

Factsheets regarding the training and test sets

1 Training sets of the metrics

Potential Caveat: Some metrics are trained on datasets we used. See below

Metric	KADID w/ Labels	KADID w/o Labels	LIVE w/ Labels	PIPAL	KonIQ w/ labels	BAPPS	DTD	CLIP data	web	Other training data
LPIPS	–	–	–	–	–	✓	–	–		
lpips-vgg+	–	–	–	–	–	✓	–	–		
DISTS	✓	–	–	–	–	–	✓	–		
AHIQ	–	–	–	✓	–	–	–	–		
BRISQUE	–	–	✓	–	–	–	–	–		
DBCNN	–	–	–	–	✓	–	–	–		synthetic-distortion pretext (label-free)
HyperIQA	–	–	–	–	✓	–	–	–		
PaQ-2-PiQ	–	–	–	–	–	–	–	–		own authentic-distortion dataset
MUSIQ	–	–	–	–	✓	–	–	–		
MANIQA	–	–	–	–	✓	–	–	–		
TReS	–	–	–	–	✓	–	–	–		
TOPIQ-NR	–	–	–	–	✓	–	–	–		
CLIP-IQA	–	–	–	–	–	–	–	✓		
LIQE	–	–	–	–	✓	–	–	✓		
ARNIQA	–	✓	–	–	✓	–	–	–		

✓ marks the training data of the shipped weights (pyiqa defaults per its `default_model_configs.py` and checkpoint names; piq defaults likewise); red cells overlap our test sets (KADID-10k, LIVE). ImageNet-pretrained backbones are not listed.

- DISTS: structure/texture parameters jointly optimized to match human ratings on KADID-10k (together with DTD textures), per the TIP 2020 paper.
- topiq_fr: default weights are cfanet_fr_kadid_res50 (KADID-10k). A KADID-clean variant exists: `pyiqa:topiq_fr-pipal` (-0 metrics.topiq_fr.impl:pyiqa:topiq_fr-pipal).
- BRISQUE: the shipped SVR model is the official LIVE-trained one (true of every public release).
- ARNIQA: encoder pre-trained self-supervised on KADID-10k images (no labels); the linear regressor is fit on KonIQ-10k.
- LIQE: default weights are the KonIQ-only checkpoint; do *not* switch to `liqe_mix` (adds KADID-10k, LIVE MD, BID, CLIVE – LIVE MD is one of Zhongling’s test sets).
- Q-Align: disabled by default; its Q-Instruct fine-tuning mixture must be audited before enabling.

2 Datasets with Degradation labels

AI Generated (Check for correctness):

Audit of the `dis_type` column of `datasets/processed/*.csv` against the earlier list. That list was almost complete; the corrections are that [LIVE MD](#) and [PIPAL](#) were missing (PIPAL's file names encode 116 type-level combinations, by far our largest labelled pool); that [MDIVL](#) is usable after all, since the distortion pair is the directory name and only the MOS table is password protected; that [KonIQ-10k](#) is not usable, being in-the-wild photographs whose extra table holds quality *attributes* rather than degradation types; and that [MDID2013](#) / [MDID2016](#) do have known component distortions, but every image carries all of them at once, so there is no per-image class. The other 11 datasets of `datasets/processed/` (AADB, AVA, BIQ2021, CID2013, FLIVE, GFIQA, KonIQ-10k, LIVE Challenge, PIQ2023, SPAQ, UHD-IQA) are authentically distorted and carry no degradation type at all.

