Latent Geometry Alignment of Independently Trained Sensor and Text Representations

Rusham Bhatt
University of Maryland Baltimore County
rbhatt4@umbc.edu

KMA Solaiman *
University of Maryland Baltimore County
ksolaima@umbc.edu

Abstract

Understanding the geometry of latent representations is central to interpreting and evaluating modern learning systems. While most multimodal approaches rely on joint training to enforce shared embedding spaces, a less explored question is whether independently trained models already encode compatible latent geometric structure. In this work, we study representation alignment between sensor time-series embeddings and textual embeddings as a controlled testbed for analyzing cross-modal latent geometry. Using the UCI Human Activity Recognition dataset, we embed sensor windows with TS2Vec and activity labels with Sentence-BERT, and analyze alignment at the class level using cosine similarity and orthogonal Procrustes analysis. We observe weak raw alignment between modalities, but find that a simple orthogonal transformation reveals strong geometric correspondence, with same-class embeddings becoming highly aligned while preserving intra-space relational structure. These results suggest that independently learned representations can exhibit shared semantic manifolds up to rigid transformations, supporting the manifold hypothesis across modalities. Our findings highlight the value of geometry-aware diagnostics for probing latent spaces and are directly applicable to broader multimodal settings, including vision–language and perception systems, where independently trained encoders are increasingly common.

1. Introduction

Understanding the geometry of latent representations is fundamental to interpreting, evaluating, and designing modern learning systems. Although high-dimensional data such as sensor signals, images, or text reside in ambient spaces with enormous dimensionality, their learned representations are often hypothesized to lie on lower-dimensional semantic manifolds whose structure governs generalization, robustness, and interpretability. Recent advances in self-supervised and pretrained models have made it possible to study such latent spaces directly; however, most multimodal approaches enforce shared representations through joint training and paired data [2, 6].

While effective, joint-training paradigms conflate representational geometry with training objectives and supervision, making it difficult to disentangle whether observed alignment arises from shared semantic structure or from explicit coupling during learning. As a result, they offer limited insight into the intrinsic organization of embedding spaces learned independently within each modality.

Representation alignment provides an alternative, geometry-driven perspective. Rather than enforcing correspondence through joint optimization, alignment analysis treats pretrained embedding spaces as fixed and asks whether their internal relational structure is compatible up to geometry-preserving transformations [5, 7]. If two embedding spaces encode similar semantic relationships, such as relative distances or angles between concepts – they may be viewed as different coordinate systems over a shared latent manifold. Under this view, alignment is not a predictive objective but a diagnostic tool for probing latent geometry across models and modalities.

In this work, we study representation alignment between sensor-based time-series embeddings and textual embeddings as a controlled setting for analyzing cross-modal latent geometry. Sensor signals and natural language differ substantially in structure and inductive biases, yet both are trained to capture semantic distinctions between human activities. We ask whether embeddings learned independently from these modalities encode compatible class-level geometry in the absence of paired supervision or joint optimization.

Using the UCI Human Activity Recognition dataset, we construct class-level representations for each modality by embedding sensor windows with TS2Vec and activity labels with Sentence-BERT. We evaluate alignment using cosine similarity and orthogonal Procrustes analysis, which identifies the optimal rigid transformation aligning two embed-

*Corresponding Author

ding spaces while preserving their internal geometry. Our analysis reveals that while raw cross-modal alignment is weak, a simple orthogonal transformation uncovers strong geometric correspondence at the class level, supporting the view that independently trained models can encode similar semantic manifolds that differ primarily in their coordinate systems. More broadly, this work highlights the value of geometry-aware diagnostics for understanding latent representations and offers a methodology applicable to other multimodal settings, including vision–language systems.

2. Related Work

Time-Series and Sensor Representation Learning. Learning general-purpose representations for time-series data has received increasing attention in recent years. TS2Vec [14] proposes a self-supervised framework that learns universal representations by contrasting temporal segments at multiple scales. Unlike supervised activity recognition pipelines, TS2Vec is designed to capture semantic structure without task-specific labels, making it well-suited for representation analysis beyond predictive accuracy. Our work leverages TS2Vec as a pretrained sensor encoder and examines the geometric properties of its learned embedding space rather than its downstream classification performance.

