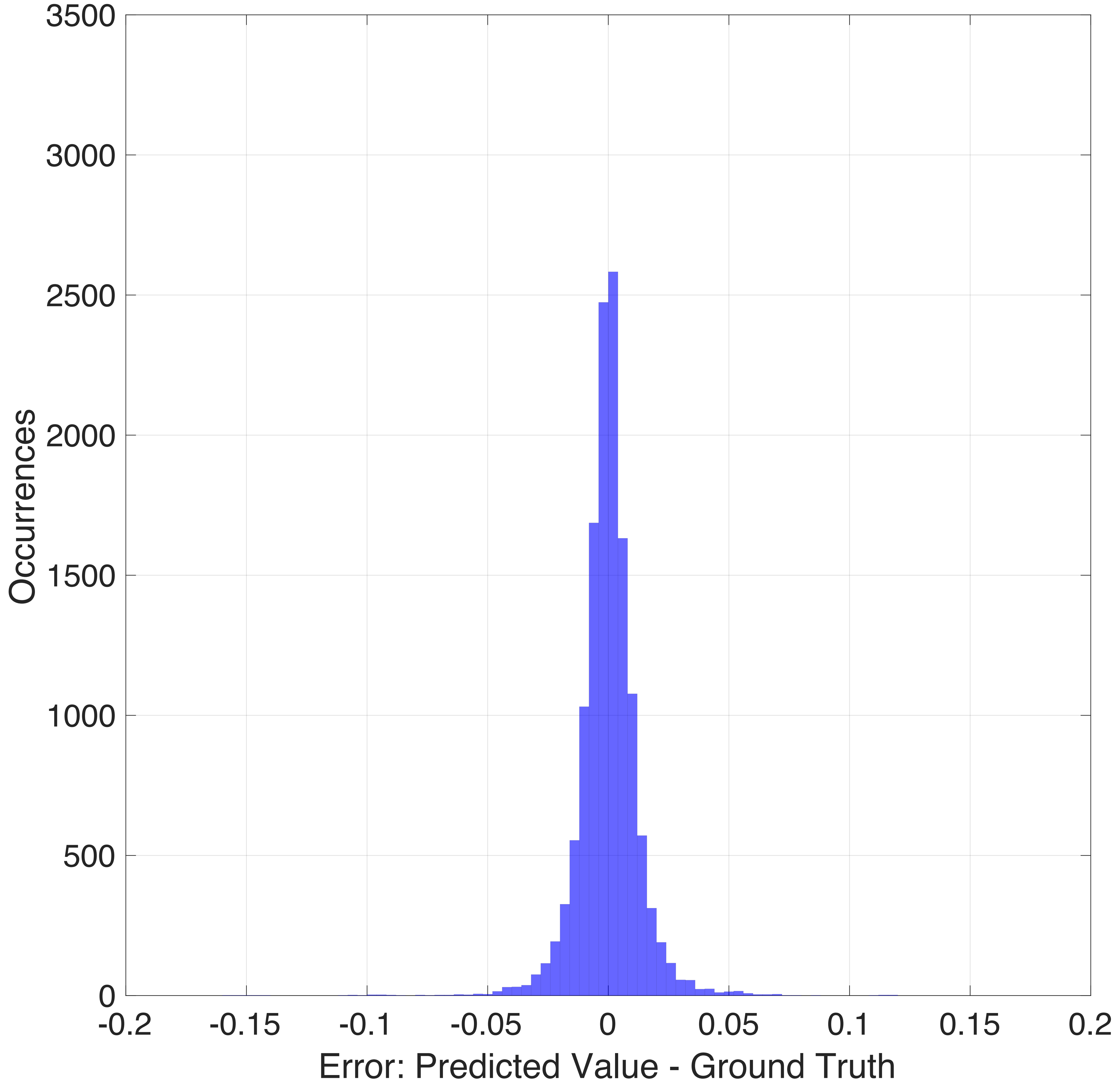


NoRVDPNet

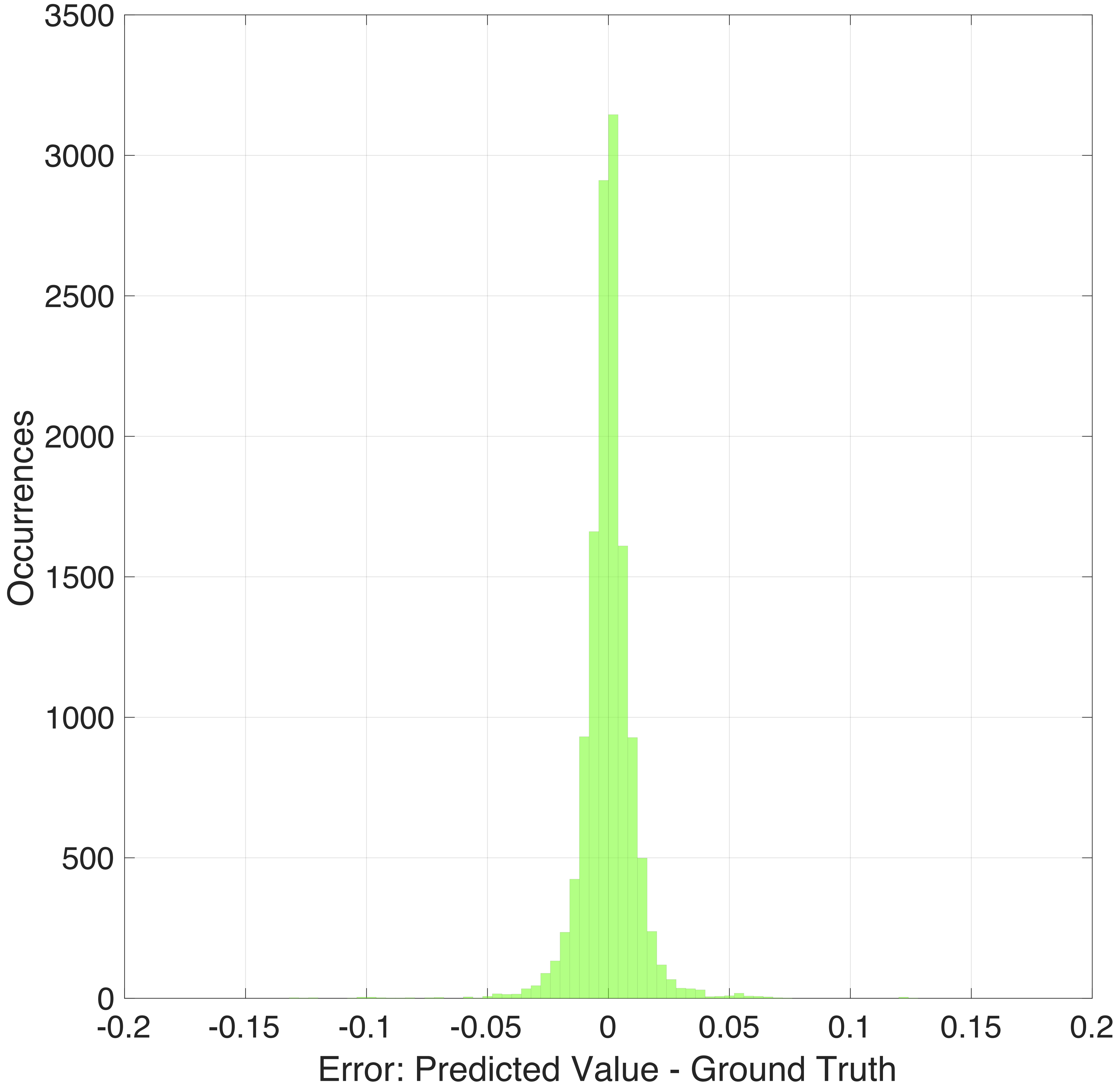


NoRVDPNet++

Results: TMOS Test Set

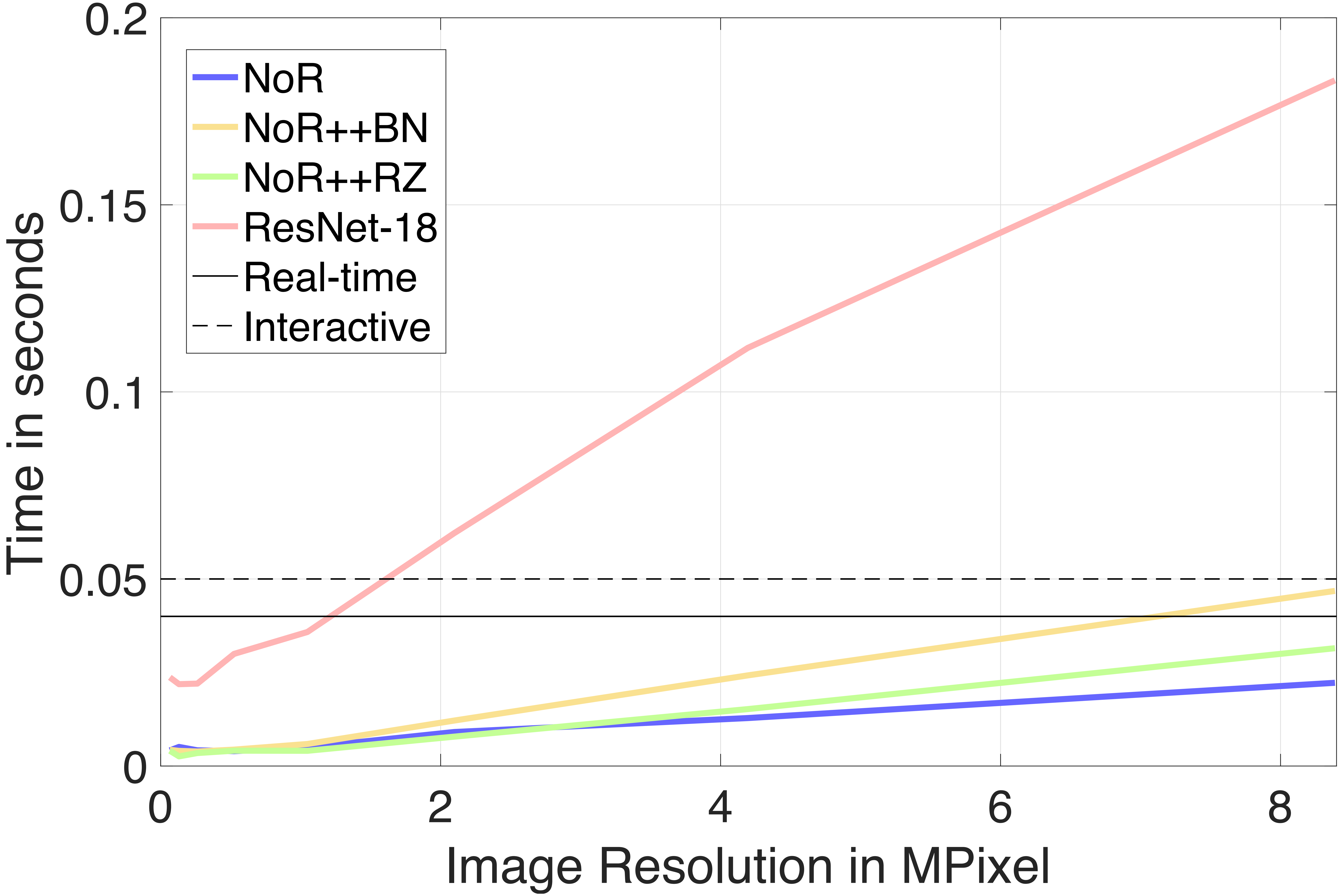