Dataset	Images	Types	Source of the label
In the repo			
PIPAL	23200	116	Filename code AXXXX_GG_LL in 7 groups; our CSV collapses all of it to PIPAL_distortions
KADID-10k	10125	25	Filename code + original table
TID2013	3000	24	Filename code + original table; our CSV merges codes 21 and 24 into Multiply_Dist. and so shows 23
TID2008	1700	17	Filename code + original table
PDAP-HDDS	12000	10	Subdir + filename
CSIQ	866	6	Subdir name
CIDIQ-MOS50	690	6	Subdir name
LIVE	779	5	Subdir name
VCLFER	552	4	Filename
SDIVL	380	2	Subdir name
Toyama	196	2	Subdir name; 28 of the images are undistorted references
LIVE MD	450	2 pairs	Subdir name (Part_1/blurjpeg, Part_2/blurnoise); our CSV collapses both to Multiply Dist.
MDIVL	750	2 pairs	Subdir name (Data_Blur_JPEG, Data_Noise_JPEG); same collapse
MDID2016	1600	1 combo	No label needed: all five types on every image at random levels
MDID2013	324	1 combo	No label needed: blur, then JPEG, then noise on every image
Not downloaded yet – no images, no scores, no MOS table in this repo			
Waterloo Exploration	94880	4	Filename + table; ships no MOS, only the D/L/P consistency tests
PieAPP	20280	75	Filename; 44 training + 31 test types, disjoint; labels are pairwise preferences, no MOS
CIDIQ-MOS100	690	6	Subdir name; the MOS50 images rated at 100 cm viewing distance

Dataset	Images	Types	Source of the label
IVC	185	4	Filename

The six degradation families below are the only ones shared by more than a handful of sets. ✓ means the set has at least one class in that family; (✓) means it occurs only in combination with the others, so it is not separable (the multiply distorted sets). For the four datasets we do not have yet, the marks come from their papers rather than from our files.

Dataset	Noise	Blur	JPEG	JP2K	Contrast	Colour
PIPAL	✓	✓	✓			
KADID-10k	✓	✓	✓	✓	✓	✓
TID2013	✓	✓	✓	✓	✓	✓
TID2008	✓	✓	✓	✓	✓	
PDAP-HDDS	✓	✓	✓	✓	✓	✓
CSIQ	✓	✓	✓	✓	✓	
CIDIQ-MOS50	✓	✓	✓	✓		✓
LIVE	✓	✓	✓	✓		
VCLFER	✓	✓	✓	✓		
SDIVL	✓		✓			
Toyama			✓	✓		
LIVE MD	(✓)	(✓)	(✓)			
MDIVL	(✓)	(✓)	(✓)			
MDID2013	(✓)	(✓)	(✓)			
MDID2016	(✓)	(✓)	(✓)	(✓)	(✓)	
Not downloaded yet						
Waterloo Exploration	✓	✓	✓	✓		
PieAPP	✓	✓	✓		✓	✓
CIDIQ-MOS100	✓	✓	✓	✓		✓
IVC		✓	✓	✓		

Degradations of a set that the matrix does not cover:

- **KADID-10k**: lens blur, motion blur, impulse noise, multiplicative noise, denoising, brighten, darken, mean shift, jitter, non-eccentricity patch, pixelate, quantization, colour block, high sharpen; its colour column stands for colour diffusion, colour shift, colour quantization and saturation shift in HSV and in Lab.
- **TID2008**: correlated (pink) noise, masked noise, high-frequency noise, impulse noise, quantization noise, denoising, JPEG and JP2K transmission errors, non-eccentricity pattern noise, local block-wise distortion, mean shift; its noise column also covers noise in the chroma components only.
- **TID2013**: all of TID2008's, plus multiplicative Gaussian noise, comfort noise, dithering, chromatic aberration, lossy compression of noisy images, and sparse sampling with reconstruction; its colour column is saturation change.
- **PDAP-HDDS**: motion blur, fast fading, brightness, scaling.
- **LIVE**: fast fading.

