Learning Shared Representations from Unpaired Data

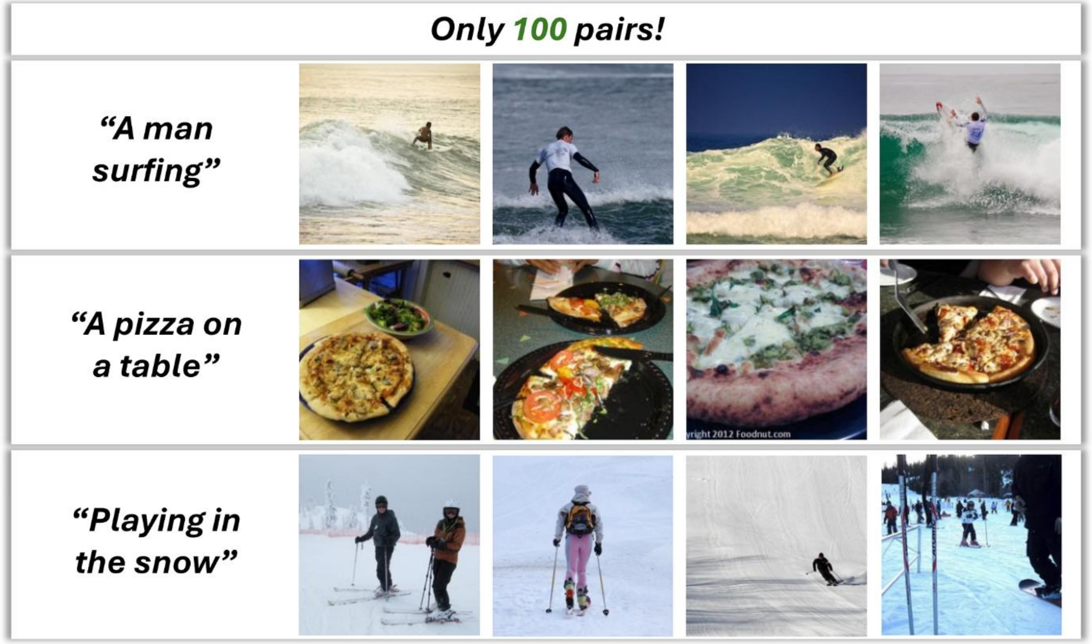
Amitai Yacobi*, Nir Ben-Ari*, Ronen Talmon, and Uri Shaham

*Equal contribution

NEURAL INFORMATION PROCESSING SYSTEMS

Bar-Ilan University
אוניברסיטת בר-אילן

TECHNION
Israel Institute of Technology



Only 100 pairs!

“A man surfing”

“A pizza on a table”

“Playing in the snow”

The Challenge of Multimodal Learning

Paired Data Dependency

Current multimodal models heavily rely on vast amounts of paired data (e.g., 400M image-caption pairs for CLIP).

Costly Acquisition

Obtaining paired data is often expensive, difficult, and sometimes impossible, especially in specialized domains like medical imaging.

Abundant Unpaired Data

Unpaired data is significantly more accessible and available, yet underutilized in shared representation learning.

