ALIGNMENT-PRESERVING LOW-DIMENSIONAL PROJECTION FOR EFFICIENT VISION-LANGUAGE RETRIEVAL

Anonymous authors
Paper under double-blind review

ABSTRACT

Vision-language models produce high-dimensional embeddings that enable effective cross-modal retrieval but incur substantial storage and similarity search costs at scale. Naive post-hoc dimensionality reduction alleviates these costs but may substantially distort the cross-modal geometry required for retrieval.

We introduce **Alignment-Preserving Low-Dimensional Projection (ALP)**, a lightweight post-hoc framework that maps frozen vision-language representations into a compact shared embedding space using modality-specific projections. We further introduce **ALP+KD**, which distills the cross-modal similarity distribution of the original representation into the compressed space.

On our primary CLIP retrieval benchmark, ALP+KD compresses 512-dimensional embeddings to 128 dimensions while improving Mean Recall@1 from 0.4389 ± 0.0025 to 0.4727 ± 0.0046 across three random seeds. At the same dimensionality, PCA achieves 0.3295 ± 0.0032 and a parameter-matched Matryoshka-style baseline achieves 0.4543 ± 0.0035.

The compact embeddings also provide practical systems benefits. On a 500K-vector FAISS database, ALP+KD reduces raw vector storage by 75%, accelerates exact inner-product search by approximately $2.9\times$, and reduces index-construction time by $4.02\times$. Zero-shot cross-dataset and cross-backbone experiments further indicate that most of the original retrieval structure is retained outside the primary training setting. These results show that explicitly preserving cross-modal similarity geometry can provide a favorable accuracy–efficiency trade-off for multimodal retrieval.

1 INTRODUCTION

Vision-language models (VLMs) provide a shared representation interface between visual and textual information. CLIP (Radford et al., 2021), for example, learns normalized image and text embeddings that can be compared directly using inner products or cosine similarity. Such representations have become a practical foundation for image-to-text retrieval, text-to-image search, zero-shot recognition, and large-scale multimodal indexing.

Despite their effectiveness, embedding-based retrieval systems face an important systems constraint: representation dimensionality. For a database containing N vectors of dimensionality D, raw FP32 vector storage scales as $O(ND)$, while exact similarity computation also scales linearly with D. At large database sizes, dimensionality therefore affects memory footprint, memory bandwidth, index construction, and retrieval throughput even when the underlying VLM is completely frozen.

A natural solution is post-hoc dimensionality reduction. PCA offers a simple and inexpensive way to retain directions explaining high global variance. However, variance preservation is not necessarily equivalent to preserving the geometry needed for cross-modal retrieval. Directions with high variance need not be retrieval-critical, while low-variance directions may encode important cross-modal distinctions.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

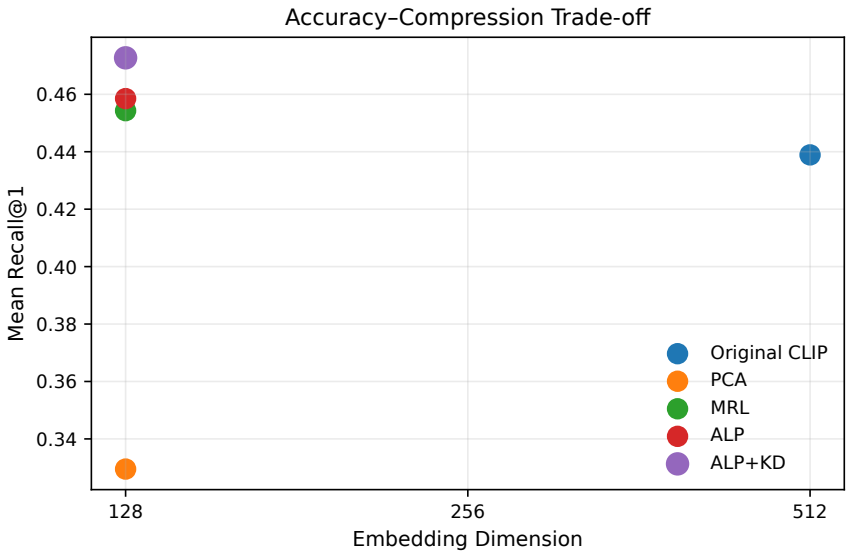

Figure 1: Accuracy–compression trade-off on the primary CLIP retrieval benchmark. ALP+KD compresses the representation from 512D to 128D while improving Mean Recall@1 over the original embedding and other 128D baselines.

This discrepancy is visible in our experiments. Reducing CLIP embeddings from 512 to 128 dimensions with PCA decreases Mean Recall@1 from 0.4389 to 0.3295, retaining only approximately 75.1% of the original performance.

This motivates our central question:

> *Can pretrained vision-language embeddings be aggressively compressed while preserving the cross-modal geometry required for retrieval?*

We introduce **Alignment-Preserving Low-Dimensional Projection (ALP)**. ALP keeps the pretrained VLM completely frozen and learns two lightweight linear projections, one for images and one for text, that map the original representations into a compact shared space. Unlike unsupervised dimensionality reduction, these projections are optimized directly for bidirectional image-text retrieval.

