Infusing invariances in neural representations

Irene Cannistraci [1] **Marco Fumero** [1] **Luca Moschella** [1] **Valentino Maiorca** [1] **Emanuele Rodolà** [1]

Abstract

It has been observed that inner representations learned by different neural networks conceal structural similarities when the networks are trained under similar inductive biases. Exploring the geometric structure of latent spaces within these networks offers insights into the underlying similarity among different neural models and facilitates reasoning about the transformations that connect them. Identifying and estimating these transformations presents a challenging task, but it holds significant potential for various downstream tasks, including merging and stitching different neural architectures for model reuse. In this study, drawing on the geometrical structure of latent spaces, we show how it is possible to define representations that incorporate invariances to the targeted transformations in a single framework. We experimentally analyze how inducing different invariances in the representations affects downstream performances on classification and reconstruction tasks, suggesting that the classes of transformations that relate independent latent spaces depend on the task at hand. We analyze models in a variety of settings including different initializations, architectural changes, and trained on multiple modalities (e.g., text, images), testing our framework on 8 different benchmarks.

1. Introduction

Discovering symmetries and conserved quantities is a core step to extract meaningful representations from raw data, in both biological and artificial systems (Higgins et al., 2022; Benton et al., 2020; Lyle et al., 2020). Achieving invariance to specific groups of transformations within neural models holds significant utility in a plethora of real-world applications (Cohen & Welling, 2016; Fawzi et al., 2016; Salamon & Bello, 2017).These desired invariances can be obtained through techniques that operate either on the input space (Benton et al., 2020; Immer et al., 2022) or on the latent space (Moschella et al., 2022).Additionally, achieving invariance unlocks the potential for comparing similar latent spaces across multiple training instances, facilitating communication and enabling model re-use (Klabunde et al., 2023). However, measuring latent space similarity remains a challenging task, with varying perspectives in the literature. Some studies suggest a limited similarity between layers of Neural Networks (NNs) trained from different random initialization (Raghu et al., 2017; Wang et al., 2018), while others demonstrate correspondences even across different architectures (Kornblith et al., 2019). Building upon the manipulation of neural representations, a recent study introduced the concept of Relative Representation (RR) (Moschella et al., 2022). This framework, in its original formulation, ensures invariance to latent isometries and rescalings, facilitating effective communication between latent spaces by projecting them into a shared relative space based on the distances between data points. Expanding upon this simple but powerful concept of RR, we present a methodology to *directly infuse invariances into the learned latent space*. This is achieved by switching from an absolute latent space to a relative one, defined by a similarity function that instills specific properties into our learned space.

2. Infusing invariances

Setting. We consider neural networks F as compositions of *encoding* and *decoding* maps $F = D \circ E$, where the encoder E is responsible for computing a latent representation $z = E(x), \quad x \in \mathcal{X}$ for some domain $\mathcal{X}$, with $dim(\mathcal{Z}) << dim(\mathcal{X})$; and the decoder D is responsible for solving the task at hand (e.g., reconstruction, generation, classification). We indicate with $M_{\mathcal{X}}$ if the module M was trained on the domain $\mathcal{X}$.

When considering different networks F, F' we are interested in modeling the transformation $\mathcal{T}$ that relates their latent spaces $\mathcal{Z}, \mathcal{Z}'$. The two networks could differ by their initialization seeds (i.e., training dynamics), by architectural changes, or even domain changes, i.e., $\mathcal{X} \neq \mathcal{X}'$. The fun-

[1]Department of Computer Science, Sapienza University of Rome. Correspondence to: Irene Cannistraci <cannistraci@di.uniroma1.it>.

Presented at the 2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML) at the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. 2023. *Copyright 2023 by the author(s).*

damental assumption of this work is that these variations induce changes in the latent representations of the models, but it exists an underlying manifold $\mathcal{M}$ where the representations are the same. When considering multiple models $\mathcal{F}_1..\mathcal{F}_n$ the manifold $\mathcal{M}$ identifies an equivalence class of encoders $\mathcal{E}_\mathcal{T}$ induced by the transformation $\mathcal{T}$, defined as $\{E \in \mathcal{E}_\mathcal{T} | \pi_\mathcal{M} \mathcal{T} E(\cdot) = \pi_\mathcal{M} E(\cdot)\}$ where $\pi_\mathcal{M}$ represent the projection on $\mathcal{M}$ and $E \neq E'$. What we look for is a representation r which independently projects the latent spaces $\mathcal{Z}_1..\mathcal{Z}_n$ into $\mathcal{M}$ and is therefore *invariant* to $\mathcal{T}$, i.e. $r(z) = r(\mathcal{T}z)$ for each $z \in \mathcal{Z}_1..\mathcal{Z}_n$, therefore $r \approx \pi_\mathcal{M}$

In other terms we assume that there exists a metric d which is preserved under the action of the transformation $\mathcal{T}$, i.e. $d_\mathcal{M}(z, z') = d_\mathcal{M}(\mathcal{T}(z), \mathcal{T}(z'))$.

Generalizing the framework of (Moschella et al., 2022) to arbitrary similarity functions or distance metrics gives us a straightforward way to define representations r, which are invariant to classes of transformations $\mathcal{T}$.

Relative representations. The RR framework (Moschella et al., 2022) provides a straightforward approach to represent each data sample in the latent space based on its similarity to a set of training samples called *anchors*. By representing data samples in the latent space as a function of these anchors, the framework transforms the absolute latent space into a *relative* one defined by the anchors.

Given a domain $\mathcal{X}$ with its corresponding learned embedding function $E_\mathcal{X} : \mathcal{X} \to \mathbb{R}^n$, a set of anchors $\mathcal{A}_\mathcal{X} \subset \mathcal{X}$, and a similarity function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. The RR for each sample $x \in \mathcal{X}$ is calculated as

$$RR(z, \mathcal{A}_\mathcal{X}) = \bigoplus_{a_i \in \mathcal{A}} d(z, a_i) \tag{1}$$

where $z = E_\mathcal{X}(x)$, and $\bigoplus$ denotes row-wise concatenation. For example, choosing the cosine similarity as d and additionally assuming that all embeddings are rescaled to unit norm, i.e., $||E(x)|| = 1$ corresponds to the choice of cosine similarity as the similarity function. This choice results in a relative space invariant to *angle-preserving transformations*. Thus, giving two different domains $\mathcal{X}$ and $\mathcal{Y}$ related by $\mathcal{T}$, and assuming to have two sets of anchors $\mathcal{A}_\mathcal{X}, \mathcal{A}_\mathcal{Y} \subset \mathcal{X} \times \mathcal{Y}$ in semantic correspondence, it is possible to transform the two representations into their relative form, to obtain two comparable spaces when using a d that induces representation invariant to $\mathcal{T}$.

Distance-induced invariances. In this work, we leveage the RR framework considering the following metrics d: Cosine (Cos.), Centered Cosine (Cen. Cos.), Euclidean (Eucl.), Normalized Euclidean (Norm. Eucl.), Wasserstein (Wass.), Change of Basis (CoB) and Geodesic (Geod.). Please refer to the Appendix A.2 for formal definitions. Specifically, we analyze the invariances infused by these metrics, taking into account the following classes of transformations $\mathcal{T}$: Isotropic Scaling (IS), Orthogonal Transformation (OT), Translation (TR), Permutation (PT), Affine Transformation (AT), Linear Transformation (LT), and Manifold Isometry (MIS). Where MIS is an isometric deformation of the manifold that preserves the geodesic distances between points.

Table 1 summarizes the invariances the different choices of d guarantee, meanwhile, Figures 1 and 5 provide intuitive visualizations of them. In particular, we generate a set of synthetic absolute latent spaces, apply various transformations to each absolute space, thus, converting the transformed spaces into relative spaces using the different measures d. The relative space generated by employing a specific metric d that infuses an invariance to the transformation T exhibits the anticipated characteristic of remaining unchanged, under transformations applied to the original space by T. Note that none of the chosen metrics d induce invariances to either LT or AT. Please refer to Appendix A.4 to visualize all the possible combinations of d and $\mathcal{T}$, and alternative absolute initializations (e.g., grid).

Table 1. Overview of the invariances infused in the relative latent space when choosing different similarity functions.

Similarity Function	IS	OT	TR	PT	AT	LT	MIS
Absolute	×	×	×	×	×	×	×
Cosine	√	√	×	√	×	×	×
Centered Cosine	√	√	√	√	×	×	×
Euclidean	√	√	√	√	×	×	×
Wasserstein	√	×	√	√	×	×	×
Change of Basis	√	×	×	√	×	×	×
Geodesic	√	√	√	√	×	×	√

3. Experiment

In this section, we perform qualitative and quantitative experiments to analyze the use of distinct similarity functions in creating representations that are inherently invariant to specific transformations. In particular, Section 3.1 analyzes the latent spaces produced by AutoEncoder (AE) and Variational AutoEncoder (VAE) architectures in reconstruction tasks on several datasets, meanwhile, Section 3.2 evaluates the zero-shot stitching performance in text and image classification tasks.