Textual Representation Learning. Sentence-BERT [8] adapts BERT-based architectures to produce semantically meaningful sentence-level embeddings through siamese and triplet training objectives. These embeddings are widely used for similarity and retrieval tasks and have been shown to encode rich semantic relationships. In this work, we use Sentence-BERT to embed textual activity labels, treating the resulting vectors as fixed representations whose geometric structure can be compared to that of sensor embeddings.

Representation Similarity and Geometry. The geometric comparison of learned representations has been explored in several contexts. Kornblith et al. [5] revisit representation similarity analysis and show that neural networks trained independently on the same task can exhibit similar relational structure despite differences in parameters. Earlier work introduced canonical correlation-based techniques such as SVCCA [7] to compare representations across layers and models. Relatedly, probing-based approaches use linear classifiers to assess the information content of intermediate representations [1]. In contrast to predictive probes or correlation-maximizing objectives, our analysis focuses on geometry-preserving alignment as a diagnostic of relational structure across modalities.

Manifold Geometry of Latent Representations. Several works have investigated the geometric and manifold structure of latent representations in deep models. Brahma et al. [4] argue that deep learning success can be understood through progressive manifold disentanglement, while Shukla et al. [11] study the geometry of latent spaces in deep generative models. These works motivate the view of learned representations as structured manifolds, though they do not address cross-modal alignment. Our work adopts this geometric perspective but focuses on comparing independently learned representations across modalities.

Multimodal Representation Learning and Alignment. Most multimodal approaches aim to construct shared embedding spaces through joint training on paired data. Methods such as Deep CCA [3] explicitly learn projections that maximize cross-modal correlation, while large-scale vision–language models such as CLIP [6] and Flamingo [2] align modalities through end-to-end optimization on paired supervision. While highly effective for downstream tasks, such approaches entangle representational geometry with training supervision and architectural coupling.

In contrast, our work does not perform joint training or enforce alignment during learning. Instead, we analyze independently trained encoders and evaluate whether their representations exhibit compatible semantic geometry up to rigid transformations. This perspective is related to analytical views of alignment in representation learning, such as the alignment – uniformity framework of Wang and Isola [13], but differs in that we do not optimize alignment as a learning objective.

3. Problem Formulation: Geometric Representation Alignment

Modern representation learning systems map high-dimensional inputs into latent embedding spaces intended to capture semantic structure. Although embeddings derived from different modalities (e.g., sensor time series and natural language) are learned using distinct architectures, objectives, and data distributions, they may nevertheless encode compatible *relational geometry* if they represent the same underlying concepts. In this work, we formalize *geometric representation alignment* as the problem of determining whether independently trained embedding spaces exhibit shared latent structure up to geometry-preserving transformations.

Let $\mathcal{C} = \{c_1, \ldots, c_K\}$ denote a set of semantic classes. We consider two embedding functions trained independently on different modalities:

$$f_s : \mathcal{X}_s \to \mathbb{R}^d, \qquad f_t : \mathcal{X}_t \to \mathbb{R}^d, \tag{1}$$

where $\mathcal{X}_s$ represents sensor time-series inputs and $\mathcal{X}_t$ represents textual descriptions. Rather than evaluating instance-

level correspondence, we focus on *class-level* latent representations, which better reflect global semantic organization and reduce sensitivity to instance-specific variation.

For each class $c_k \in \mathcal{C}$, we compute a class centroid in each embedding space:

$$\mathbf{s}_k = \mathbb{E}_{x \sim c_k}\big[f_s(x)\big], \qquad \mathbf{t}_k = f_t\big(\text{name}(c_k)\big), \tag{2}$$

yielding sensor and text embedding matrices

$$S = [\mathbf{s}_1, \ldots, \mathbf{s}_K]^\top \in \mathbb{R}^{K \times d}, \tag{3}$$

$$T = [\mathbf{t}_1, \ldots, \mathbf{t}_K]^\top \in \mathbb{R}^{K \times d}. \tag{4}$$