NoRVDPNet



NoRVDPNet++

Timings



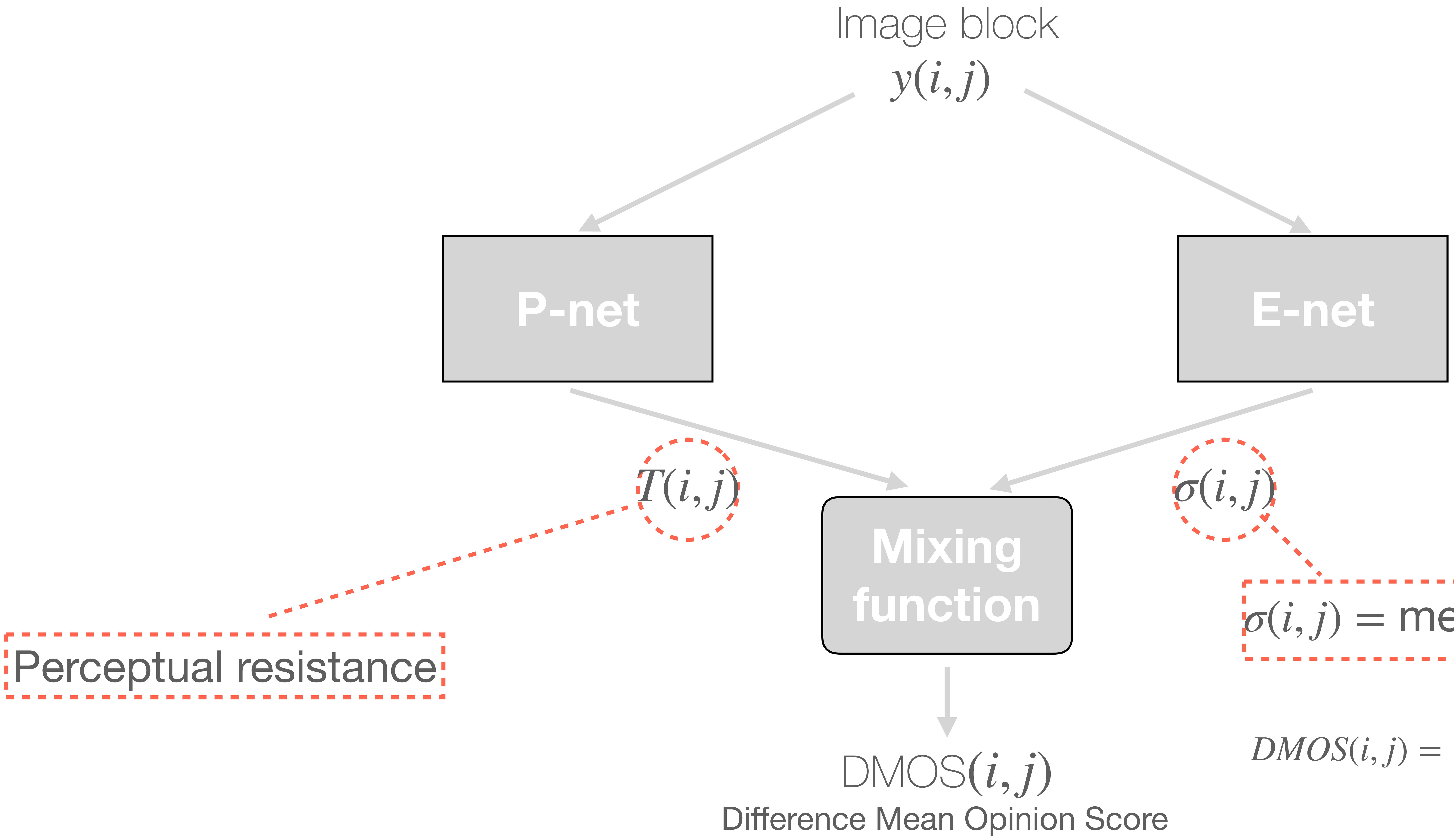
0.2
0.15
0.1
0.05
0
Time in seconds
0
2
4
6
8
Image Resolution in MPixel
NoR
NoR++BN
NoR++RZ
ResNet-18
Real-time
Interactive

NoR-VDPNet(++): Conclusions

- We can go from reference to no-reference;
- When we model several distortions we have a larger error than a single distortion;
- Layer normalization increases quality;
- This scheme works for TMQI-I (SSIM-based);
- Still real-time performance.