3.1. Latent space analysis

Experimental setting. In this experiment, we perform an image reconstruction task using AE on `CIFAR-10`, `CIFAR-100` (Krizhevsky et al., 2009), `MNIST`, and `FashionMNIST` (Xiao et al., 2017) datasets. We train five end-to-end instances of an AE and a VAE using different seed values, utilizing the Absolute (Abs.) representations until convergence. The generated representations are then

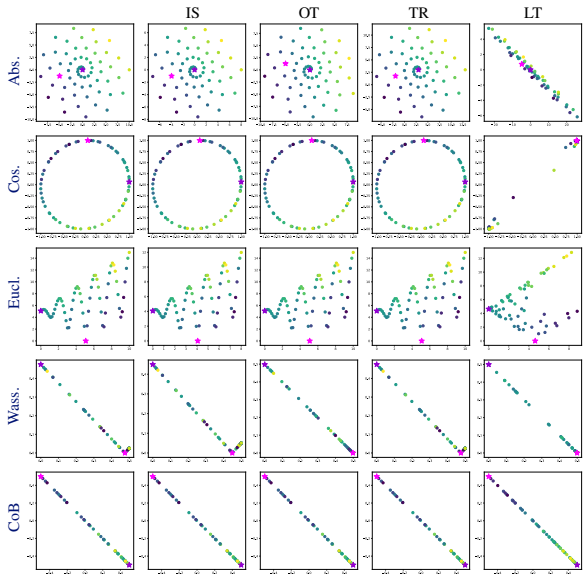

Figure 1. Qualitative synthetic results using a spiral initialization. Refer to Figure 6 for the visualization of all the possible combinations.

transformed into relative representations by projecting the embeddings onto 800 randomly selected but fixed anchors. For the relative projection, we employ five different similarity functions defined in Section 2. Finally, for each projection type, we compute the similarities across the different seeds. See Appendix A.4 for the additional results.

Table 2. Similarity space scores (± std), across 5 seeds, of VAE and AE on the `CIFAR-100` dataset. See Appendix A.4 for the complete results table.

	Projection	Cosine ↑	L1 ↓	MSE ↓	Spearman ↑
AE	Abs.	-0.00 ± 0.00	0.03 ± 0.00	$4.42\text{E-}03 \pm 0.00$	0.00 ± 0.00
	Cen. Cos.	0.99 ± 0.00	0.02 ± 0.00	$9.79\text{E-}04 \pm 0.00$	$\mathbf{0.98} \pm 0.01$
	CoB	0.06 ± 0.00	0.21 ± 0.00	$9.71\text{E-}02 \pm 0.10$	0.03 ± 0.02
	Cos.	0.98 ± 0.01	0.02 ± 0.00	$1.89\text{E-}03 \pm 0.00$	0.97 ± 0.01
	Eucl.	$\mathbf{1.00} \pm 0.00$	0.16 ± 0.13	$8.43\text{E-}02 \pm 0.11$	$\mathbf{0.98} \pm 0.01$
	Norm. Eucl.	$\mathbf{1.00} \pm 0.00$	0.01 ± 0.00	$5.16\text{E-}04 \pm 0.00$	$\mathbf{0.98} \pm 0.01$
	Wass.	0.99 ± 0.00	$\mathbf{0.00} \pm 0.00$	$\mathbf{8.80\text{E-}12} \pm 0.00$	0.84 ± 0.10
VAE	Abs.	0.00 ± 0.00	0.14 ± 0.00	$5.87\text{E-}02 \pm 0.01$	0.00 ± 0.01
	Cen. Cos.	0.99 ± 0.00	0.02 ± 0.00	$6.71\text{E-}03 \pm 0.01$	0.94 ± 0.06
	CoB	0.04 ± 0.00	0.21 ± 0.00	$1.16\text{E-}01 \pm 0.04$	0.04 ± 0.02
	Cos.	0.99 ± 0.00	0.02 ± 0.00	$6.83\text{E-}03 \pm 0.01$	0.94 ± 0.06
	Eucl.	$\mathbf{1.00} \pm 0.00$	0.13 ± 0.01	$1.01\text{E-}01 \pm 0.08$	$\mathbf{0.96} \pm 0.04$
	Norm. Eucl.	$\mathbf{1.00} \pm 0.00$	0.02 ± 0.00	$3.37\text{E-}03 \pm 0.03$	0.94 ± 0.06
	Wass.	0.89 ± 0.03	$\mathbf{0.00} \pm 0.00$	$\mathbf{1.86\text{E-}09} \pm 0.00$	0.48 ± 0.22

Result analysis. Table 2 and Figure 2 indicates that, apart from the CoB, all other projections demonstrate high similarity and low distance scores. This finding suggests that the latent spaces generated by altering only the seed values are strongly correlated within a specific transformation class.

Interestingly, both the VAE and AE achieve the best relative space compatibility across different seeds with the same projection type, even though the VAE is regularized. Our hypothesis is that the emergence of a transformation class,

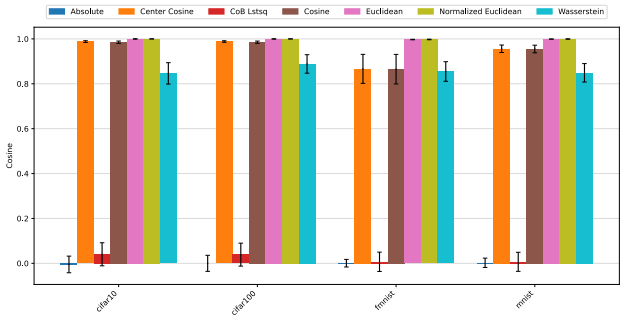

Figure 2. Cosine similarity scores (± std), across 5 seeds, of the VAE architecture on the four different datasets. See Figure 9 analogous plot on AE.

which establishes correlations among independently trained models with diverse seeds, relies primarily on the downstream task at hand (e.g., reconstruction), rather than being influenced by the VAE regularization techniques.

Furthermore, it is noteworthy that the Cos., despite being the default similarity function employed in RR (Moschella et al., 2022), never attains optimal performance as the chosen projection type. This interesting finding challenges the assumption made in RR that angle-preserving transformations are the primary drivers of correlation among the latent spaces of models trained with different seeds.

Takeaway. The emerging transformation class that relates latent spaces obtained by changing only the seed is highly dependent on the specific downstream task being addressed.

3.2. Zero-Shot Stitching

Experimental setting. In this experiment, we conduct zero-shot stitching for the classification task (text and image) on multiple datasets using various architectures. Our stitched models consist of an encoder, which embeds the data, and a specialized relative decoder responsible for the classification task. The relative decoders are trained with different seed values, and the resulting representations are transformed into relative representations by projecting the embeddings onto 1280 randomly selected but fixed anchors. The stitching process is performed in a zero-shot manner, without any training or fine-tuning, and the accuracy score for the classification task is evaluated on each assembled model.

Text Classification. We perform experiments on three different datasets `TREC (coarse)` (Hovy et al., 2001), `DBpedia` (Zhang et al., 2015), and `N24News (Text)` (Wang et al., 2022), adopting 11 different pre-trained transformers. Additional details are provided in Appendix A.3.

Image Classification. We perform experiments on six different datasets, adopting 17 different pre-trained models where 11 are transformer-based architectures while 6 are

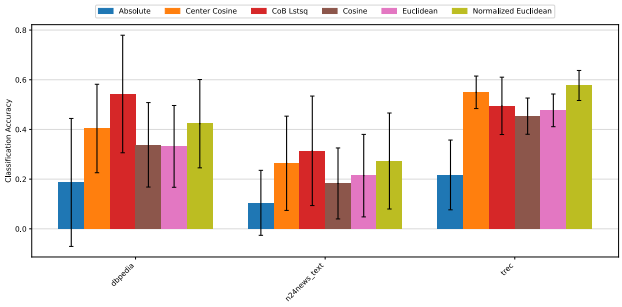

Figure 3. Cross-architecture stitching performance comparison for the text classification task. The figure shows the mean weighted Accuracy (± std) for each dataset across 5 different seeds.

ResNet-based architectures. Please refer to Appendix A.3 for a detailed list of models and datasets.

Table 3. Classification accuracy scores (± std), across 5 seeds on the `CIFAR-100 (fine)` dataset. See Appendix A.4 for the complete results table.