The central question we address is *not* whether these embeddings coincide in a shared coordinate system, but whether they encode similar latent geometry. Specifically, we ask whether the relative relationships between classes (e.g., angles and distances) are preserved across modalities, even if their embedding spaces differ by a rotation or reflection. We define alignment as the existence of an orthogonal transformation that best maps one space onto the other:

$$R^\star = \arg \min_{R \in \mathbb{R}^{d \times d} : R^\top R = I} \|SR - T\|_F \,, \tag{5}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Orthogonal transformations preserve distances and angles within each embedding space, ensuring that alignment reflects compatibility of *relational structure* rather than arbitrary distortion.

Under this formulation, successful alignment suggests that the two embedding spaces act as different coordinate systems over a shared latent semantic manifold, whereas weak alignment indicates that the modalities encode fundamentally different geometric organization. Importantly, this view treats representation alignment as a *geometry-aware diagnostic* rather than a predictive objective: it enables comparison of independently learned representations without joint training or paired supervision. In the remainder of the paper, we instantiate this formulation in a sensor–text setting and quantify alignment using cosine similarity and orthogonal Procrustes analysis.

4. Methodology

4.1. Sensor Representation Learning (TS2Vec)

We encode sensor windows using TS2Vec [14], a self-supervised time-series representation learning framework designed to learn generalizable latent representations without requiring labeled supervision. Each 2.56-second multivariate sensor window is mapped to a fixed-length embedding in $\mathbb{R}^{384}$, capturing temporal and cross-channel structure in the input signals.

4.2. Text Representation Learning (Sentence-BERT)

We encode textual activity labels (e.g., "walking", "sitting") using Sentence-BERT [8], which produces semantically meaningful sentence-level embeddings suitable for similarity analysis. Each activity label is embedded into the same 384-dimensional latent space, enabling direct geometric comparison with sensor embeddings.

4.3. Class-Level Representation Aggregation

To analyze latent geometry at the semantic level, we operate on class-level representations rather than individual instances. For the sensor modality, we compute a centroid embedding for each activity class by averaging TS2Vec embeddings over all sensor windows belonging to that class. For the text modality, each activity class is represented by the embedding of its corresponding label. This aggregation yields one embedding per class for each modality and reduces sensitivity to instance-level variability, allowing us to focus on global relational structure.

4.4. Geometry-Aware Alignment Metrics

We evaluate representation alignment using geometry-aware similarity measures that probe relational structure rather than pointwise correspondence.

- **Cosine Similarity**: We compute cosine similarity between sensor and text class embeddings to assess raw angular correspondence between modalities prior to alignment.
- **Orthogonal Procrustes Alignment**: To test whether the two embedding spaces encode compatible latent geometry up to isometric transformation, we apply orthogonal Procrustes analysis [12]. This procedure identifies the optimal orthogonal transformation that aligns the sensor embedding space to the text embedding space while preserving distances and angles within each space. Improvements in alignment after this transformation indicate shared relational structure across modalities.

5. Experimental Setup

We design our experiments to evaluate geometric representation alignment under controlled conditions. Rather than benchmarking task performance, our goal is to probe whether independently trained sensor and text embedding spaces exhibit compatible latent structure at the class level, both before and after geometry-preserving alignment.

5.1. Dataset: UCI Human Activity Recognition

We conduct our analysis on the UCI Human Activity Recognition (HAR) dataset [9], which contains multivariate sensor recordings collected from 30 participants performing six daily activities: *walking*, *walking upstairs*, *walking*

downstairs, *sitting*, *standing*, and *laying*. Accelerometer and gyroscope signals are sampled at 50 Hz and segmented into fixed-length windows of 2.56 seconds, yielding a total of 10,299 sensor windows.

The dataset provides a controlled multimodal setting in which semantic activity classes are shared across modalities while the input structures differ substantially. This makes it well suited for studying cross-modal latent geometry without confounding factors introduced by complex annotation schemes or task-specific objectives.

