Supplementary Information for

Predicting Moisture-Driven Resistivity Evolution Using a CLIP-Inspired Dual-Encoder Framework with Cave Microclimate Conditioning

Xitao Wang[1], Qingsheng Wen[1], Zeyun Gao[1], Huakang Bian[1,*], Shunbo Hu[1,2,*], Jizhong Huang[1,2]

[1]Material Genome Institute, Institute for the Conservation of Cultural Heritage, Institute for Quantum Science and Technology, Shanghai University, Shanghai 200444, China.

[2]Key Laboratory of Silicate Cultural Heritage Conservation (Shanghai University), Ministry of Education, Shanghai 200444, China.

[*]Corresponding authors: bianhuakang@shu.edu.cn; shunbohu@shu.edu.cn

This Supplementary Information provides additional material supporting the main manuscript: split-wise cave microclimate statistics, deployment-oriented runtime profiling, and a controlled condition-responsive explainability example.

S1. Split-wise cave microclimate statistics

Variable	Train ($n = 1173$)	Validation ($n = 251$)	Test ($n = 253$)
Northeast RH (%)	70.40 ± 16.97	81.78 ± 10.59	84.03 ± 16.20
Northeast air temp. (deg C)	12.45 ± 3.20	14.03 ± 0.92	15.16 ± 0.72
Northeast surface temp. (deg C)	11.11 ± 2.73	12.12 ± 0.47	13.09 ± 0.42
Northwest air temp. (deg C)	12.52 ± 3.08	14.06 ± 1.02	15.24 ± 0.78
Northwest surface temp. (deg C)	13.21 ± 2.14	14.31 ± 0.42	15.23 ± 0.42

Table 1: Supplementary Table S1. Split-wise cave microclimate statistics for the corrected chronological protocol. Values are reported as mean ± standard deviation over the synchronized 1,677-sample pool.

Under the corrected chronological protocol, the held-out block is not merely a random subset of the synchronized triplets; it occupies a clearly warmer and more humid microclimate regime than the earlier training block. Relative humidity rises from 70.4% in the training subset to 84.0% in the test subset, while mean air temperature increases from 12.5 deg C to 15.2 deg C at the northeast sensor and from 12.5 deg C to 15.2 deg C at the northwest sensor. This shift supports the main-text interpretation that chronological evaluation exposes a genuine seasonal domain shift rather than a mere reduction in sample count.

S2. Runtime and memory profile

Profiling the retained CLIPWithVecFusion checkpoint shows that deployment is dominated by the conditioning-side forward pass rather than by gallery search. On the profiled hardware, building the 253-item candidate cache from preloaded tensors took 0.167 s in total, while a single conditioned query had a median end-to-end latency of 2.66 ms (mean 3.07 ms; p95 2.69 ms). Of that budget, the current-state encoder and fusion block consumed a median of 2.56 ms, whereas the similarity scan over the cached gallery remained at 0.10 ms median. This supports the manuscript claim that the retrieval formulation is operationally feasible once future-difference embeddings are precomputed.

Device	NVIDIA GeForce RTX 4090 (cuda)
Gallery size	253
Gallery precompute	0.167 s total (0.66 ms/candidate)
Single-query encode	mean 2.56 ms; median 2.56 ms
Similarity scan	mean 0.50 ms; median 0.10 ms
End-to-end query	mean 3.07 ms; median 2.66 ms; p95 2.69 ms
Query throughput	326.09 queries/s
Model parameters	36,569,193
Parameter memory	139.50 MB
Cached gallery embedding memory	0.49 MB
Observed peak memory	938.08 MB

Table 2: Supplementary Table S2. Deployment-oriented runtime profile for the retained CLIP-WithVecFusion checkpoint under the corrected chronological protocol. Both gallery-precompute and query timings are measured on preloaded tensors, excluding disk I/O and assuming a precomputed gallery cache during online retrieval.