HDR NR-IQA

HDR NR-IQA Principle

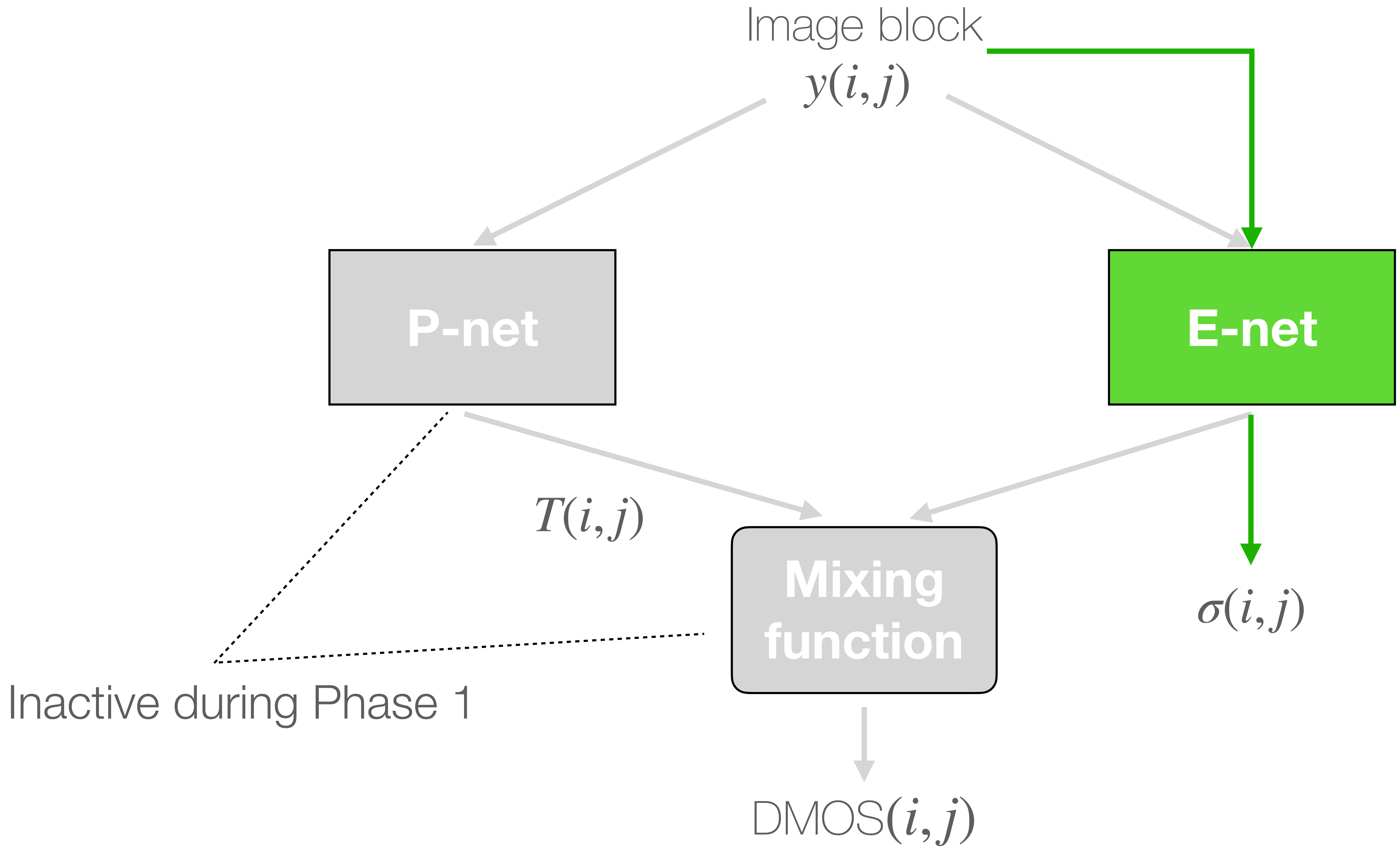


- Perceptual threshold
  - Not from psychophysical experiments
  - Data driven method