Key Insight: Universality

Modality-Invariant Structure

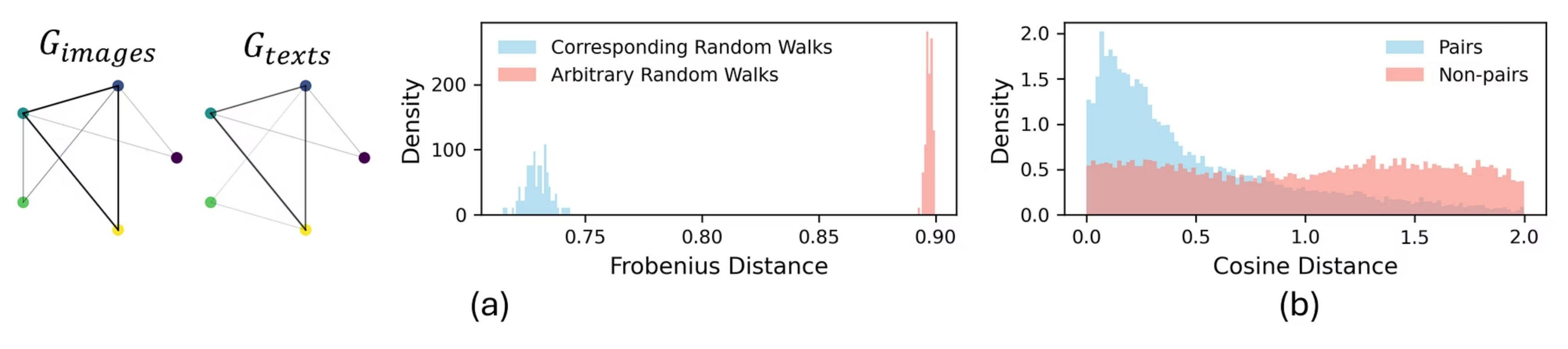
This universal structure is independent of the specific data modality, allowing for seamless comparisons between images, text, and other data types.

Random Walk Processes

The shared structure is captured by analyzing intrinsic geometric properties through random walk processes on each modality's high-dimensional data manifold.

Spectral Embedding

The shared structure is then extracted and represented in a low-dimensional space using spectral embedding techniques, revealing inherent relationships across modalities.

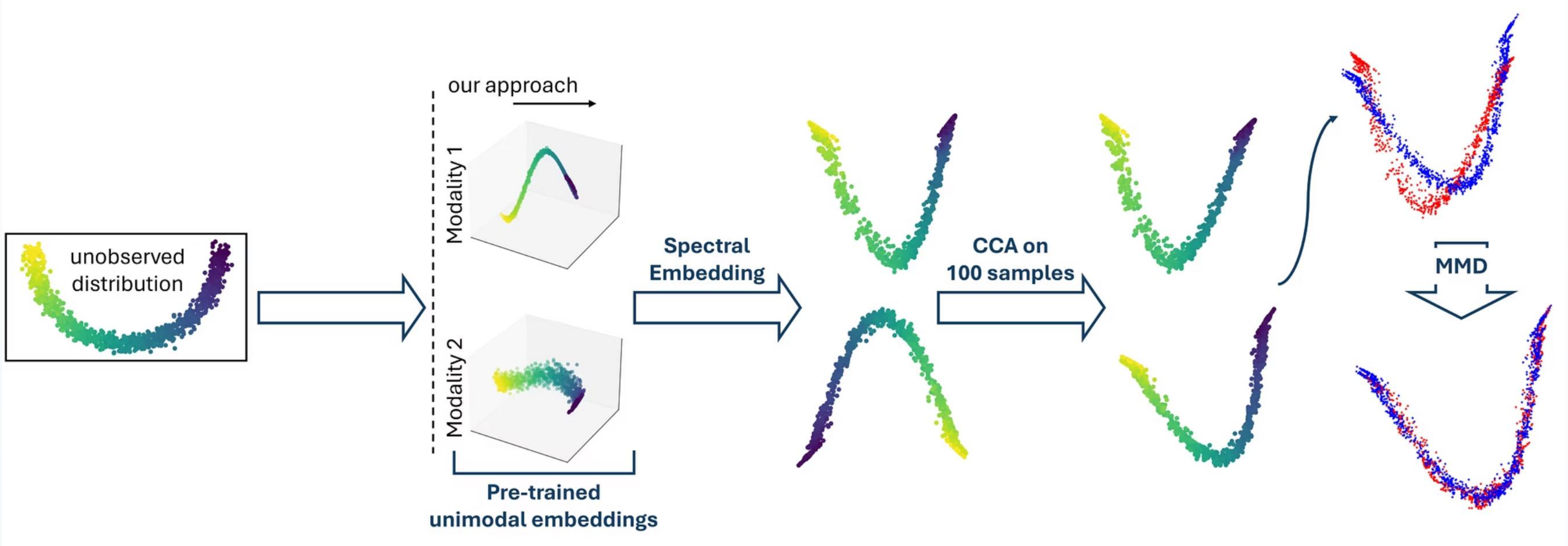


(a)

(b)

Spectral Universal Embedding (SUE)

SUE demonstrates that shared representations can be learned almost exclusively from unpaired data, challenging conventional approaches.