5.2. Representation Construction

5.2.1. Textual Embeddings

For the text modality, each activity class is represented by the Sentence-BERT embedding of its corresponding label. We attain the textual embeddings by utilizing Sentence-BERT's all-MiniLM-L6-v2 model [10], which maps sentences and paragraphs to a 384-dimensional vector space. Because the text modality provides a single embedding per class, no further aggregation is required. Both modalities thus yield one embedding per class in a shared ambient dimensionality, enabling direct geometric comparison.

5.2.2. Sensor Embeddings

When embedding the corresponding sensor signals, TS2Vec provides the capability to specify the dimensionality of the output embeddings. The output dimension was set to 384 to match the Sentence-BERT model used for text embeddings.

Each HAR window contains 128 time steps across 9 channels (accelerometer and gyroscope axes). We reshape each instance to 128×9 as required by TS2Vec. This reshaping preserves the temporal ordering and channel structure required by TS2Vec, enabling the model to learn representations that capture both temporal dynamics and cross-channel relationships. TS2Vec is optimized using a self-supervised hierarchical temporal contrastive loss. We monitor this loss during training and observe convergence after approximately 28 epochs.

5.3. Experimental Questions

Our experiments are designed to answer the following questions:

1. Do independently trained sensor and text embeddings exhibit measurable alignment in their original coordinate systems?
2. Does geometry-preserving alignment reveal shared relational structure between the two embedding spaces?
3. Are embeddings corresponding to the same semantic class more closely aligned than those from different classes after orthogonal transformation?

These questions explicitly target the presence or absence of compatible latent geometry rather than predictive accuracy.

5.4. Evaluation Protocol

We evaluate alignment using cosine similarity and orthogonal Procrustes analysis, as described in Section 4.4. Cosine similarity is computed between all pairs of sensor and text class embeddings to assess raw angular correspondence. For Procrustes alignment, we center both embedding matrices, compute the optimal orthogonal transformation aligning the sensor space to the text space, and re-evaluate cosine similarity in the aligned space. Algorithm 1 summarizes this process.

Algorithm 1 Orthogonal Procrustes Alignment and Similarity Score

Input: Sensor centroid matrix $S \in \mathbb{R}^{C \times d}$, text centroid matrix $T \in \mathbb{R}^{C \times d}$
Output: Alignment score s, rotation matrix $R \in \mathbb{R}^{d \times d}$, aligned embeddings $S_{\text{aligned}} \in \mathbb{R}^{C \times d}$
Center embeddings:
$S_c \leftarrow S - \text{mean}(S),\ T_c \leftarrow T - \text{mean}(T)$
Cross-covariance: $M \leftarrow S_c^\top T_c$
Compute SVD: $(U, \Sigma, V^\top) \leftarrow \text{SVD}(M)$
Compute orthogonal rotation: $R \leftarrow UV^\top$
Align sensor space: $S_{\text{aligned}} \leftarrow S_c R$
Numerator: $n \leftarrow \sum_i \Sigma_{ii}$ (i.e., $\text{trace}(\Sigma)$)
Denominator: $d \leftarrow \|S_c\|_F \cdot \|T_c\|_F$
Alignment score: $s \leftarrow n/d$
Return: s, R, S_{aligned}

To facilitate qualitative inspection of latent geometry, we additionally visualize embeddings before and after alignment using Principal Component Analysis (PCA). PCA is fit once on the combined set of centered sensor and text class embeddings and applied consistently across conditions to ensure that observed differences reflect alignment effects rather than changes in projection.

Together, this experimental setup enables a focused analysis of geometric representation alignment under minimal assumptions, isolating the role of latent structure from training procedures or task-specific objectives.

6. Results and Geometric Analysis

In this section, we analyze representation alignment between sensor and text embeddings from a geometric perspective. Our objective is not to assess predictive performance, but to evaluate whether independently trained embedding spaces encode compatible latent structure at the class level. We report results before and after geometry-preserving alignment and interpret them in terms of relational correspondence between modalities.

Table 1. Mean cosine similarity between TS2Vec sensor class centroids and Sentence-BERT text label embeddings, before and after Procrustes alignment.