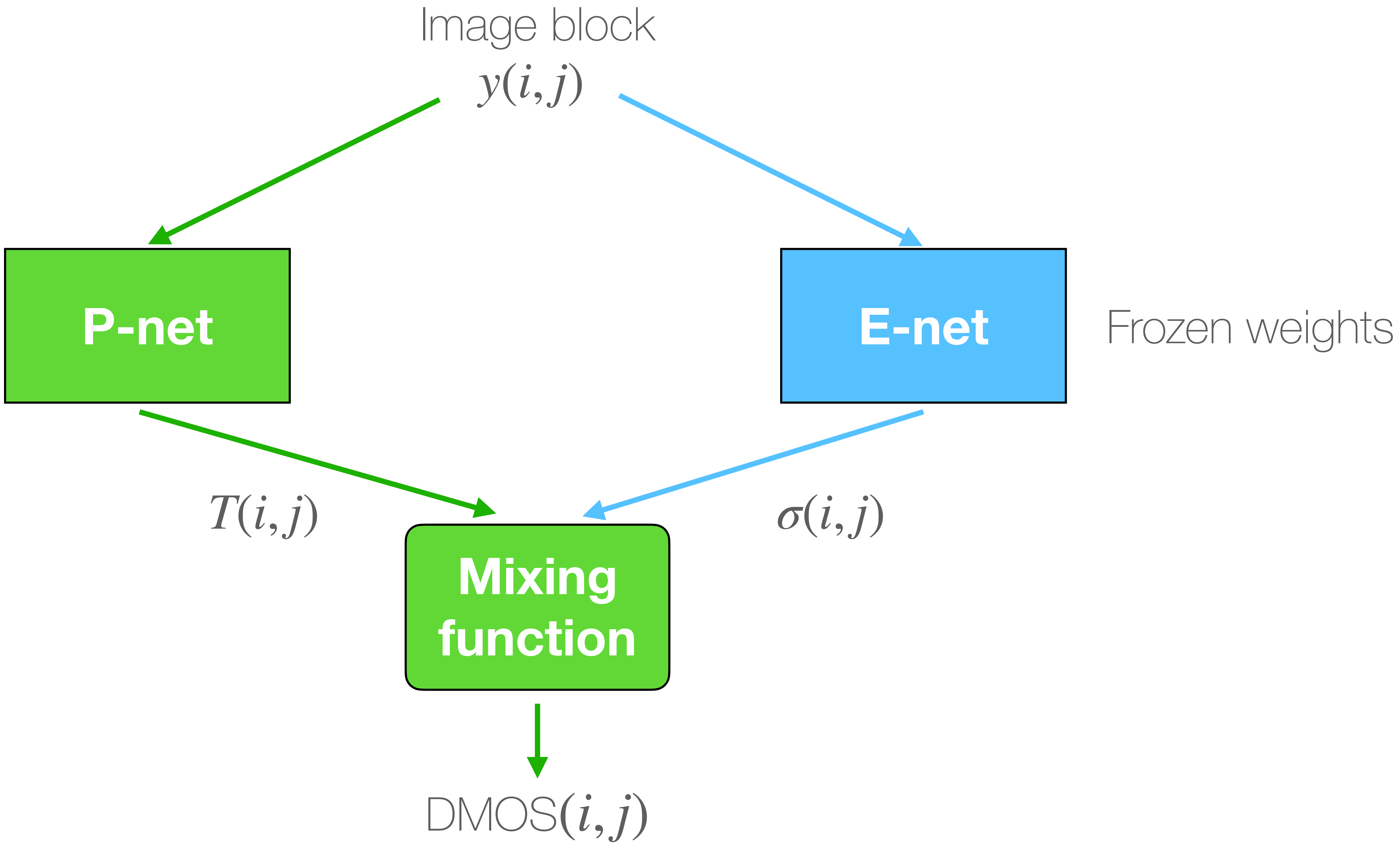
$$\sigma(i,j) = \text{mean}\left(\left|Y - Y_{\text{distorted}}\right|\right)$$

$$DMOS(i,j) = 1 - exp\left(-\left|\frac{k \times \tilde{\sigma}(i,j)}{T(i,j)}\right|\right)$$

HDR NR-IQA Training - Phase 1



HDR NR-IQA Training - Phase 2

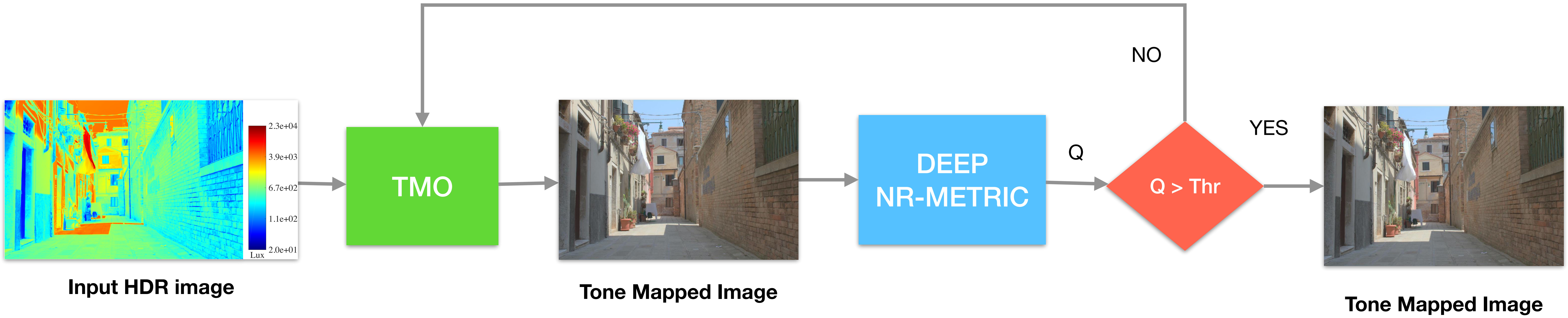


HDR NR-IQA: Conclusions

- Computational performances are not real-time, but it can be still optimized.
- It outperforms other NR-IQA methods.
- It is comparable to HDR FR-IQA:
    - without the need of a reference image.