Projection	Cosine ↑	L1 ↓	MSE ↓	Spearman ↑	Accuracy ↑
Abs.	0.15 ± 0.32	1.74 ± 0.83	7.43 ± 5.19	0.13 ± 0.31	0.17 ± 0.29
Cen. Cos.	0.60 ± 0.10	0.08 ± 0.01	$\mathbf{0.01} \pm 0.00$	$\mathbf{0.53} \pm 0.11$	0.65 ± 0.08
CoB	0.06 ± 0.08	0.28 ± 0.33	0.36 ± 0.55	0.04 ± 0.07	0.38 ± 0.26
Cos.	0.95 ± 0.02	0.18 ± 0.13	0.06 ± 0.06	0.48 ± 0.14	0.62 ± 0.09
Eucl.	$\mathbf{1.00} \pm 0.00$	40.67 ± 35.92	2969.75 ± 3643.53	0.40 ± 0.15	0.60 ± 0.09
Norm. Eucl.	$\mathbf{1.00} \pm 0.00$	$\mathbf{0.06} \pm 0.01$	$\mathbf{0.01} \pm 0.00$	$\mathbf{0.53} \pm 0.11$	$\mathbf{0.66} \pm 0.08$

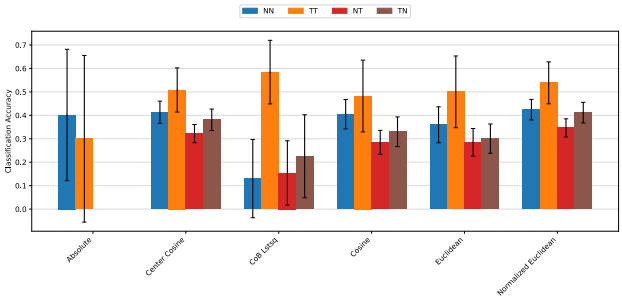

Figure 4. Stitching performance comparison with different encoding and decoding techniques: T stands for transformer-based and T for ResNet-based, thus possible combinations are: NN, TT, NT, and TN. The figure reports the mean weighted Accuracy (± std) on `CIFAR-100 (fine)`, across 5 seeds, and 17 architectures.

Result analysis. Figure 3 and Table 3 depict the performance of different projection functions in the text classification and image classification tasks, respectively. The projection function that guarantees the highest accuracy is not shared across these two modalities: in the text classification task, the CoB projection generally performs better, while when used for the image classification task performs the worst. Additionally, Figure 4 illustrates the stitching performance when employing various encoding and decoding techniques. When using the same architecture type for both encoding and decoding, the performance improves across all the projection functions, particularly when employing transformer-based architectures. This outcome validates the expectation that eliminating architectural variations as a source of diversity leads to improved similarities between independently trained models.

Remarkably, the Cos. consistently attains the highest overall performance, with only the Norm. Eucl. exhibiting comparable results. However, the Norm. Eucl. projection entails centering and unit norm normalization, rendering it conceptually similar to the Cos.. These findings support the assumption proposed in (Moschella et al., 2022) regarding the existence of an angle-preserving transformation that emerges between diverse models.

Takeaway. Results suggest that distinct latent spaces exhibit strong correlations through angle-preserving transformations when models are trained using a downstream classification task.

4. Conclusion

This work introduces a simple methodology for constructing representations that guarantee invariance to specific classes of transformations by leveraging the framework of RR. Furthermore, a preliminary analysis is conducted to explore the types of transformations that arise in classification and reconstruction tasks, with the aim of identifying suitable similarity functions that directly infuse the required invariance into the representation. Empirical findings from this study demonstrate that the transformation class exhibiting correlation among models in classification tasks corresponds to an angle-preserving one, while a distinct transformation class arises in reconstruction tasks.

Limitations and Future Works Finding an appropriate similarity function that enforces invariance to arbitrary transformations presents a significant challenge, and theoretical guarantees regarding the specific class of transformations that emerge between latent spaces under practical scenarios remain elusive. Indeed, A potential avenue for future research is to analyze a collection of models to directly estimate the range of transformations that relates them and subsequently search for the corresponding similarity function that infuses the required invariance. This approach could offer insights into the empirical bounds of the transformation classes that emerge between independently trained networks across diverse settings. Moreover, it is possible that alternative closed-form similarity functions exist that impart invariances to other transformations, which could be highly valuable in practical applications.

Acknowledgements

This work is supported by the ERC grant no.802554 (SPEC-GEO), PRIN 2020 project no.2020TA3K9N (LEGO.AI), and PNRR MUR project PE0000013-FAIR.

References

Benton, G., Finzi, M., Izmailov, P., and Wilson, A. G. Learning invariances in neural networks from training data. *Advances in neural information processing systems*, 33: 17605–17616, 2020.

Cohen, T. S. and Welling, M. Group equivariant convolutional networks, 2016.

Crane, K., Weischedel, C., and Wardetzky, M. The heat method for distance computation. *Commun. ACM*, 60(11): 90–99, October 2017. ISSN 0001-0782. doi: 10.1145/3131280. URL http://doi.acm.org/10.1145/3131280.

Davari, M., Horoi, S., Natik, A., Lajoie, G., Wolf, G., and Belilovsky, E. Reliability of cka as a similarity measure in deep learning. *arXiv preprint arXiv:2210.16156*, 2022.

Fawzi, A., Samulowitz, H., Turaga, D., and Frossard, P. Adaptive data augmentation for image classification. In *2016 IEEE international conference on image processing (ICIP)*, pp. 3688–3692. Ieee, 2016.

Gandikota, K. V., Geiping, J., Lähner, Z., Czapliński, A., and Moeller, M. Training or architecture? how to incorporate invariance in neural networks. *arXiv preprint arXiv:2106.10044*, 2021.

Higgins, I., Racanière, S., and Rezende, D. Symmetry-based representations for artificial and biological general intelligence, 2022.

Hotelling, H. Relations between two sets of variates. *Breakthroughs in statistics: methodology and distribution*, pp. 162–190, 1992.

Hovy, E., Gerber, L., Hermjakob, U., Lin, C.-Y., and Ravichandran, D. Toward semantics-based answer pinpointing. In *Proceedings of the first international conference on Human language technology research*, 2001.

Immer, A., van der Ouderaa, T., Rätsch, G., Fortuin, V., and van der Wilk, M. Invariance learning in deep neural networks with differentiable laplace approximations. *Advances in Neural Information Processing Systems*, 35: 12449–12463, 2022.

Kandi, H., Jain, A., Velluva Chathoth, S., Mishra, D., and Subrahmanyam, G. R. S. Incorporating rotational invariance in convolutional neural network architecture. *Pattern Analysis and Applications*, 22:935–948, 2019.

Klabunde, M., Schumacher, T., Strohmaier, M., and Lemmerich, F. Similarity of neural network models: A survey of functional and representational measures. *arXiv preprint arXiv:2305.06329*, 2023.

Kornblith, S., Norouzi, M., Lee, H., and Hinton, G. Similarity of neural network representations revisited. In *International Conference on Machine Learning*, pp. 3519–3529. PMLR, 2019.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

Kuprieiev, R., skshetry, Petrov, D., Redzyński, P., Rowlands, P., da Costa-Luis, C., Schepanovski, A., Shcheklein, I., Taskaya, B., Gao, Orpinel, J., de la Iglesia Castro, D., Santos, F., Sharma, A., Berenbaum, D., Zhanibek, Hodovic, D., daniele, Kodenko, N., Grigorev, A., Earl, Dash, N., Vyshnya, G., Lamy, R., maykulkarni, Hora, M., Vera, and Mangal, S. Dvc: Data version control - git for data & models, 2022. URL https://doi.org/10.5281/zenodo.7083378.

Lyle, C., van der Wilk, M., Kwiatkowska, M., Gal, Y., and Bloem-Reddy, B. On the benefits of invariance in neural networks. *arXiv preprint arXiv:2005.00178*, 2020.

Morcos, A., Raghu, M., and Bengio, S. Insights on representational similarity in neural networks with canonical correlation. *Advances in Neural Information Processing Systems*, 31, 2018.

Moschella, L., Maiorca, V., Fumero, M., Norelli, A., Locatello, F., and Rodolà, E. Relative representations enable zero-shot latent space communication. *arXiv preprint arXiv:2209.15430*, 2022.

Raghu, M., Gilmer, J., Yosinski, J., and Sohl-Dickstein, J. Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. *Advances in neural information processing systems*, 30, 2017.

Rath, M. and Condurache, A. P. Deep neural networks with efficient guaranteed invariances. In *International Conference on Artificial Intelligence and Statistics*, pp. 2460–2480. PMLR, 2023.

Salamon, J. and Bello, J. P. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal processing letters*, 24(3):279–283, 2017.

Shao, H., Kumar, A., and Fletcher, P. T. The riemannian geometry of deep generative models, 2017.

Tenenbaum, J. B., Silva, V. d., and Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

van der Ouderaa, T. F. and van der Wilk, M. Learning invariant weights in neural networks. In *Uncertainty in Artificial Intelligence*, pp. 1992–2001. PMLR, 2022.

Wang, L., Hu, L., Gu, J., Hu, Z., Wu, Y., He, K., and Hopcroft, J. Towards understanding learning representations: To what extent do different neural networks learn the same representation. *Advances in neural information processing systems*, 31, 2018.

Wang, Z., Shan, X., Zhang, X., and Yang, J. N24News: A new dataset for multimodal news classification. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 6768–6775, Marseille, France, June 2022. European Language Resources Association. URL https://aclanthology.org/2022.lrec-1.729.

Xiao, H., Rasul, K., and Vollgraf, R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

Zhang, X., Zhao, J., and LeCun, Y. Character-level convolutional networks for text classification. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/250cf8b51c773f3f8dc8b4be867a9a02-Paper.pdf.