SETTING	SAME-CLASS MEAN	DIFF-CLASS MEAN
RAW	0.028	0.026
ALIGNED	≈ 0.94	≈ -0.19

6.1. Raw Cross-Modal Alignment

We first examine raw alignment between sensor and text class embeddings using cosine similarity, without applying any transformation to either embedding space. This analysis probes whether the two modalities share a common orientation in latent space.

As summarized in Table 1, raw cosine similarities are close to zero for both same-class and different-class pairs. The mean cosine similarity for same-class pairs is 0.028, while the mean similarity for different-class pairs is 0.026. The absence of separation between these distributions indicates that sensor and text embeddings are not directly aligned in their original coordinate systems.

From a geometric standpoint, this result suggests that although each embedding space may encode meaningful internal structure, their coordinate axes are misaligned. Importantly, near-zero cosine similarity does not imply an absence of shared semantic organization, but rather reflects a mismatch in orientation between the two latent spaces.

6.2. Alignment After Orthogonal Transformation

We next apply orthogonal Procrustes alignment to the sensor embedding space and re-evaluate cosine similarity. Because orthogonal transformations preserve distances and angles within each embedding space, improvements in alignment reflect compatibility of relational geometry rather than arbitrary distortion.

After alignment, we observe a substantial increase in same-class similarity, with mean cosine similarity rising to approximately 0.94. In contrast, the mean similarity for different-class pairs becomes strongly negative (approximately -0.19). This pronounced separation indicates that, once coordinate systems are aligned, embeddings corresponding to the same activity class occupy closely aligned directions in latent space, while embeddings from different classes are well separated.

These results provide strong evidence that the two modalities encode similar class-level relational structure up to a rigid transformation. In other words, the latent spaces appear to represent comparable semantic organization, differing primarily in their coordinate systems rather than their intrinsic geometry.

Algorithm 2 PCA-Based Projection for Sensor–Text Embedding Visualization

Input: Sensor centroids $S \in \mathbb{R}^{C\times d}$, text embeddings $T \in \mathbb{R}^{C\times d}$, aligned sensor embeddings $S_{\text{aligned}} \in \mathbb{R}^{C\times d}$
Output: 3D projections $S^{(3)}, T^{(3)}, S^{(3)}_{\text{aligned}} \in \mathbb{R}^{C\times 3}$
Center embeddings:
  $S_c \leftarrow S - \text{mean}(S)$
  $T_c \leftarrow T - \text{mean}(T)$
Stack centered embeddings:
  $Z \leftarrow [S_c; T_c]$
Fit PCA model with 3 components:
  PCA $\leftarrow \texttt{PCA}(n\_components = 3)$
  PCA.fit(Z)
Project embeddings to 3D:
  $S^{(3)} \leftarrow$ PCA.transform(S_c)
  $T^{(3)} \leftarrow$ PCA.transform(T_c)
  $S^{(3)}_{\text{aligned}} \leftarrow$ PCA.transform(S_{aligned})
Return: $S^{(3)}, T^{(3)}, S^{(3)}_{\text{aligned}}$

6.3. Geometric Interpretation of Alignment

The contrast between weak raw alignment and strong post-alignment correspondence is a central finding of this study. Because orthogonal Procrustes alignment preserves intra-space geometry, the observed improvement cannot be attributed to arbitrary warping or dimensional collapse. Instead, it indicates that relative relationships between classes—such as angular separation and neighborhood structure—are consistently encoded across modalities.

From a manifold perspective, these findings support the hypothesis that independently trained models can learn compatible semantic organization even when trained on fundamentally different input modalities. The sensor and text embedding spaces can thus be viewed as different coordinate charts over a shared latent semantic manifold at the class level.

6.4. Visualization of Latent Geometry

To complement the quantitative analysis, we visualize the embedding spaces before and after geometry-preserving alignment using Principal Component Analysis (PCA) to qualitatively assess latent geometric correspondence (Algorithm 2). PCA is applied to the combined set of centered sensor and text class embeddings to obtain a shared three-dimensional projection, ensuring that observed differences arise from alignment rather than changes in the projection basis.