S3. Condition-responsive explainability

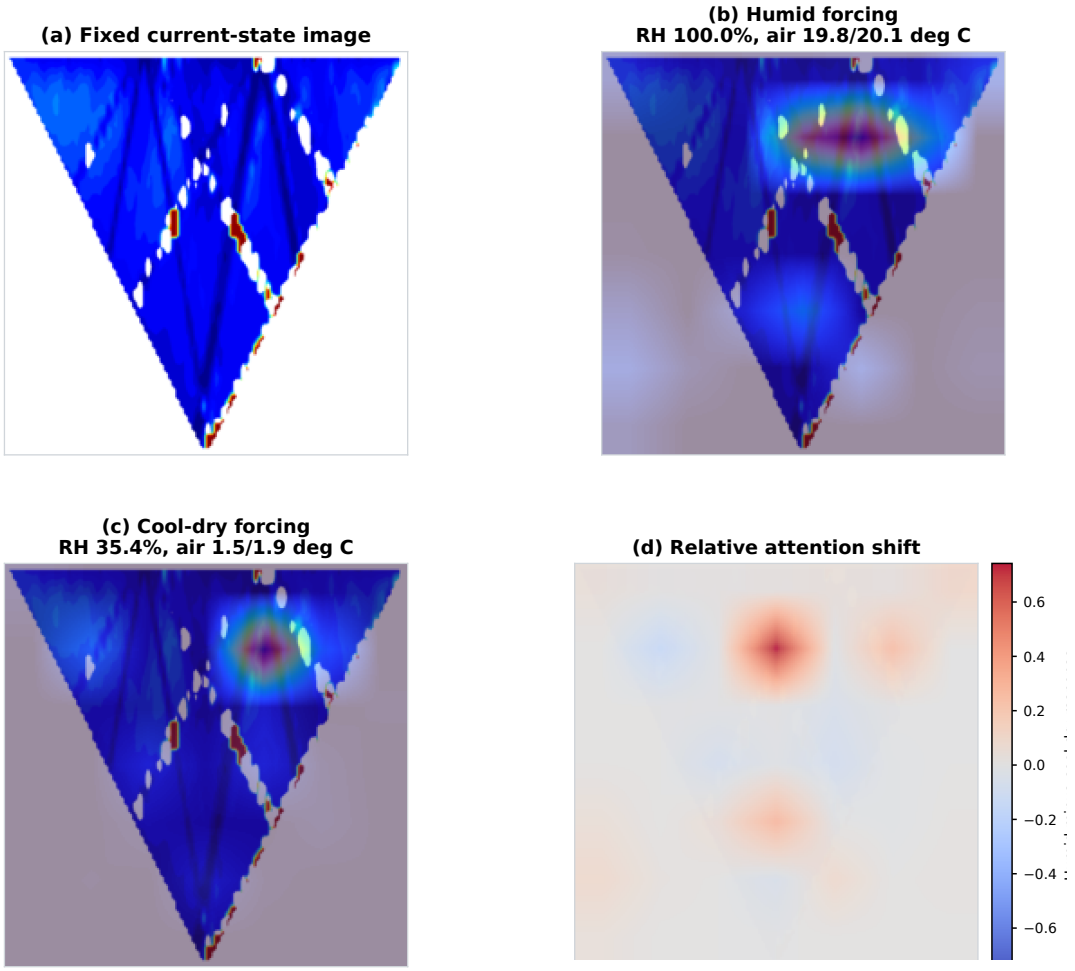

Figure 1: Supplementary Figure S1. Controlled condition-responsive activation comparison for the same current-state HDRM image. (a) Fixed current-state query image. (b) Normalized qualitative activation overlay under a humid forcing vector (RH 100.0%, air 19.8/20.1 deg C). (c) Normalized qualitative activation overlay under a cool-dry forcing vector (RH 35.4%, air 1.5/1.9 deg C). (d) Relative attention shift obtained by subtracting the cool-dry map from the humid map, highlighting the upper-central region that becomes more salient under wetter forcing.

To test whether the conditioning vector changes the spatial evidence used by the model rather than merely rescaling the final score, we fixed one current-state HDRM image and queried the network under two contrasting but physically observed cave microclimate vectors. Under the humid forcing (RH 100.0%, air 19.8/20.1 deg C), the activation expands into a broader shallow band and the retrieved match score rises to 14.30. Under the cool-dry forcing (RH 35.4%, air 1.5/1.9 deg C), the map contracts into a more localized hotspot and the top retrieval score drops to 13.33. The humid-minus-cool-dry difference map therefore visualizes a genuine condition-driven attention shift, supporting the interpretation that cave microclimate measurements actively steer where the model searches for moisture-sensitive resistivity evolution patterns within the same present-state image.