A. Appendix

A.1. Related work

Comparison methods. Several metrics have been proposed to facilitate the comparison of latent spaces generated by independent NNs, aiming to capture their inherent similarity up to some transformation that correlates the spaces. A classical statistical method is Canonical Correlation Analysis (CCA) (Hotelling, 1992), which is invariant to linear transformations, and its variations to improve robustness via Singular Value Decomposition (SVD), Singular Value CCA (SVCCA) (Raghu et al., 2017), and to reduce the sensitivity of CCA to perturbations, Projection Weighted CCA (PWCCA) (Morcos et al., 2018). Closely related to these metrics, the Centered Kernel Alignment (CKA) metric (Kornblith et al., 2019) measures the similarity between latent spaces while disregarding orthogonal transformations. However, a recent study (Davari et al., 2022) demonstrates its sensitivity to transformations that shift a subset of data points in the representation space.

These metrics provide a way to quantify the similarity of the shared structure across seemingly dissimilar representations. In this work, we do not introduce new metrics that are invariant to specific transformations. Instead, *we leverage the RR framework to construct representations that are inherently invariant to certain transformations*, allowing for direct comparisons between them.

Learning and incorporating invariance and equivariance into representation. Invariances in NN can be achieved through various techniques operating at different levels, including model architecture adjustment, training constraints, or input manipulation (Lyle et al., 2020). Some studies focus on learning these invariances, particularly in scenarios where the specific invariances in the data are unknown or the desired level of invariance to a particular symmetry group is uncertain. (Benton et al., 2020) propose a method that parameterizes an augmentation distribution and jointly optimizes the training loss with respect to the network and augmentation parameters to learn invariances and equivariances; (Immer et al., 2022) introduce a gradient-based approach that effectively captures inherent invariances in the data, improving generalization and data efficiency on image datasets; while (van der Ouderaa & van der Wilk, 2022) enable the training of NNs with invariance to specific transformations by learning weight-space equivalents instead of modifying the input data. Alternatively, other works directly incorporate invariances into the model or the representations through specific constraints. (Rath & Condurache, 2023) enforce a multi-stream architecture to exhibit invariance to various symmetry transformations without relying on data-driven learning, while (Kandi et al., 2019) propose an improved Convolutional Neural Network (CNN) architecture for better rotation invariance. (Gandikota et al., 2021) introduce a method for designing network architectures that are invariant or equivariant to structured transformations associated with a group action. Whereas, (Moschella et al., 2022) propose an alternative representation of the latent space that guarantees invariance to latent isometries and rescalings without requiring additional training.

In our approach, rather than modifying the NN architecture or imposing additional training constraints, we directly *infuse the desired invariances into the learned latent space*. This enables the achievement of a consistent latent representation across different architectures or datasets.

A.2. Distance-induced invariances details

This section provides additional details about the metrics we described in section Section 2.

Cosine (Cos.) Given two vectors *u, v*, the cosine similarity is defined as:

$$\cos(u, v) = \frac{u \cdot v}{\|u\| \, \|v\|} \tag{2}$$

Centered Cosine (Cen. Cos.) Given two vectors *u, v* belonging to the latent space $X \in R^{n \times d}$, the centered cosine similarity is defined as:

$$\cos(u, v) = \frac{(u - mean(\mathbf{X})) \cdot (v - mean(\mathbf{X}))}{\|u - mean(\mathbf{X})\| \, \|v - mean(\mathbf{X})\|} \tag{3}$$

where $mean(\mathbf{X})$ is calculated across the n dimension in order to perform center normalization.

Euclidean (Eucl.) Given two vectors *u, v*, the Euclidean distance is defined as:

$$d(u, v) = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2} \tag{4}$$

Normalized Euclidean (Norm. Eucl.) Given two vectors *u, v*, the Normalized Euclidean distance is defined as:

$$d(u,v) = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2} \tag{5}$$

where u, v are first normalized with $norm(x) = \frac{x - mean(\mathbf{X})}{\|x - mean(\mathbf{X})\|}$ to have centered unit norm vectors.

Change of Basis (CoB) Given two spaces *U, V*, where *U* contains the **anchors** to the RR computation, the Change of Basis is defined as:

$$CoB(u_i, v_j) = lstsq(U^T, V^T)_{i,j} \tag{6}$$

where *lstsq* computes a solution to the least squares problem for a linear system $UX = V$ with $U \in \mathbb{K}^{m \times n}$, $V \in \mathbb{K}^{m \times k}$ is defined as $\min_{X \in \mathbb{K}^{n \times k}} \|UX - V\|_F$, where $\| - \|_F$ denotes the Frobenius norm. Additionally a normalizing function $norm(x) = \frac{x - mean(\mathbf{X})}{\|x - mean(\mathbf{X})\|}$ is applied to each vector $x \in X$ to have centered unit norm vectors.

Wasserstein (Wass.) Given two vectors *u, v*, considered as two distributions normalized with the *Softmax* function $\theta : \mathbb{R}^K \to (0,1)^K$, $\theta(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$, for $i = 1, ..., K$ and $z = (z_1, ..., z_k) \in \mathbb{R}^K$, the Wasserstein distance is defined as:

$$W(u,v) = \inf_{\pi \in \Gamma(u,v)} \int_{\mathbb{R} \times \mathbb{R}} |x - y| d\pi(x,y) \tag{7}$$

Geodesic distance Given a manifold $\mathcal{M}$ and its parametrization $g : \mathcal{Z} \mapsto \mathcal{X}$ we can represent the Riemannian metric as symmetric, positive definite matrix $G(z)$ defined at each point in Z. $G(z)$ can be obtained as $G(z) = J_g(z)^T J_g(z)$, where $J_g(z)$ indicates the Jacobian of g evaluated at z. This metric enables us to define an inner product on tangent spaces on $\mathcal{M}$. Considering a smooth curve $\gamma : [a,b] \mapsto \mathcal{Z}$ this corresponds to a curve on $\mathcal{M}$ via $g \circ \gamma(t)$. Its arc length is defined as:

$$L(\gamma) = \int_a^b \sqrt{\dot{\gamma}(t)^T G_{\gamma(t)} \dot{\gamma}(t) dt} \tag{8}$$

A *geodesic* curve is a curve that locally minimizes the arc length, corresponding to minimizing the following energy functional:

$$E(\gamma) = \frac{1}{2} \int_a^b \dot{\gamma}(t)^T G_{\gamma(t)} \dot{\gamma}(t) dt \tag{9}$$

In Figure 5 we show how geodesic distance is preserved under several classes of transformations, including manifold isometries, i.e., possibly nonlinear transformations that preserve the metric on $\mathcal{M}$. In the synthetic experiment geodesic distances are computed using the heat method of (Crane et al., 2017) and the manifold isometry is computed using Isomap (Tenenbaum et al., 2000). Possible approaches to extend geodesic computation to real cases when $dim(\mathcal{Z}) > 3$ include (Shao et al., 2017). We leave this promising direction for future work.

A.3. Implementation Details

This section provides further details about the experiments conducted in Section 3. Table 4 contains the full list of pre-trained models used, while Table 5 contains dataset information.

A.3.1. Tools & Technologies

We use the following tools in all the experiments presented in this work:

- *PyTorch Lightning*, to ensure reproducible results while also getting a clean and modular codebase;
- *Transformers by HuggingFace*, to get ready-to-use transformers for both text and images;
- *Datasets by HuggingFace*, to access most of the datasets;
- *DVC* (Kuprieiev et al., 2022), for data versioning;

A.4. Additional results

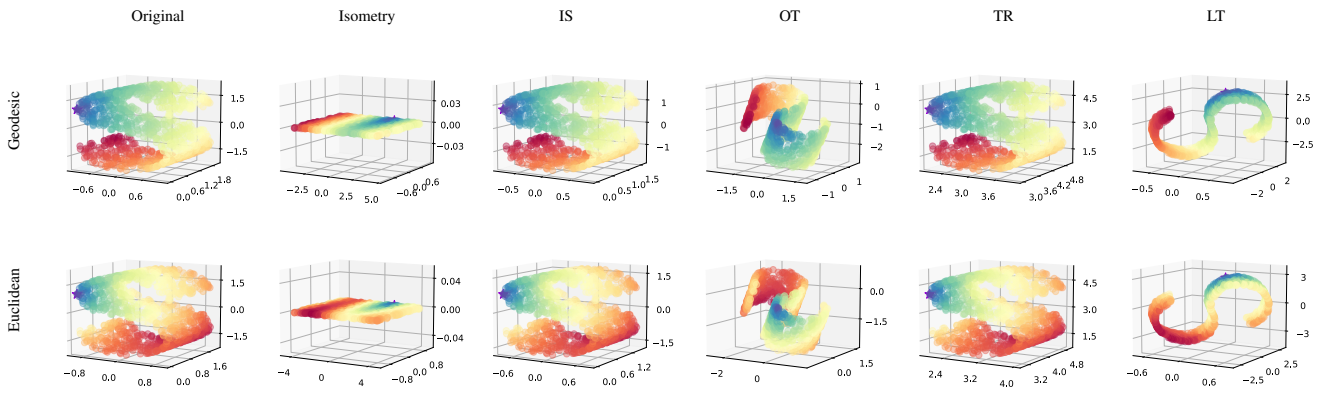

Figure 5. Qualitative synthetic results demonstrating invariances induced by using a geodesic distance-based representation. We plot geodesic distances (*top row*) from the violet star point with values going from blue (closer) to red (farther). On the *bottom row* we compare with Euclidean distances showing that the latter does not estimate nor preserve well the metric information under transformations of the manifold.

