

BoilingBench-Multimodal (NED3-017)

License: cc-by-4.0

BoilingBench-Multimodal is a family of research datasets from the NED³ laboratory for machine learning, computer vision, acoustic sensing, and multimodal heat-transfer analysis. The family contains four multimodal pool-boiling datasets, one human-annotated image dataset, one hydrophone-only pool-boiling dataset, and one infrared immersion-cooling dataset.

This folder is a data distribution, not a Python package. The original acquisition files are preserved where redistribution is authorized; the *_Processed directories contain analysis products generated by BoilingLab. Treat every processed target as **derived data** unless its metadata explicitly identifies it as a direct measurement.

Start here

1. Read the dataset-specific README.md in the folder you plan to use.
2. Decide whether your task needs raw data, processed data, or only a lightweight feature table.
3. Inspect the corresponding manifest, summary, time-alignment record, units, and quality flags before training a model.
4. Split by experiment/run or physical condition before creating temporal windows. Never distribute adjacent windows from one run across train, validation, and test.
5. Cite the dataset-specific publication and the archival DOI where one is provided. Also cite the BoilingBench release and the software used to generate derived products.

What the seven datasets contain

Dataset	Physical system	What is measured	What is derived or annotated	Recommended first tools
BoilingBench-1	Ambient, subcooled, flat-Cu pool boiling	High-speed video, temperature, pressure, DC power, hydrophone, AE hits/time/source, AE waveform	BoilingLab heat flux, surface temperature, synchronization, spectra, MEB and hysteresis products	BoilingLab, BubbleID/BubbleID-Flow, SeqReg
BoilingBench-2	Subatmospheric, saturated, flat-Cu pool boiling	High-speed video, temperature, pressure, DC power, hydrophone, AE hits/time/source, AE waveform	BoilingLab heat flux, surface temperature, synchronization, spectra, transition and hysteresis products	BoilingLab, BubbleID/BubbleID-Flow, SeqReg
BoilingBench-3	Ambient, saturated, Cu-foam pool boiling	Video, temperature, hydrophone, microphone, legacy MISTRAS USB AE hit/source files	BoilingLab thermal regression, acoustic features, event screening, clock alignment	BoilingLab, SeqReg, acoustic feature workflows
BoilingBench-4	Ambient, saturated, flat-Cu pool boiling	Video, temperature, hydrophone, microphone, legacy MISTRAS USB AE hit/source files	BoilingLab thermal regression, acoustic features, event screening, clock alignment	BoilingLab, SeqReg, acoustic feature workflows
BoilingBench-5	Human-annotated boiling images	RGB images and human annotation files	Bubble masks, detections, morphology, vapor-area labels; optional model checkpoints	BubbleID, BubbleID-Flow
BoilingBench-6	Ambient, saturated, flat-Cu pool boiling	Hydrophone and temperature	Acoustic-versus-heat-flux processed series	BoilingLab/feature scripts, SeqReg
BoilingBench-7	Closed HFE-7100 and water immersion cooling	IR video/temperature and electrical/thermal logs	IR temperature features and event labels	IR/video tools, SeqReg

BoilingBench-1 and -2 have continuous AE waveform files and waveform-derived products. BoilingBench-3 and -4 use the older MISTRAS USB AE Node: hit/source files are available, but no

continuous .wfs export is present. BoilingBench-5 is an image-annotation resource rather than a synchronized experiment. BoilingBench-6 and -7 are deliberately single-modality tracks.

Directory organization

Each dataset generally contains a raw acquisition directory and a processed directory:

```
BoilingBench-N_.../
├── README.md
├── *_Raw/           # original acquisition files; measured/source data
└── *_Processed/    # BoilingLab or other derived products
```

Typical raw files include .lvm LabVIEW measurements, .DTA/.TXT AE exports, .wfs waveform streams, .cine Phantom recordings, video files, image files, and IR/electrical logs. Typical processed files include CSV/NPZ time series, acoustic or AE spectra, spectrograms, event tables, hysteresis workbooks, plots, and JSON metadata.