According to Figure 1, prior to alignment, sensor and text embeddings corresponding to each of the six activity classes occupy distinct regions of the projected space, with large Euclidean distances between corresponding class

pairs. After Procrustes alignment, sensor embeddings move closer to their corresponding text embeddings and exhibit increased clustering by class. This shift reflects the effect of the orthogonal rotation, which brings the sensor embedding space into closer geometric alignment with the text embedding space, consistent with the observed increase in cosine similarity.

We quantify this effect by computing the mean Euclidean distance between matched sensor – text class pairs in the three-dimensional PCA space. This distance decreases from 2.47 before alignment to 0.13 after alignment, representing an order-of-magnitude reduction. Although PCA is a lossy projection of the original embedding space, this substantial decrease provides further evidence that the Procrustes transformation reveals compatible relational geometry between the two modalities.

6.5. Summary of Findings

Taken together, these results demonstrate that independently trained sensor and text embeddings exhibit weak apparent alignment in their original coordinate systems, but strong geometric compatibility after orthogonal transformation. This pattern indicates that semantic structure is preserved across modalities at the class level, even in the absence of joint training or paired supervision. The findings underscore the value of geometry-aware diagnostics for probing latent representations and comparing embedding spaces across models and modalities.

7. Discussion

This work investigates whether independently trained sensor and text embedding spaces exhibit compatible latent geometry at the class level. By framing alignment as the existence of a geometry-preserving transformation rather than direct coordinate correspondence, we separate semantic structure from modality-specific training artifacts.

Our results show that weak raw cosine similarity can obscure strong underlying relational structure. After orthogonal Procrustes alignment, embeddings corresponding to the same semantic class become highly aligned while preserving intra-space distances and angles. This indicates that independently trained models may encode similar semantic organization up to a rigid transformation, even without joint training or paired supervision.

We operate deliberately at the class level, aggregating instance-level sensor embeddings into centroids. This abstraction reduces intra-class variability and enables a focused examination of global semantic geometry. While this choice precludes instance-level correspondence analysis, it provides a controlled setting for probing latent structure without introducing ill-defined cross-modal pairing assumptions.

Although our experiments focus on a sensor–text setting, the methodology is broadly applicable. Geometry-aware alignment diagnostics offer a principled tool for analyzing representational compatibility in systems that combine independently trained encoders, such as vision–language or audio–text pipelines.

Our study is intentionally limited in scope. We restrict alignment to linear, orthogonal transformations and analyze a small number of semantic classes. PCA visualizations provide qualitative insight but remain lossy projections. Future work may explore nonlinear alignment, richer textual descriptions, and finer-grained geometric analysis.

8. Conclusion

We presented a geometry-first analysis of cross-modal representation alignment, treating independently trained sensor and text embeddings as latent spaces whose compatibility can be evaluated through geometry-preserving transformations. By operating at the class level and applying orthogonal Procrustes alignment, we demonstrated that weak apparent correspondence in raw coordinates can mask strong underlying relational structure.

Our results support the view that semantic organization may be encoded consistently across modalities, even in the absence of joint training or paired supervision, and that geometric diagnostics provide a valuable lens for probing such representations. Rather than proposing a new multimodal learning method, this work highlights the importance of understanding latent geometry as a prerequisite for principled model design, evaluation, and interpretation.

We hope this study encourages further exploration of geometry-aware analysis tools for latent representations, particularly in settings where models are trained independently but expected to interoperate in downstream systems.

References

[1] Guillaume Alain and Yoshua Bengio. Understanding intermediate layers using linear classifier probes. *arXiv preprint arXiv:1610.01644*, 2016. 2

[2] Jean-Baptiste Alayrac et al. Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems, 35, 23716-23736*, 2022. 1, 2

[3] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *International conference on machine learning*, pages 1247–1255. PMLR, 2013. 2

[4] Pratik Prabhanjan Brahma, Dapeng Wu, and Yiyuan She. Why deep learning works: A manifold disentanglement perspective. *IEEE transactions on neural networks and learning systems*, 27(10):1997–2008, 2015. 2

[5] Simon Kornblith et al. Similarity of neural network representations revisited. *ICML*, 2019. 1, 2

[6] Alec Radford et al. Learning transferable visual models from natural language supervision. *ICML*, 2021. 1, 2