Table 4. The pre-trained feature extractors used in the various experiments, with their HuggingFace name and encoding dim.

Modality	HuggingFace model name	Encoding Dim
Language	bert-base-cased	768
	bert-base-uncased	768
	google/electra-base-discriminator	768
	roberta-base	768
	albert-base-v2	768
	xlm-roberta-base	768
	openai/clip-vit-base-patch32	768
Vision	rexnet_100	1280
	cspdarknet53	1024
	vit_small_patch16_224	384
	vit_base_patch16_224	768
	vit_base_patch16_384	768
	vit_base_resnet50_384	768
	openai/clip-vit-base-patch32	768
	efficientnet_b1_pruned	1280
	regnety_002	368
	cspresnext50	2048

Table 5. HuggingFace datasets utilized in the classification experiments, with their number of classes.

	Dataset	**Number of Classes**
Image	`MNIST`	10
	`Fashion MNIST`	10
	`CIFAR 10`	10
	`Cifar 100`	20 (coarse) — 100 (fine)
	`N24News`	24
Text	`TREC`	6 (coarse) — 50 (fine)
	`DBPEDIA 14`	14
	`N24News`	24

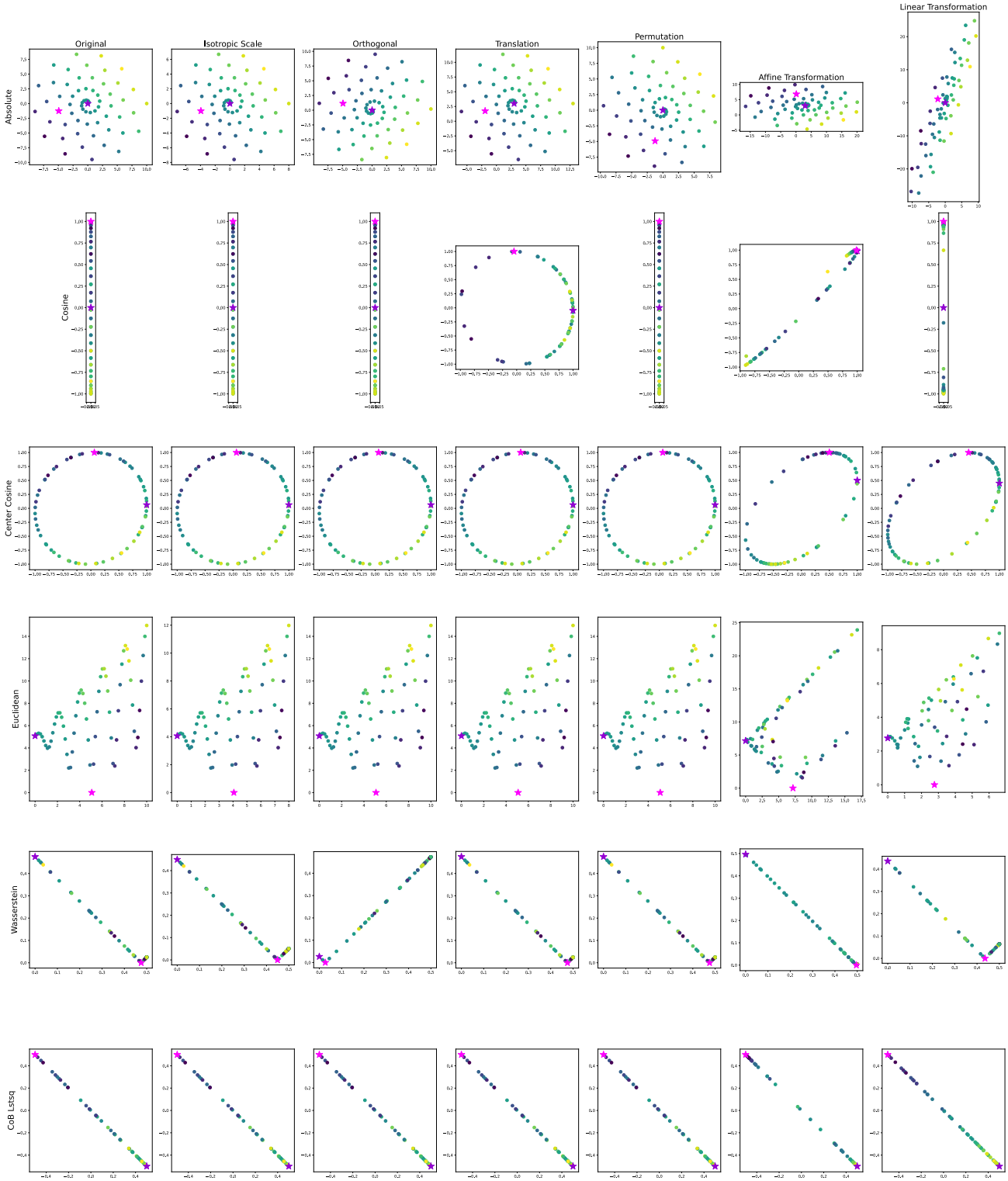

Figure 6. Synthetic experiments using a spiral initialization.

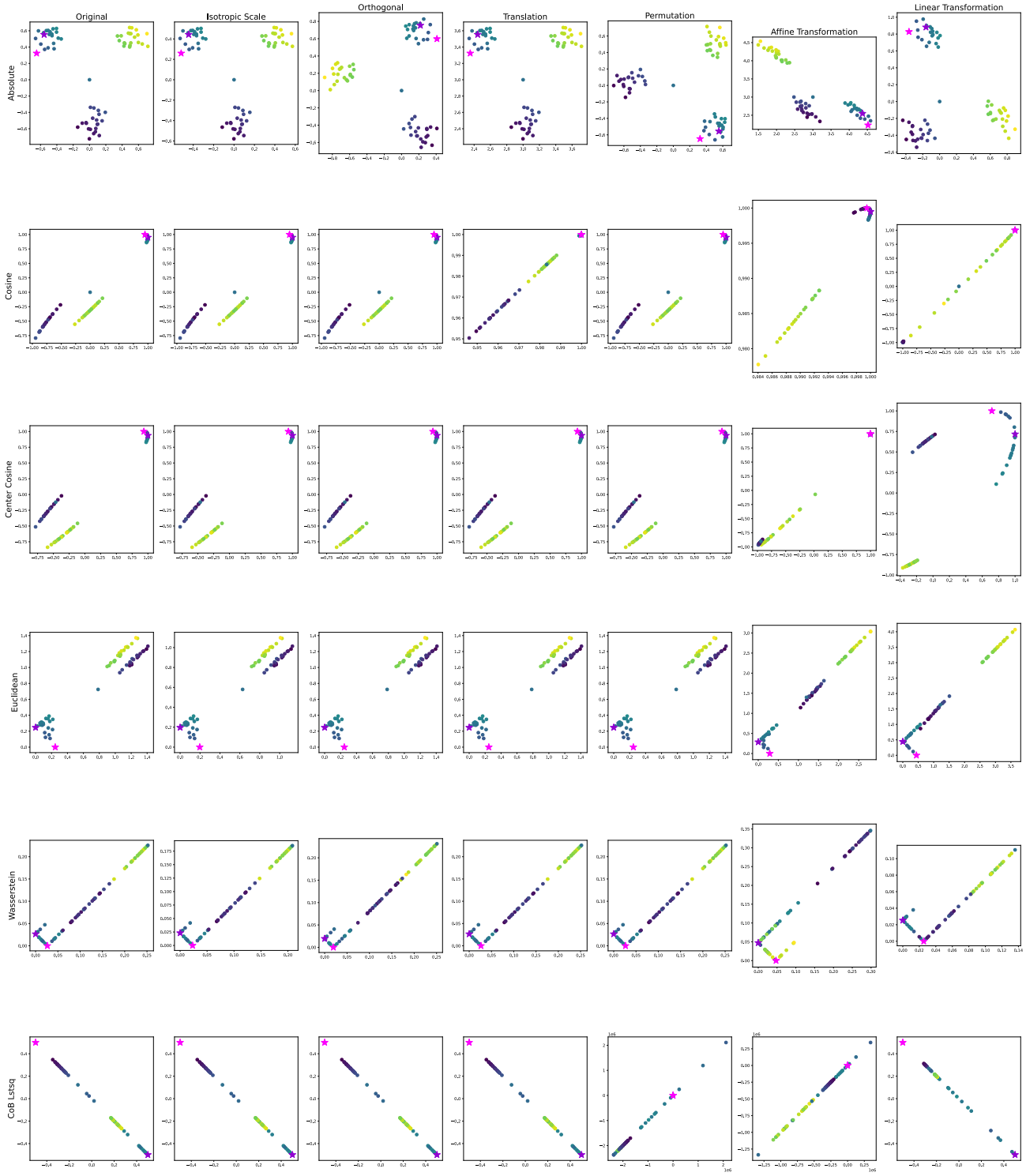

Figure 7. Synthetic experiments using random clustered initialization.