Which tool should I use?

BoilingLab — thermal-fluid analysis and synchronization

Use <https://github.com/UARK-NED3/BoilingLab> for:

- parsing and aligning temperature, pressure, DC power, hydrophone, microphone, and AE data;
- linear-regression thermal reconstruction of surface temperature and heat flux;
- heat-transfer coefficient and wall-superheat calculations;
- critical-event and transition screening;
- hydrophone/microphone spectral features and spectrograms;
- hysteresis analysis when a valid heat-load path and shutoff/DC-power trace exist;
- exporting analysis tables, plots, and machine-readable metadata.

BoilingLab outputs are not automatically validated ground truth. Check thermocouple locations, surface extrapolation distance, material/effective conductivity, pressure source, calibration, sign convention, and the R^2 /quality fields before using heat flux as a label.

BubbleID — bubble interface analysis

Use <https://github.com/cldunlap73/BubbleID> for image-based bubble detection, segmentation, interface geometry, and frame-level bubble statistics. It is most directly applicable to BoilingBench-5 and to video frames extracted from BoilingBench-1 through -4.

BubbleID-Flow — vapor-area and flow-boiling image analysis

Use <https://github.com/UARK-NED3/BubbleID-Flow> for machine-vision quantification of projected vapor area fraction and related image-derived features. Keep human annotations, model predictions, and manually corrected labels in separate files or columns.

SeqReg — sequence regression and temporal prediction

Use <https://github.com/cldunlap73/SeqReg> for sequence-to-value or sequence-to-sequence regression using thermal, acoustic, AE-feature, image-derived, or multimodal time histories. For a credible evaluation, fit normalization and feature-selection parameters on the training runs only and report run-held-out performance.

Waveform decoders and general-purpose tools

For BoilingBench-1 and -2 continuous AE files, use the decoder documented with the release and preserve waveform metadata, sampling rate, channel, trigger time, and calibration. For BoilingBench-3

and -4, do not claim continuous AE waveform analysis: only hit-level/source-file analysis is available. Standard Python tools such as NumPy, pandas, SciPy, scikit-learn, PyTorch, and Hugging Face Datasets can consume the exported CSV/NPZ/Parquet products.

Suggested workflows

Acoustic heat-flux regression

Use BoilingBench-1, -2, -3, -4, or -6. Start with hydrophone/microphone or AE-hit features, align them to the temperature reference clock, and predict a clearly identified derived heat-flux target. Compare persistence, mean, and linear baselines before neural sequence models.

Multimodal fusion

Use BoilingBench-1 or -2 first because they contain the richest synchronized modalities. Add BoilingBench-3 and -4 only after accounting for missing pressure/DC-power channels, legacy AE acquisition, Cu-foam versus flat-Cu geometry, and calibration differences.

Bubble segmentation and vapor-area fraction

Use BoilingBench-5 with BubbleID or BubbleID-Flow. Keep image-level splits grouped by experiment/facility/source sequence. Do not randomly split near-duplicate frames if they share the same acquisition sequence.

Cross-domain generalization

Use BoilingBench-3 versus -4 for Cu-foam/flat-Cu transfer, BoilingBench-1 versus -2 for pressure/regime transfer, and BoilingBench-7 for fluid/device IR transfer. Report the domain shift explicitly; a random-window score is not evidence of cross-domain validity.

Data status and scientific limitations

- **Measured versus derived:** raw sensor/video/image files are acquisition data; heat flux, surface temperature, event times, spectral summaries, MEB indicators, and hysteresis metrics are derived unless stated otherwise.
- **BoilingBench-3 and -4:** current packages lack pressure and DC-power traces. Constant-pressure fallbacks and provisional thermocouple ordering may be present in processed outputs. These heat-flux values are screening-level pending geometry and calibration confirmation.
- **Cu foam:** do not use flat-copper conductivity or surface extrapolation assumptions without checking their applicability to the foam structure.
- **AE coverage:** BoilingBench-3 and -4 have no continuous AE waveform export or AE-waveform spectrogram.
- **BoilingBench-5:** checkpoint files may have different licensing from images and annotations.
- **BoilingBench-6:** raw files and processed files use a simpler legacy layout; inspect the local card before automated ingestion.
- **BoilingBench-7:** this is an IR immersion-cooling dataset, not a conventional multimodal pool-boiling episode.
- **Clock alignment:** recorded clock offsets do not prove zero trigger latency, sensor response delay, or absence of clock drift.

