MATCH: Multimodal Alignment Through Contrastive Learning of Heterogeneous Sensor Streams at Static Nodes

Julia Gersey[1,2], Nicola Ferrier[2,3]

[1]University of Michigan, Ann Arbor, MI USA
[2]Argonne National Laboratory, Lemont, IL USA
[3]Northwestern University, Evanston, IL USA

What is Sage?

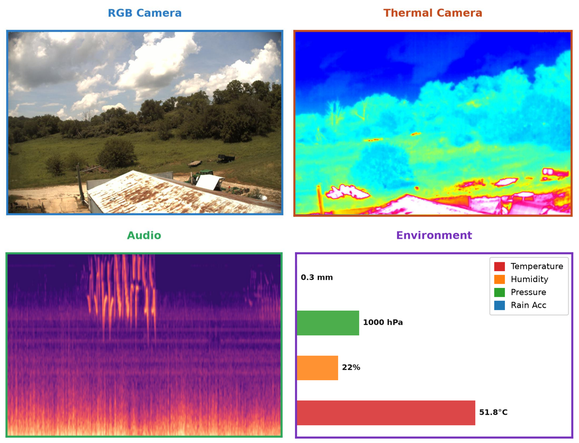
- NSF-funded open AI testbed for edge computing & intelligent sensing; **300+ nodes across all 50 states**
- Multimodal nodes: GPU compute + RGB, thermal, microphone, air quality & more!



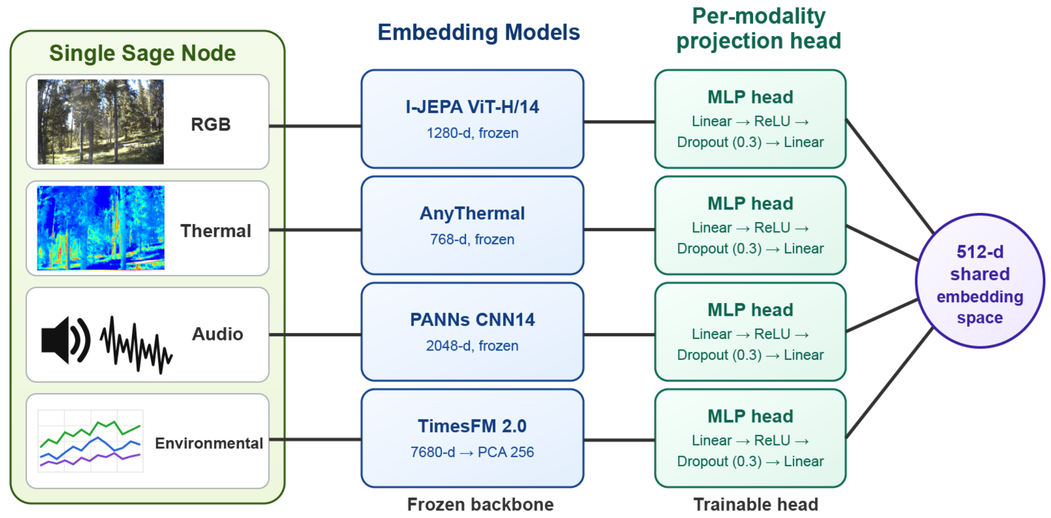
Sage Node Installation (W069) in Hawaii.

MATCH Architecture

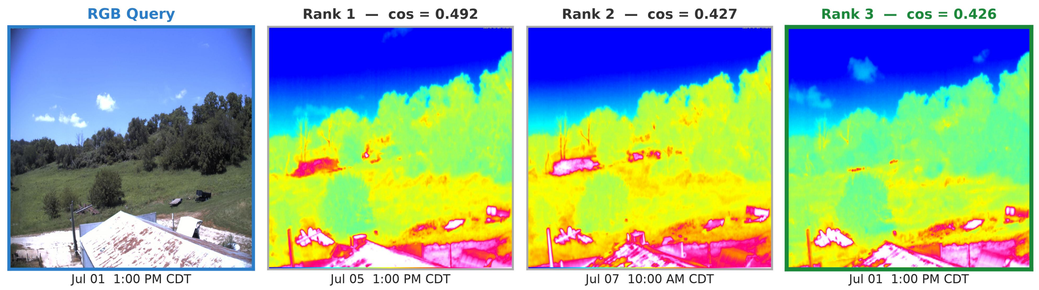
- *MATCH* learns a shared 512-d embedding space with no task-specific labels
- Frozen, domain-matched backbones (I-JEPA [1], AnyThermal [2], PANNs [3], TimesFM [4])
- Trainable MLP projection heads map all four into shared space via multi-view InfoNCE



Data samples from each modality.



Contrastive Learning for Multimodal Alignment Results



RGB query embedding retrieving nearest-neighbor thermal matches, ranked by cosine similarity.

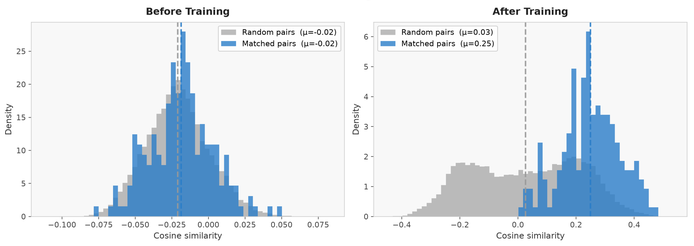
- All six modality pairs retrieve above random chance, RGB ↔ Thermal strongest
- 73.3% of errors land within ±2 hrs of correct time-of-day (vs. 20.4% chance); model learns diurnal structure but confuses similar moments across days

Pair	z-score	Mean rank	Hub max rate
RGB ↔ Thermal	1.212	28.8	0.221
RGB ↔ Audio	1.017	39.7	0.308
RGB ↔ Env	1.082	36.1	0.262
Thermal ↔ Audio	0.846	49.4	0.349
Thermal ↔ Env	1.014	39.9	0.256
Audio ↔ Env	0.780	53.1	0.292

Retrieval performance across modalities. Lower mean rank and z-score = stronger alignment.

Evidence of Learned Alignment

- Cosine similarity between matched pairs separates clearly from random after training



Current & Future Work

- Node-specific adaptation for deployment across multiple Sage nodes
- Downstream applications: anomaly detection, cross-modal sensor imputation
- Implement edge embedding and federated retrieval for network queries

Acknowledgements

Supported by NSF-OAC #2436842 and the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Award Number DE-SC0026073.

Learn more about Sage:

sagecontinuum.org

References

[1] M. Assran et al., "Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture," in Proc. IEEE/CVF CVPR, Vancouver, BC, Canada, Jun. 2023, pp. 15619–15629.

[2] P. Maheshwari et al., "AnyThermal: Towards Learning Universal Representations for Thermal Perception," arXiv:2602.06203, Feb. 2026.

[3] Q. Kong et al., "PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition," arXiv:1912.10211, Dec. 2019.

[4] A. Das et al., "A Decoder-Only Foundation Model for Time-Series Forecasting," arXiv:2310.10688, Apr. 2024.