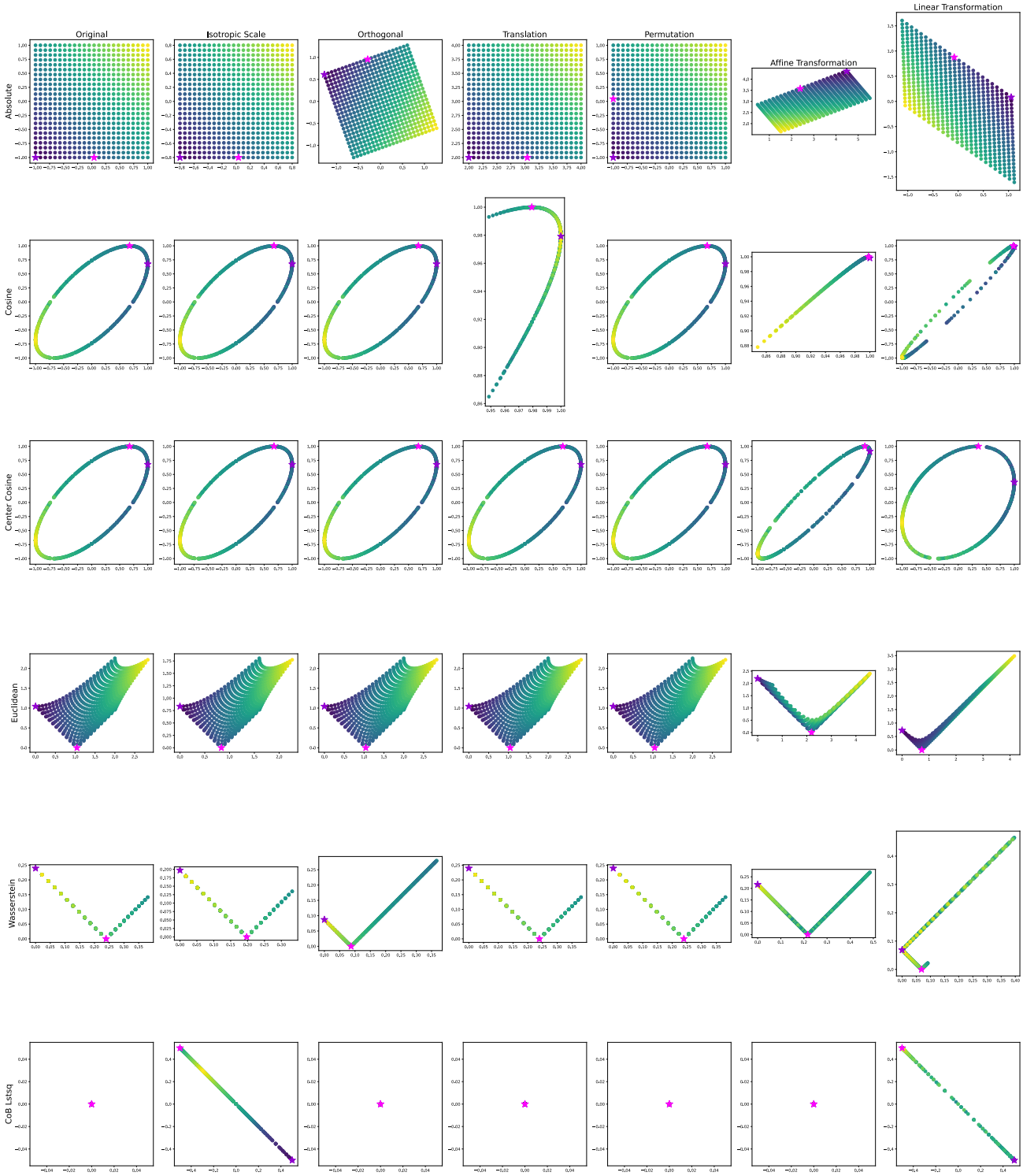

Figure 8. Synthetic experiments using a grid initialization.

Table 6. Overiew of the synthetic relative latent space invariances with respect to Isometric Scaling (IS), Orthogonal Transformation (OT), Translation (TR), Permutation (PT), Affine Transformation (AT), and Linear Transformation (LT).

RR	$\mathcal{T}$	**MSE** $\downarrow$	**L1** $\downarrow$	**Cosine** $\uparrow$
Cos.	AT	0.688 ± 0.716	0.691 ± 0.442	0.215 ± 0.647
	IS	0.000 ± 0.000	0.000 ± 0.000	0.984 ± 0.125
	LT	0.156 ± 0.284	0.211 ± 0.314	0.453 ± 0.881
	OT	0.000 ± 0.000	0.000 ± 0.000	0.984 ± 0.125
	PT	0.000 ± 0.000	0.000 ± 0.000	0.984 ± 0.125
	TR	0.544 ± 0.433	0.655 ± 0.341	0.231 ± 0.558
Cen. Cos.	AT	0.318 ± 0.434	0.368 ± 0.360	0.785 ± 0.185
	IS	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
	LT	0.603 ± 0.868	0.557 ± 0.540	0.594 ± 0.395
	OT	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
	PT	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
	TR	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
Eucl.	AT	63.084 ± 90.853	5.076 ± 4.349	0.979 ± 0.030
	IS	1.860 ± 1.789	1.197 ± 0.657	1.000 ± 0.000
	LT	26.680 ± 36.388	3.754 ± 2.797	0.994 ± 0.009
	OT	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
	PT	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
	TR	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
Wass.	AT	0.033 ± 0.052	0.121 ± 0.125	0.870 ± 0.190
	IS	0.001 ± 0.001	0.018 ± 0.014	0.998 ± 0.003
	LT	0.014 ± 0.023	0.063 ± 0.067	0.930 ± 0.122
	OT	0.024 ± 0.047	0.098 ± 0.119	0.880 ± 0.229
	PT	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
	TR	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
CoB	AT	0.055 ± 0.070	0.142 ± 0.138	0.646 ± 0.707
	IS	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
	LT	0.027 ± 0.037	0.093 ± 0.097	0.781 ± 0.582
	OT	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
	PT	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000
	TR	0.000 ± 0.000	0.000 ± 0.000	1.000 ± 0.000