- [CSIQ](#): pink (1/f) noise.
- [CIDIQ](#) (both halves): its noise is Poisson rather than Gaussian, and its colour column is ΔE and SGCK gamut mapping.
- [PIPAL](#): the bulk of its 116 combinations are super-resolution, denoising and GAN outputs and mixtures thereof; the group names must be taken from the official type list.
- [PieAPP](#): distortions aimed at the HVS (non-eccentricity, contrast sensitivity) and at algorithm artifacts (deblurring, denoising, super-resolution, geometric and colour transforms).
- [IVC](#): LAR coding.
- [VCLFER](#), [SDIVL](#), [Toyama](#), [Waterloo Exploration](#): nothing beyond the matrix.

3 Overview of the Datasets

This table outlines the size of the datasets, specifically,

- Number of reference images
- Number of distorted images
- Number of distortion types (or combinations)
- Type of original labels (MOS, DMOS, pairwise / 2AFC, aesthetics, ...)
- Observed range of those original labels
- Range of our alignment-experiment MOS (the labelled subset)

AI Generated (Check for correctness):

Counts are taken from *datasets/processed/*.csv* (one row per distorted image; references counted as distinct `ref_img_path` values) and agree with the combined score tables of *datasets/iqa_scored/combined/*. The grouping follows the experiment configs: the all-metrics models are tested on the FR+NR sets only, while the NR-metrics models (*fusion/options/wqi_x1_nr*.yaml*) add the authentic, reference-less sets.

Original labels are the `MOS_orig` / `DMOS_orig` column (higher-is-better MOS, Elo, JOD and aesthetics; higher-is-worse DMOS). Original ranges are the min–max actually present in those CSVs, not the nominal scale of the rating protocol. **Our MOS** is the alignment-experiment score on the labelled subset (40 images per set, 39 for *Toyama*), on a common 0–100 scale; the rest of each CSV has that cell empty.

Dataset	References	Distorted	Distortion types	Image size	Labels	Original range	Our MOS
Training set (unsupervised, simulated)							
WQI x1	18381	18381	random chains†	1920 × 1080	–	–	–
Test sets — FR+NR metrics							
CIDIQ-MOS50	23	690	6	800 x 800	MOS (DS, 9-cat.)	[1.18, 7.65]	[18.56, 82.07]
CSIQ	30	866	6	512 x 512	DMOS (pairwise)	[0, 1]	[7.89, 61.56]
KADID-10k	81	10125	25	512 × 384	MOS (ACR)	[1.00, 4.93]	[17.83, 64.88]
LIVE	29	779	5	~700 x 500 (variable)	DMOS (DSCQS)	[−2.64, 111.8]	[2.39, 74.49]
LIVE MD	15	450	2 combinations	1280 x 720	DMOS (SS-HR)	[0.61, 84.67]	[21.21, 67.63]
MDID2013	12	324	1 combination	768x512	DMOS (SS)	[0.32, 0.55]	[21.52, 48.27]
MDID2016	20	1600	1 combination	512x384	MOS (pairwise)	[0.08, 7.92]	[16.35, 58.33]
MDIVL	10	750	2 combinations	886x591	MOS (SS)	[1.41, 97.97]	[20.44, 79.68]
PDAP-HDDS	250	12000	10	1920x1080	MOS (ACR)	[1, 5]	[16.78, 84.28]