Spectral Embedding (SE)

Extracts universal features by mapping unimodal embeddings into corresponding eigenspaces, capturing global structure.

Canonical Correlation Analysis (CCA)

Aligns features using a minimal number of paired samples to resolve SE ambiguity.

MMD-net

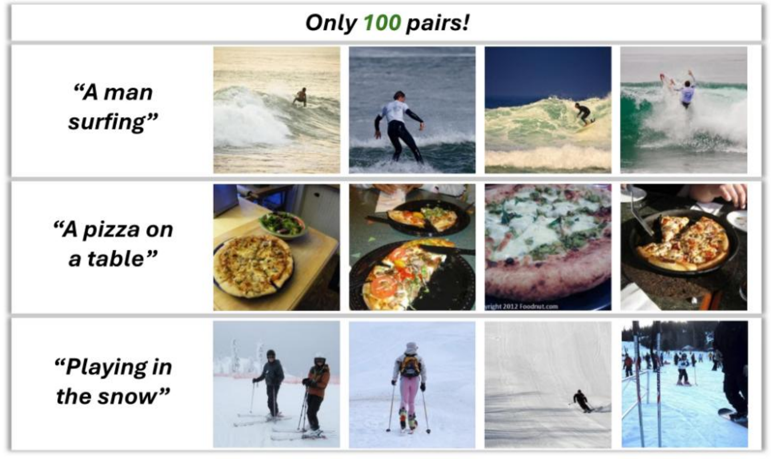
Provides additional non-linear alignment by minimizing Maximum Mean Discrepancy between distributions.

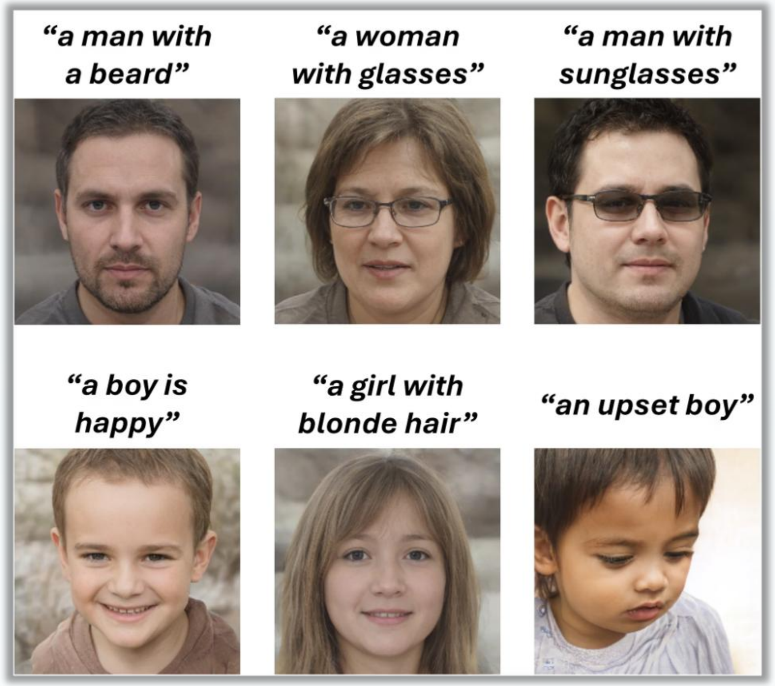
SUE's Empirical Superiority in Retrieval

SUE significantly outperforms contrastive learning and other baselines in retrieval tasks, when only limited paired data is avialble.