Table 7. AE VAE analysis results

model	dataset	projection	**Cosine ↑**	**L1 ↓**	**MSE ↓**	**Pearson ↑**	**Spearman ↑**
AE	CIFAR 10	Abs.	-0.00 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	-0.00 ± 0.00	-0.00 ± 0.00
		Cen. Cos.	0.99 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
		CoB	0.06 ± 0.00	0.20 ± 0.00	0.07 ± 0.00	0.06 ± 0.00	0.06 ± 0.00
		Cos.	0.98 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.98 ± 0.00
		Eucl.	1.00 ± 0.00	0.12 ± 0.07	0.02 ± 0.02	0.99 ± 0.00	0.99 ± 0.00
		Norm. Eucl.	1.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
		Wass.	0.99 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.94 ± 0.01	0.91 ± 0.01
	CIFAR 100	Abs.	-0.00 ± 0.00	0.03 ± 0.00	0.00 ± 0.00	-0.00 ± 0.00	0.00 ± 0.00
		Cen. Cos.	0.99 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
		CoB	0.06 ± 0.00	0.21 ± 0.00	0.07 ± 0.00	0.06 ± 0.00	0.05 ± 0.00
		Cos.	0.98 ± 0.01	0.02 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.98 ± 0.00
		Eucl.	1.00 ± 0.00	0.16 ± 0.13	0.05 ± 0.06	0.99 ± 0.00	0.99 ± 0.00
		Norm. Eucl.	1.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
		Wass.	0.99 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.94 ± 0.01	0.91 ± 0.01
	fmnist	Abs.	-0.00 ± 0.01	0.03 ± 0.00	0.00 ± 0.00	-0.00 ± 0.01	-0.00 ± 0.00
		Cen. Cos.	0.98 ± 0.00	0.03 ± 0.00	0.00 ± 0.00	0.98 ± 0.00	0.97 ± 0.00
		CoB	0.01 ± 0.00	0.26 ± 0.00	0.12 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
		Cos.	0.97 ± 0.01	0.04 ± 0.00	0.00 ± 0.00	0.98 ± 0.00	0.96 ± 0.01
		Eucl.	1.00 ± 0.00	0.28 ± 0.18	0.11 ± 0.12	0.99 ± 0.00	0.98 ± 0.00
		Norm. Eucl.	1.00 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.98 ± 0.00	0.97 ± 0.00
		Wass.	0.98 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.91 ± 0.01	0.87 ± 0.01
	mnist	Abs.	0.00 ± 0.01	0.06 ± 0.00	0.01 ± 0.00	0.00 ± 0.01	0.00 ± 0.00
		Cen. Cos.	0.98 ± 0.01	0.03 ± 0.00	0.00 ± 0.00	0.98 ± 0.01	0.97 ± 0.01
		CoB	0.01 ± 0.00	0.28 ± 0.01	0.13 ± 0.01	0.01 ± 0.00	0.01 ± 0.00
		Cos.	0.95 ± 0.03	0.05 ± 0.02	0.00 ± 0.00	0.97 ± 0.01	0.96 ± 0.01
		Eucl.	1.00 ± 0.00	0.34 ± 0.19	0.16 ± 0.14	0.98 ± 0.01	0.97 ± 0.01
		Norm. Eucl.	1.00 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.98 ± 0.00	0.97 ± 0.01
		Wass.	0.97 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.75 ± 0.04	0.67 ± 0.04
VAE	CIFAR 10	Abs.	-0.00 ± 0.01	0.15 ± 0.00	0.07 ± 0.00	-0.01 ± 0.01	-0.00 ± 0.01
		Cen. Cos.	0.99 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
		CoB	0.04 ± 0.00	0.21 ± 0.00	0.08 ± 0.00	0.04 ± 0.00	0.04 ± 0.00
		Cos.	0.99 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.98 ± 0.00
		Eucl.	1.00 ± 0.00	0.15 ± 0.03	0.04 ± 0.01	0.99 ± 0.00	0.99 ± 0.00
		Norm. Eucl.	1.00 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
		Wass.	0.85 ± 0.08	0.00 ± 0.00	0.00 ± 0.00	0.54 ± 0.12	0.70 ± 0.03
	CIFAR 100	Abs.	0.00 ± 0.00	0.14 ± 0.00	0.07 ± 0.00	0.00 ± 0.00	0.00 ± 0.01
		Cen. Cos.	0.99 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
		CoB	0.04 ± 0.00	0.21 ± 0.00	0.07 ± 0.00	0.04 ± 0.00	0.04 ± 0.00
		Cos.	0.99 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.98 ± 0.00
		Eucl.	1.00 ± 0.00	0.13 ± 0.01	0.03 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
		Norm. Eucl.	1.00 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
		Wass.	0.89 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.58 ± 0.11	0.68 ± 0.04
	FMNIST	Abs.	0.00 ± 0.00	0.08 ± 0.00	0.04 ± 0.00	0.00 ± 0.00	0.00 ± 0.01
		Cen. Cos.	0.87 ± 0.02	0.11 ± 0.01	0.02 ± 0.00	0.87 ± 0.02	0.85 ± 0.02
		CoB	0.01 ± 0.00	0.30 ± 0.01	0.16 ± 0.01	0.01 ± 0.00	0.01 ± 0.00
		Cos.	0.87 ± 0.02	0.11 ± 0.01	0.02 ± 0.00	0.87 ± 0.02	0.85 ± 0.02
		Eucl.	1.00 ± 0.00	0.36 ± 0.02	0.23 ± 0.02	0.91 ± 0.01	0.89 ± 0.01
		Norm. Eucl.	1.00 ± 0.00	0.08 ± 0.00	0.01 ± 0.00	0.88 ± 0.01	0.85 ± 0.02
		Wass.	0.85 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.46 ± 0.08	0.34 ± 0.06
	MNIST	Abs.	0.00 ± 0.01	0.08 ± 0.00	0.06 ± 0.00	0.00 ± 0.01	0.00 ± 0.00
		Cen. Cos.	0.96 ± 0.01	0.05 ± 0.00	0.00 ± 0.00	0.96 ± 0.01	0.95 ± 0.01
		CoB	0.01 ± 0.00	0.30 ± 0.01	0.15 ± 0.01	0.01 ± 0.00	0.01 ± 0.00
		Cos.	0.95 ± 0.01	0.05 ± 0.00	0.00 ± 0.00	0.96 ± 0.01	0.95 ± 0.01
		Eucl.	1.00 ± 0.00	0.26 ± 0.02	0.11 ± 0.02	0.96 ± 0.00	0.95 ± 0.01
		Norm. Eucl.	1.00 ± 0.00	0.04 ± 0.00	0.00 ± 0.00	0.96 ± 0.01	0.95 ± 0.01
		Wass.	0.85 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.24 ± 0.03	0.20 ± 0.04

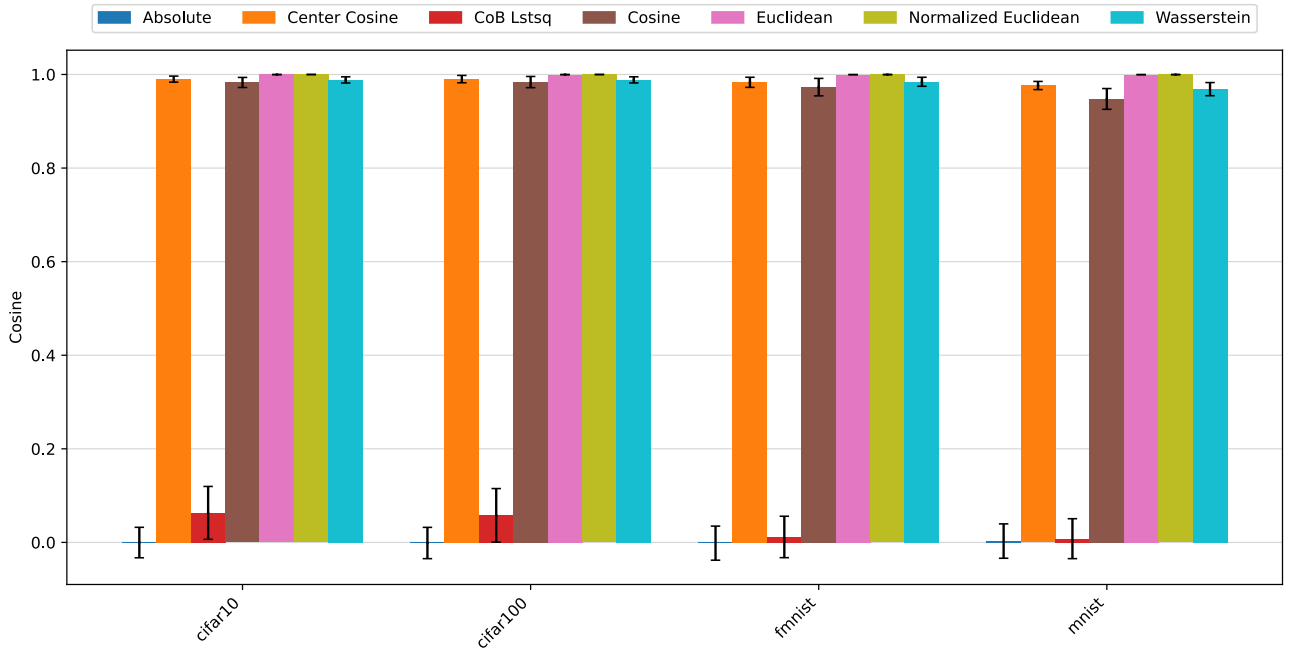

Figure 9. Cosine similarity scores (± std), over 5 seeds, of the AE architecture on the four different datasets.

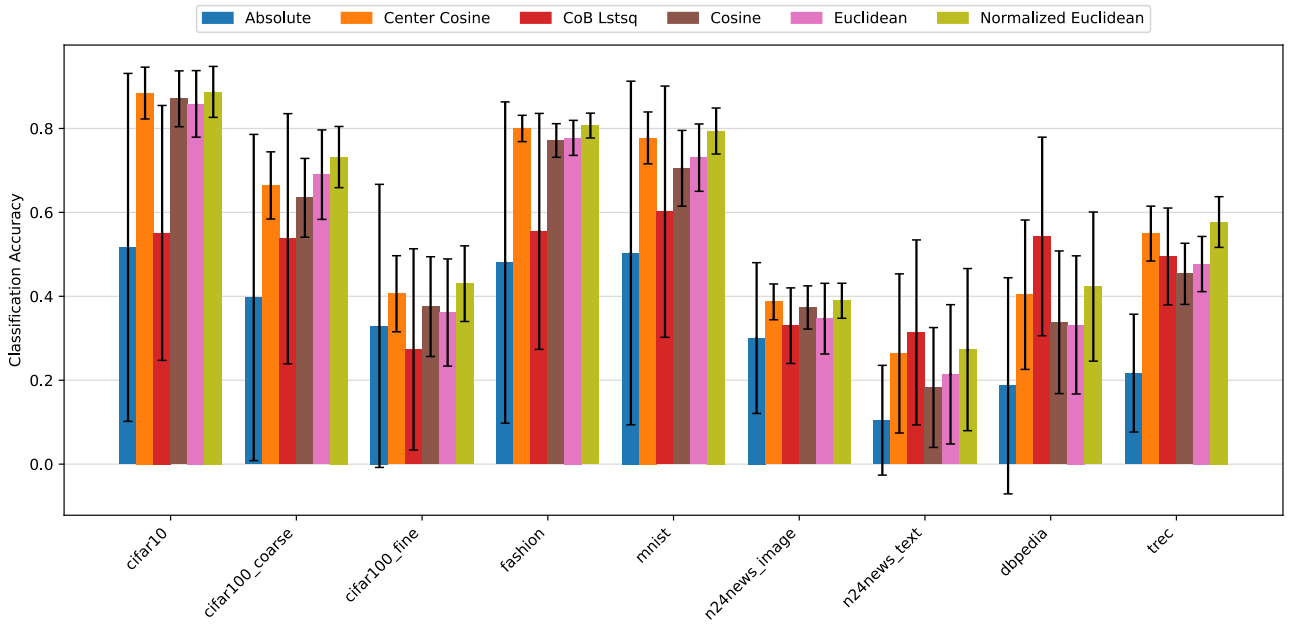

Figure 10. Cross-architecture stitching performance comparison for both image and text classification task. The figure shows the mean weighted Accuracy (± std) for each dataset, across 5 seeds.