Machine-learning hygiene

At minimum, retain the following fields with every training example: `dataset_id`, `run_id`, `source_file`, `processing_version`, `time_reference`, `modality`, `sampling_rate`, `target_name`, `target_units`, `target_status`, and `quality_flag`.

Use grouped splits by run, specimen, surface, fluid, pressure, facility, or heat-load path as appropriate. Do not let processed copies, duplicate frames, adjacent windows, or waveform caches from one run cross partitions. Report the split type and the number of independent runs—not only the number of windows.

Distribution profiles

The full archive may be large. A Lite profile can omit original `.cine` recordings while retaining compressed videos and tabular/acoustic products. An Ultra-Lite profile can additionally omit processed waveform `.npy` caches while retaining waveform metadata and, where authorized, raw `.wfs` files. Every reduced profile must include an omission manifest and point to the complete archival version.

Citation and licensing

Cite the relevant dataset-specific publication(s), the source archival DOI where provided, and the exact BoilingBench release/profile. The following references identify the current provenance:

- BoilingBench-1: Prince, Hossain, Pierson, and Hu, “Coupled Thermal and Acoustic Signatures of Slow Modulation in Microbubble Emission Boiling,” manuscript in preparation.
- BoilingBench-2: Hossain, Prince, Pierson, Dunlap, Parjuli, and Hu, “Pressure-Adjusted Post-CHF Apparent Base-Wall Superheat Correlates with Boiling Hysteresis in Subatmospheric Pool Boiling,” manuscript in preparation.
- BoilingBench-3/-4: Pandey, Li, Dunlap, and Hu, “Unveiling Hysteresis of Transient Boiling: A Multimodal Perspective,” *Applied Thermal Engineering*, 262, 125259 (2025); Dryad DOI <https://doi.org/10.5061/dryad.ksn02v7h2>.
- BoilingBench-5: Dunlap et al., “BubbleID: A Deep Learning Framework for Bubble Interface Dynamics Analysis,” *Journal of Applied Physics*, 136, 014902 (2024); and Fahim et al., “BubbleID-Flow,” manuscript in preparation.
- BoilingBench-6: Dunlap, Pandey, Weems, and Hu, “Nonintrusive Heat Flux Quantification Using Acoustic Emissions During Pool Boiling,” *Applied Thermal Engineering*, 228, 120558 (2023); Dryad DOI <https://doi.org/10.5061/dryad.q573n5tvq>.
- BoilingBench-7: Pandey et al., “Two-Phase Immersion Cooler for Medium-Voltage Silicon Carbide MOSFETs,” 2024 IEEE ITherm Conference; Dryad DOI <https://doi.org/10.5061/dryad.k98sf7mk9>.

The final dataset license is subject to a rights audit covering raw files, derived products, human annotations, and model checkpoints. Do not assume that the license of BoilingLab, BubbleID, BubbleID-Flow, or SeqReg automatically licenses the data. Preserve any third-party or archival terms that are more restrictive.

Related resources

- Benchmark control plane: <https://github.com/UARK-NED3/BoilingBench-Multimodal>
- Thermal-fluid processing: <https://github.com/UARK-NED3/BoilingLab>
- Bubble interface analysis: <https://github.com/cldunlap73/BubbleID>
- Flow-boiling vapor-area analysis: <https://github.com/UARK-NED3/BubbleID-Flow>
- Temporal regression: <https://github.com/cldunlap73/SeqReg>