Dataset	References	Distorted	Distortion types	Image size	Labels	Original range	Our MOS
PIPAL	200	23200	116 combinations	288x288	Elo (2AFC)	[917, 1836]	[9.01, 46.44]
SDIVL	20	380	2	886x591	MOS (SS)	[8.17, 87.73]	[38.92, 75.56]
TID2008	25	1700	17	512x284	MOS (Swiss 2AFC)	[0, 7.71]	[7.92, 61.98]
TID2013	25	3000	24	512x384	MOS (Swiss 2AFC)	[0.24, 7.21]	[7.79, 61.21]
Toyama	14	168	2	768x512	MOS (ACR)	[1.00, 4.88]	[26.23, 69.34]
VCLFER	23	552	4	683x512	MOS (SS)	[1.57, 96.52]	[18.33, 76.02]
Test sets – NR metrics only (authentic distortions, no references)							
AADB	–	9831	authentic	1024 on long side	aesthetics	[0, 1]	[37.47, 83.65]
AVA	–	255502	authentic	~500 - 700	aesthetics	[1.81, 8.60]	[9.13, 81.39]
BIQ2021	–	12000	authentic	512	MOS (ACR, scaled)	[0, 1]	[17.64, 62.69]
CID2013	–	474	authentic	1600x1200	MOS (DR-ACR)	[−9.55, 87.88]	[12.79, 88.50]
FLIVE	–	159240	authentic	256x256 for major- ity. 64-280	MOS (crowd-sourced)	[0, 100]	[13.10, 84.17]
GFIQA	–	19998	authentic	512x512	MOS (scaled)	[0.005, 0.941]	[20.41, 79.16]
KonIQ-10k	–	10073	authentic	1024 × 768	MOS (ACR)	[1.10, 4.31]	[17.42, 87.77]
LIVE Challenge	–	1161	authentic	500x500	MOS (crowd-sourced)	[3.42, 92.43]	[17.09, 63.27]
PIQ2023	–	15348	authentic	around 1080p	JOD (2AFC)	[−7.28, 5.14]	[12.44, 85.89]
SPAQ	–	11125	authentic	around 720p	MOS (SS)	[2, 96]	[8.94, 88.20]
UHD-IQA	–	6073	authentic	around UHD	MOS (scaled)	[0.29, 0.89]	[24.63, 94.76]

- † WQI x1: one distorted image per reference (generate_dir.sh REFS OUT 1). 19787 candidates passed the similarity filter before the 1080p preparation (see doc/design.typ “Dataset Filtering”); the reference count here is the number actually distorted, i.e. the size of the scored x1 table. Each image carries 1–5 distortion groups drawn from the 33 traditional ops (colour, noise, blur, resolution and geometry buckets) plus the 6 neural ops, every group closed by one codec step (5 codecs or none), so the number of type combinations is combinatorial rather than a fixed list.
- **TID2013**: 24 types officially; our CSV merges codes 21 and 24 into Multiply_Dist. and so shows 23.
- **LIVE MD**: blur+JPEG and blur+noise pairs; **MDIVL**: blur+JPEG and noise+JPEG pairs. Our CSVs collapse each to Multiply_Dist.
- **MDID2013**: blur, then JPEG, then noise on every image; **MDID2016**: all five types on every image at random levels.

- [PIPAL](#): 116 type-level combinations, encoded in the file names; our CSV collapses them to PIPAL_distortions.
- [FLIVE](#): 159240 rows = 39810 full images + 119430 patches.
- [Toyama](#): the score table has 196 rows, of which 28 are pristine copies of the references (Reference), leaving 168 distorted (84 JPEG + 84 JP2K).
- [TID2008](#), [SDIVL](#), [Toyama](#) and [UHD-IQA](#) are scored ([datasets/iqa_scored/combined/](#)) but not yet wired into the fusion test configs of [fusion/options/](#). The cache-warming toy set CacheWarm (4 images) is omitted from the table.
- Label protocols, in brief: ACR / SS = single-stimulus category or slider rating; DS = double stimulus; SS-HR = single stimulus with hidden reference; DSCQS = double-stimulus continuous quality scale (LIVE’s session-realigned DMOS); Swiss 2AFC = TID’s nine-round Swiss-system pairwise votes, averaged to MOS on a nominal [0, 9]; pairwise = pair-comparison sorting (MDID2016, three-way better/worse/equal) or CSIQ’s side-by-side displacement; 2AFC Elo / JOD = forced-choice pairwise aggregated by Elo (PIPAL, init. 1400) or psychometric JOD (PIQ2023, per-scene, not a global scale). DR-ACR is CID2013’s dynamic-reference ACR with a later cluster realignment. “Scaled” means the released table is min-max or [0, 1]-normalized after a wider raw rating (BIQ2021 ACR 1–5; UHD-IQA slider [1, 100]; GFIQA likewise).
- LIVE DMOS can fall slightly outside [0, 100] after z -alignment across sessions; CID2013 MOS can be slightly negative after cluster realignment. [MDID2013](#) is entirely triple-distorted, so its DMOS occupies only a slice of the nominal [0, 1]. [Toyama](#)’s original-range row includes the 28 pristine copies (MOS near 5).