Table 8. Results of the stitching experiment on classification tasks.

		Cosine ↑	L1 ↓	MSE ↓	Pearson ↑	Accuracy ↑
CIFAR 10	Abs.	0.15 ± 0.32	1.74 ± 0.83	7.39 ± 5.17	0.13 ± 0.31	0.23 ± 0.31
	Cen. Cos.	0.69 ± 0.08	0.08 ± 0.01	0.01 ± 0.00	0.62 ± 0.08	0.88 ± 0.06
	CoB	0.05 ± 0.07	0.33 ± 0.41	0.54 ± 0.89	0.04 ± 0.07	0.53 ± 0.30
	Cos.	0.96 ± 0.02	0.17 ± 0.13	0.05 ± 0.06	0.55 ± 0.12	0.87 ± 0.07
	Eucl.	1.00 ± 0.00	39.87 ± 35.01	2847.94 ± 3488.28	0.48 ± 0.13	0.85 ± 0.08
	Norm. Eucl.	1.00 ± 0.00	0.06 ± 0.01	0.01 ± 0.00	0.62 ± 0.08	0.88 ± 0.06
CIFAR 100 Coarse	Abs.	0.15 ± 0.32	1.74 ± 0.83	7.43 ± 5.19	0.13 ± 0.31	0.17 ± 0.29
	Cen. Cos.	0.60 ± 0.10	0.08 ± 0.01	0.01 ± 0.00	0.53 ± 0.11	0.65 ± 0.08
	CoB	0.06 ± 0.08	0.28 ± 0.33	0.36 ± 0.55	0.04 ± 0.07	0.38 ± 0.26
	Cos.	0.95 ± 0.02	0.18 ± 0.13	0.06 ± 0.06	0.48 ± 0.14	0.62 ± 0.09
	Eucl.	1.00 ± 0.00	40.67 ± 35.92	2969.75 ± 3643.53	0.40 ± 0.15	0.60 ± 0.09
	Norm. Eucl.	1.00 ± 0.00	0.06 ± 0.01	0.01 ± 0.00	0.53 ± 0.11	0.66 ± 0.08
CIFAR 100 Fine	Abs.	0.15 ± 0.32	1.74 ± 0.83	7.43 ± 5.19	0.13 ± 0.31	0.12 ± 0.27
	Cen. Cos.	0.60 ± 0.10	0.08 ± 0.01	0.01 ± 0.00	0.53 ± 0.11	0.39 ± 0.08
	CoB	0.06 ± 0.08	0.35 ± 0.49	0.77 ± 2.03	0.04 ± 0.07	0.25 ± 0.22
	Cos.	0.95 ± 0.02	0.18 ± 0.13	0.06 ± 0.06	0.48 ± 0.14	0.35 ± 0.10
	Eucl.	1.00 ± 0.00	40.66 ± 35.90	2968.71 ± 3642.09	0.41 ± 0.15	0.34 ± 0.11
	Norm. Eucl.	1.00 ± 0.00	0.06 ± 0.01	0.01 ± 0.00	0.53 ± 0.11	0.42 ± 0.08
DBPEDIA 14	Abs.	-0.00 ± 0.04	0.52 ± 0.23	0.85 ± 0.41	-0.00 ± 0.01	0.07 ± 0.01
	Cen. Cos.	0.39 ± 0.23	0.26 ± 0.10	0.11 ± 0.08	0.36 ± 0.23	0.37 ± 0.15
	CoB	0.05 ± 0.02	0.04 ± 0.01	0.00 ± 0.00	0.02 ± 0.01	0.51 ± 0.22
	Cos.	0.96 ± 0.02	0.36 ± 0.21	0.19 ± 0.15	0.28 ± 0.22	0.29 ± 0.12
	Eucl.	0.96 ± 0.02	12.08 ± 7.59	212.93 ± 211.77	0.28 ± 0.23	0.28 ± 0.11
	Norm. Eucl.	0.98 ± 0.01	0.20 ± 0.08	0.08 ± 0.05	0.36 ± 0.23	0.39 ± 0.15
Fashion MNIST	Abs.	0.13 ± 0.32	1.74 ± 0.83	7.68 ± 5.39	0.12 ± 0.31	0.21 ± 0.27
	Cen. Cos.	0.82 ± 0.05	0.10 ± 0.02	0.02 ± 0.01	0.75 ± 0.06	0.80 ± 0.03
	CoB	0.05 ± 0.05	0.24 ± 0.27	0.27 ± 0.43	0.03 ± 0.04	0.53 ± 0.28
	Cos.	0.99 ± 0.01	0.16 ± 0.09	0.04 ± 0.04	0.69 ± 0.08	0.76 ± 0.03
	Eucl.	1.00 ± 0.00	30.42 ± 26.65	1676.46 ± 2058.34	0.65 ± 0.08	0.77 ± 0.04
	Norm. Eucl.	1.00 ± 0.00	0.07 ± 0.01	0.01 ± 0.00	0.75 ± 0.06	0.80 ± 0.03
MNIST	Abs.	0.12 ± 0.34	1.69 ± 0.85	6.83 ± 4.90	0.11 ± 0.34	0.21 ± 0.28
	Cen. Cos.	0.64 ± 0.10	0.15 ± 0.02	0.04 ± 0.01	0.60 ± 0.11	0.76 ± 0.04
	CoB	0.04 ± 0.04	0.17 ± 0.18	0.12 ± 0.19	0.03 ± 0.03	0.58 ± 0.30
	Cosine	0.99 ± 0.00	0.15 ± 0.08	0.03 ± 0.03	0.59 ± 0.13	0.68 ± 0.06
	Eucl.	0.99 ± 0.00	18.10 ± 16.21	620.56 ± 764.42	0.57 ± 0.12	0.71 ± 0.06
	Norm. Eucl.	0.99 ± 0.00	0.11 ± 0.02	0.02 ± 0.01	0.60 ± 0.11	0.78 ± 0.04
N24News (image)	Abs.	0.30 ± 0.42	2.07 ± 0.64	8.72 ± 4.54	0.28 ± 0.41	0.18 ± 0.19
	Cen. Cos.	0.64 ± 0.11	0.06 ± 0.01	0.01 ± 0.00	0.57 ± 0.12	0.38 ± 0.04
	CoB	0.12 ± 0.11	0.08 ± 0.19	0.09 ± 0.32	0.09 ± 0.10	0.32 ± 0.09
	Cos.	0.90 ± 0.03	0.08 ± 0.02	0.01 ± 0.00	0.60 ± 0.13	0.36 ± 0.05
	Eucl.	1.00 ± 0.00	54.16 ± 33.24	4061.31 ± 3753.04	0.47 ± 0.22	0.33 ± 0.08
	Norm. Eucl.	1.00 ± 0.00	0.04 ± 0.01	0.00 ± 0.00	0.57 ± 0.12	0.38 ± 0.04
N24News (text)	Abs.	0.03 ± 0.04	0.53 ± 0.29	0.90 ± 0.53	-0.00 ± 0.03	0.05 ± 0.01
	Cen. Cos.	0.18 ± 0.14	0.42 ± 0.36	0.34 ± 0.44	0.15 ± 0.13	0.23 ± 0.16
	CoB	0.14 ± 0.12	0.00 ± 0.00	0.00 ± 0.00	0.10 ± 0.10	0.29 ± 0.21
	Cos.	1.00 ± 0.00	0.15 ± 0.11	0.04 ± 0.04	0.18 ± 0.17	0.15 ± 0.11
	Eucl.	0.69 ± 0.44	9.45 ± 8.39	167.88 ± 231.19	0.17 ± 0.16	0.18 ± 0.13
	Norm. Eucl.	0.71 ± 0.44	0.50 ± 0.57	0.59 ± 0.88	0.15 ± 0.13	0.24 ± 0.17
TREC	Abs.	0.01 ± 0.03	0.47 ± 0.22	0.73 ± 0.31	-0.00 ± 0.02	0.16 ± 0.07
	Cen. Cos.	0.48 ± 0.14	0.36 ± 0.11	0.21 ± 0.12	0.53 ± 0.12	0.54 ± 0.06
	CoB	0.09 ± 0.03	0.03 ± 0.01	0.00 ± 0.00	0.04 ± 0.01	0.47 ± 0.10
	Cos.	0.98 ± 0.02	0.17 ± 0.07	0.05 ± 0.04	0.57 ± 0.13	0.44 ± 0.07
	Eucl.	0.95 ± 0.02	6.67 ± 3.97	73.78 ± 75.43	0.58 ± 0.13	0.47 ± 0.06
	Norm. Eucl.	0.97 ± 0.02	0.28 ± 0.10	0.15 ± 0.10	0.53 ± 0.12	0.57 ± 0.06