Applications

Applications: TMO Optimization Task



Applications: Optimized TMO

TMO without optimized parameters

TMO with optimized parameters

Video Courtesy of Jan Fröhlich - Stuttgart HDR Video Dataset

Application: Optimized TMO


(a) $\hat{Q} = 0.903\ /Q = 0.885$


(b) $\hat{Q} = 0.906/Q = 0.930$


(c) $\hat{Q} = 0.933\ /Q = 0.914$


(d) $\hat{Q} = 0.918/Q = 0.903$


(e) $\hat{Q} = 0.902/Q = 0.889$


(f) $\hat{Q} = 0.841/Q = 0.771$


(g) $\hat{Q} = 0.951/Q = 0.831$


(h) $\hat{Q} = 0.875/Q = 0.909$

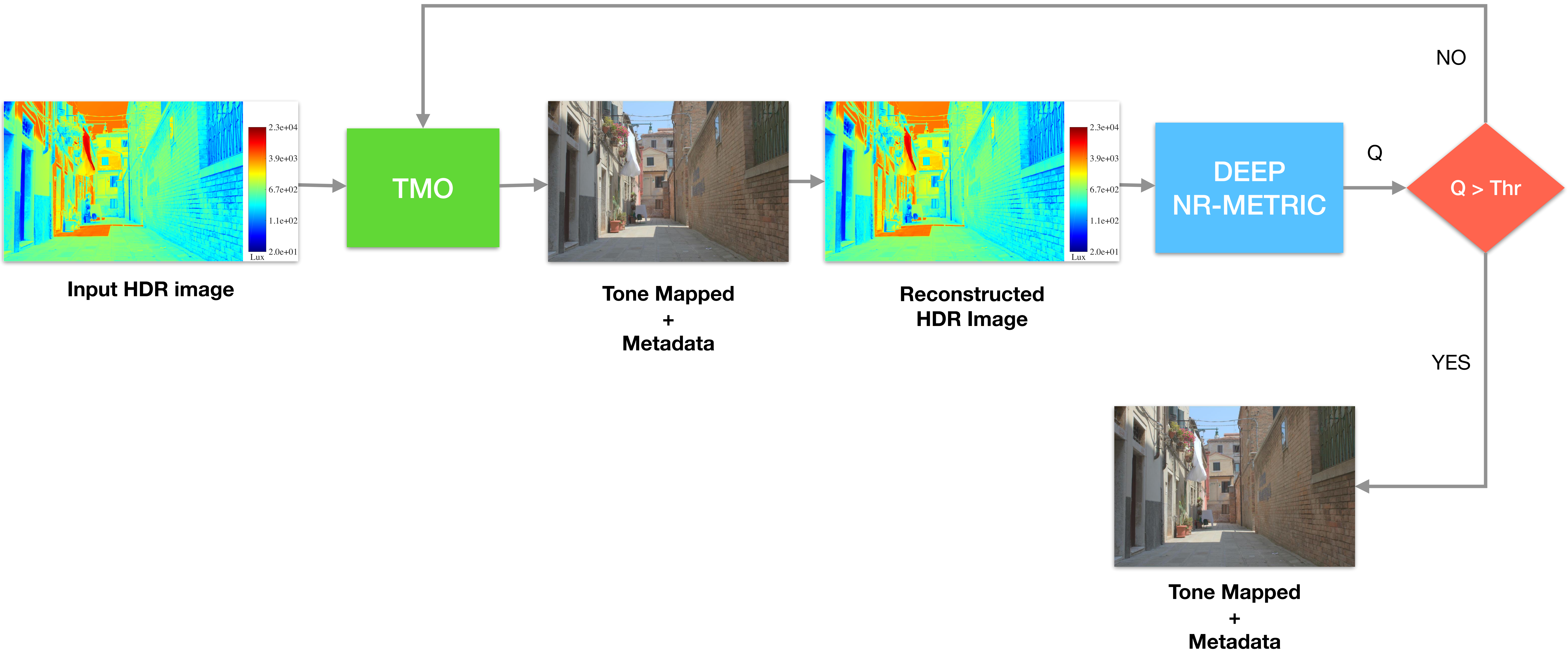

(i) $\hat{Q} = 0.951/Q = 0.967$


(j) $\hat{Q} = 0.958/Q = 0.974$


(k) $\hat{Q} = 0.967/Q = 0.976$


(l) $\hat{Q} = 0.997/Q = 0.979$

Applications: JPEG-XT Compression Task



Applications: Results JPEG-XT Compression

Input HDR image

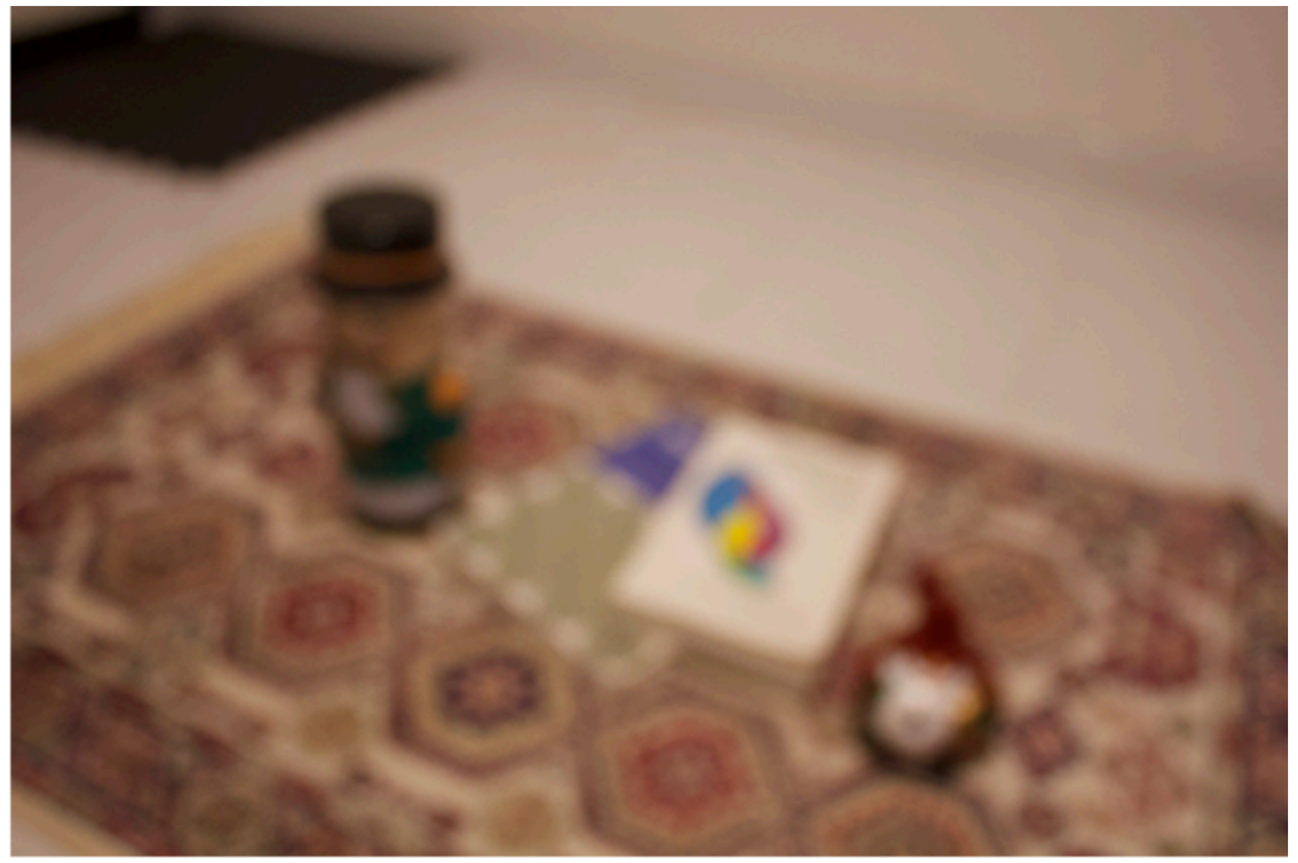
Reinhard et al.'s TMO
optimized with NoRVDPNet

Tone Mapped HDR image
for JPEG-XT

Applications: Photo Selection

Q=86.99

Q=86.92



Q=56.46

Q=91.39

Q=76.26

Q=59.9

Applications: Photo Selection

Q=86.99

Q=86.92

Q=56.46

Q=91.39

Q=76.26

Q=59.9

Future Directions

Future Directions

- Going in the temporal domain.
- Extend approaches to perceptual uniform domains.
- Mix perceptual experiments results and metrics.

Thank you for your attention!

Please contact us at:

a.artusi@cyens.org.cy francesco.banterle@isti.cnr.it

or visit us:

https://deepacamera.org.cy http://vcg.isti.cnr.it

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 739578 and No. 820434, and from the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

This project is co-financed by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (POST-DOC/0916/0034).