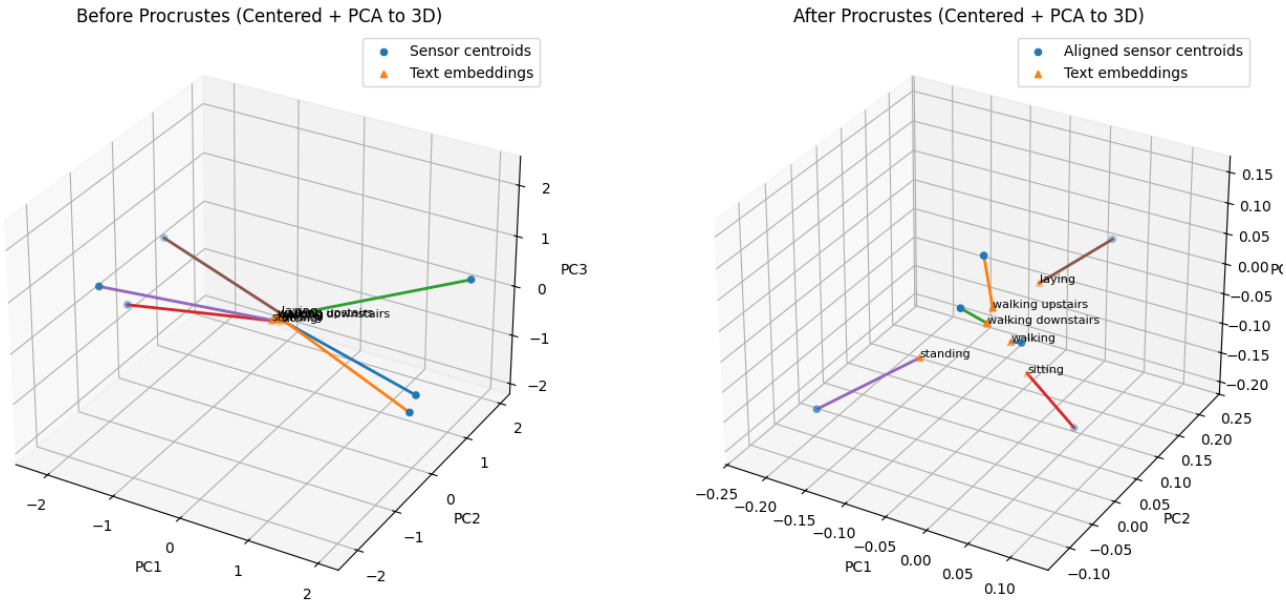

Figure 1. Three-dimensional PCA visualization of sensor and text class embeddings before and after Procrustes alignment. Each point corresponds to a class-level embedding, and lines connect matching sensor–text pairs. Prior to alignment, sensor centroids and text embeddings occupy distinct regions of the projected space. After the Procrustes rotation, the sensor embeddings move closer to their corresponding text embeddings, reflecting improved geometric alignment.

[7] Maithra Raghu, Justin Gilmer, Jason Yosinski, and Jascha Sohl-Dickstein. Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. *Advances in neural information processing systems*, 30, 2017. 1, 2

[8] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *EMNLP*, 2019. 2, 3

[9] Jorge Luis Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto, and Xavier Parra. Human activity recognition using smartphones dataset, 2013. 3

[10] sentence-transformers. sentence-transformers/all-minilm-l6-v2. https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2, 2025. 4

[11] Ankita Shukla, Shagun Uppal, Sarthak Bhagat, Saket Anand, and Pavan Turaga. Geometry of deep generative models for disentangled representations. In *Proceedings of the 11th Indian Conference on Computer Vision, Graphics and Image Processing*, pages 1–8, 2018. 2

[12] Chang Wang and Sridhar Mahadevan. Manifold alignment using procrustes analysis. Technical report, University of Massachusetts Amherst, 2008. 3

[13] Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International conference on machine learning*, pages 9929–9939. PMLR, 2020. 2

[14] Zhihan Yue et al. Ts2vec: Towards universal representation of time series. *AAAI*, 2022. 2, 3