#Pairs	Dataset	**SUE**	Contrastive	CSA	Improvement
		R@10	R@10	R@10	(over best)
100	MSCOCO I2T	**34.25**	13.00	3.00	+257.20%
500	Flickr30k I2T	**32.00**	16.20	2.50	+103.32%
50	Edges2Shoes E2S	**25.25**	14.00	2.25	+200.51%

For MSCOCO, SUE achieves over 250% better performance than contrastive methods using the same minimal number of pairs.





Versatile Applications of SUE

Text-to-Image Generation

Generates semantically aligned images from text queries with minimal text-image correspondence.

Semantic Arithmetics

Performs intuitive vector summations of text and image embeddings to create new images.

Zero-Shot Classification

Acts as a classifier for unseen categories without explicit classification training.

Cross-Domain Classification

Enables knowledge transfer between labeled and unlabeled domains with minimal paired data.

The Hidden Power of Unpaired Data

SUE redefines efficiency by leveraging inherent semantic similarities. This innovative approach allows SUE to achieve remarkable results with significantly less data, transforming the landscape of AI training.

Reduced Paired Data Requirement

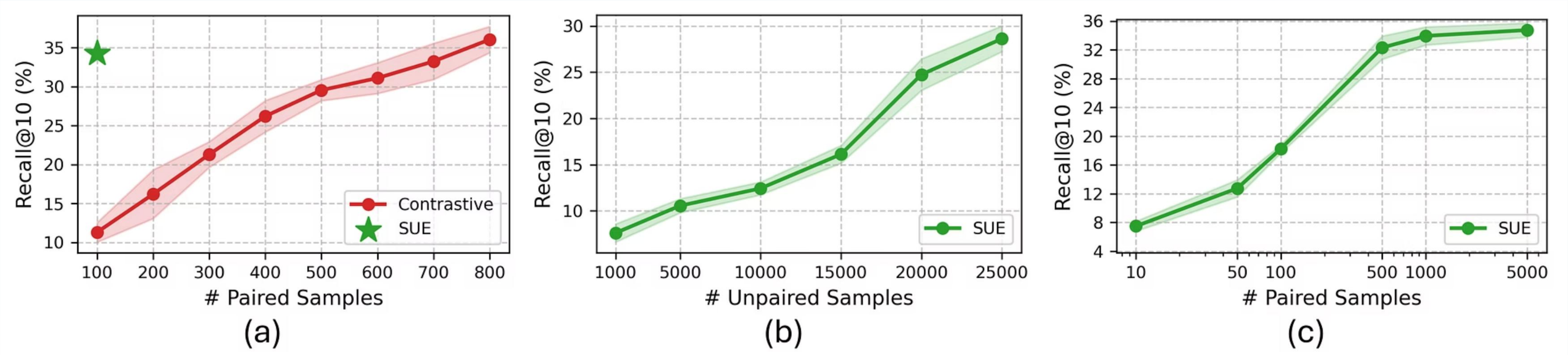
SUE achieves comparable performance using an astonishing **10 times less paired data** than traditional methods, critically beneficial when such data is scarce.

Leveraging Abundant Unpaired Data

SUE thrives on **abundant unpaired data**, drastically improving overall performance and minimizing reliance on expensive, hard-to-find paired datasets.

Treuly relying on unpaired data

Its independence from costly paired data makes SUE a more **robust and highly scalable** solution for a wide array of applications, much less affected by paired data.



Conclusions and Future

Unlocking Unpaired Data

SUE demonstrates the ability to learn shared representations almost exclusively from unpaired data.

Universal Embedding

Spectral Embedding (SE) is key to uncovering modality-invariant properties.

Future Directions

Extending to more complex modalities and achieving performance comparable to massively paired models.

This work lays a foundation for a fully unpaired multimodal learning framework, opening new avenues for research.

Thank You!

We appreciate your attention and hope SUE inspires new directions in multimodal learning.

Contact Us

amitaiyacobi@gmail.com

nirnirba@gmail.com

Project Page

Explore the code and examples.

GitHub Repo

Read the Paper

Dive into the full research.

arXiv Preprint