- [PIQ2023](#)’s score for one image is not comparable to another image’s score, unless they are on the same “scene”

The following four sets were not used in the preliminary experiments, but should be included in future tests:

- SDIVL
- TID2008
- Toyama
- UHD-IQA

4 Distortion types of KADID-10k, TID2008 and TID2013

The three laboratory sets share the same design (one labelled synthetic type per image, several levels). The table groups types by family and aligns equivalent types across sets; ✓ means the set contains that type (CSV dis_type names in parentheses when they differ from the row title). KADID-10k has 25 types, TID2008 17, TID2013 24.

TID2013's CSV merges official types 21 and 24 into Multiply_Dist.. They are recovered from the filename iXX_YY_Z (YY = 21 or 24): 21 is lossy compression of noisy images, 24 is sparse sampling and reconstruction. TID's impulse-noise label is misspelt Impluse_Noise in the CSV.

Category	Distortion	KADID-10k	TID2008	TID2013
Noise	Additive white Gaussian noise AWGN	✓	✓	✓
	Additive noise in colour components AWGN_YCbCr / Chroma_Noise	✓	✓	✓
	Spatially correlated (pink) noise Pink_Noise		✓	✓
	Masked noise Masked_Noise		✓	✓
	High-frequency noise High_Frequency_Noise		✓	✓
	Impulse noise Impulse_Noise / Impluse_Noise	✓	✓	✓
	Quantization / quantization noise Quantization / Quantization_Noise	✓	✓	✓
	Multiplicative (speckle) noise Multiplicative_Noise / Multi_Gaussian_Noise	✓		✓
	Comfort noise Comfort_Noise			✓
Blur	Gaussian blur GBLUR	✓	✓	✓
	Lens blur Lens_Blur	✓		
	Motion blur Motion_Blur	✓		
Compression artefacts	JPEG JPEG	✓	✓	✓
	JPEG 2000 JP2K	✓	✓	✓
	JPEG transmission errors JPEG_Transmission_Error		✓	✓
	JPEG 2000 transmission errors JP2K_Transmission_Error		✓	✓
Colour	Colour diffusion Color_Diffusion	✓		
	Colour shift Color_Shift	✓		
	Colour quantization (with dither) Color_Quantization / Dithering	✓		✓
	Colour saturation† Saturation_Dist._HSV, Saturation_Dist._Lab / Saturation_Dist.	✓		✓
	Colour block Color_Block	✓		
	Chromatic aberrations Chroma_Aberrations			✓

Category	Distortion	KADID-10k	TID2008	TID2013
Intensity	Contrast change Contrast_Dist. / Contrast_Dist.	✓	✓	✓
	Brighten Brighten	✓		
	Darken Darken	✓		
	Mean shift Mean_Shift	✓	✓	✓
Spatial / structural	Non-eccentricity patch / pattern noise Non-eccentricity_patch / Non_Eccentricity_Pattern_Noise	✓	✓	✓
	Jitter Jitter	✓		
	Pixelate Pixelate	✓		
	Local block-wise distortion Block_Dist.		✓	✓
	Sparse sampling and reconstruction filename *_24_*; CSV Multiply_Dist.			✓
Restoration artefacts	Denoising Denoising	✓	✓	✓
	High sharpen High_Sharpener	✓		
Compound	Lossy compression of noisy images filename *_21_*; CSV Multiply_Dist.			✓

† KADID-10k implements colour saturation as two types, in HSV and in Lab; TID2013 has a single saturation-change type. Aligning them as one row is why the KADID column shows 24 ticks for 25 official types.