We further propose **ALP+KD**. Contrastive supervision primarily identifies matched image-text pairs, whereas the pretrained representation contains richer relational information in the relative similarities between all examples. ALP+KD therefore transfers the cross-modal similarity distribution of the original teacher representation into the compressed student space. This formulation preserves retrieval-relevant neighborhood structure without requiring individual student coordinates to reproduce teacher coordinates.

Across three random image-level splits, ALP+KD improves Mean Recall@1 from 0.4389 ± 0.0025 for the original 512D CLIP embedding to 0.4727 ± 0.0046 at 128 dimensions. At the same dimensionality, PCA achieves 0.3295 ± 0.0032, an alignment-only ALP variant achieves 0.4585 ± 0.0038, and a parameter-matched Matryoshka-style baseline achieves 0.4543 ± 0.0035.

These representation-level improvements translate into direct systems benefits. On a 500K-vector database, 128D embeddings require 244.14 MB rather than 976.56 MB of raw FP32 vector storage. Exact FAISS search is approximately $2.9\times$ faster, while index construction improves by $4.02\times$.

Our main contributions are:

- We introduce **ALP**, a lightweight post-hoc method for compressing frozen vision-language embeddings using modality-specific retrieval-aware projections.
- We introduce a **cross-modal similarity distillation** objective that transfers the pairwise retrieval geometry of the original embedding space to the compact representation.

- We provide controlled comparisons against PCA, alignment-only learning, and a parameter-matched Matryoshka-style baseline. ALP+KD achieves the best 128D Mean Recall@1 across all three evaluated random seeds.
- We demonstrate practical efficiency improvements: 75% lower embedding storage, approximately $2.9\times$ faster exact retrieval, and $4.02\times$ faster index construction.
- We evaluate cross-dataset and cross-backbone transfer, showing that the compressed representation preserves most of the original retrieval structure outside its primary training setting.

2 Related Work

2.1 Vision-Language Representation Learning

Contrastive vision-language pretraining provides a scalable method for learning shared image-text representations. CLIP (Radford et al., 2021) trains image and text encoders to discriminate matched image-caption pairs and demonstrates broad zero-shot transfer. Related large-scale dual-encoder approaches include ALIGN (Jia et al., 2021) and LiT (Zhai et al., 2022), which further demonstrate the effectiveness of independently encoded image and text representations for transfer and retrieval. Open-source scaling efforts such as OpenCLIP (Ilharco et al., 2021) and LAION-based training (Schuhmann et al., 2022) have made such representations widely available.

SigLIP (Zhai et al., 2023) replaces the global softmax contrastive objective with a pairwise sigmoid loss and provides another strong vision-language representation family. Our work does not modify either pretraining framework. Instead, pretrained models are treated as frozen feature extractors and we study post-hoc compression of their shared embedding spaces.

2.2 Embedding Compression

Classical dimensionality reduction methods such as PCA (Jolliffe & Cadima, 2016) reduce storage and computation by projecting high-dimensional vectors into compact subspaces. However, PCA optimizes explained variance rather than downstream retrieval quality.

Our experiments show that this distinction is substantial for multimodal representations: PCA retains much less image-text retrieval performance than a projection explicitly optimized for cross-modal alignment.

2.3 Matryoshka Representation Learning

Matryoshka Representation Learning (MRL) (Kusupati et al., 2022) learns representations whose nested prefixes remain useful at multiple dimensionalities. This allows a single representation to support multiple accuracy and efficiency budgets and has demonstrated substantial retrieval speedups. The same principle has subsequently been explored in multimodal recommendation settings (Wang et al., 2024), indicating broader interest in dimension-adaptive multimodal representations.

We construct a parameter-matched Matryoshka-style baseline using the same image and text $512 \rightarrow 128$ projection matrices as ALP and nested contrastive supervision at 32D, 64D, and 128D. This comparison isolates differences in supervision rather than differences in parameter count.

2.4 Knowledge Distillation

Knowledge distillation (Hinton et al., 2015) transfers information from a teacher into a smaller or more efficient student model. Beyond matching output probabilities, relational distillation methods preserve relationships among examples. Similarity-preserving knowledge distillation (Tung & Mori, 2019), for example, transfers pairwise representation structure.

Our method adopts this relational perspective specifically for multimodal retrieval. We transfer bidirectional image-to-text and text-to-image similarity distributions from the original VLM embedding space into the compressed representation.

3 METHOD

3.1 PROBLEM FORMULATION

Let a frozen VLM encode an image x_i and text y_i as

$$\mathbf{v}_i = f_{\text{img}}(x_i) \in \mathbb{R}^D, \qquad \mathbf{t}_i = f_{\text{txt}}(y_i) \in \mathbb{R}^D. \tag{1}$$

We seek compact representations

$$\mathbf{z}_i^v, \mathbf{z}_i^t \in \mathbb{R}^d, \qquad d \ll D, \tag{2}$$

without updating the pretrained encoders.

Our primary configuration uses

$$D = 512, \qquad d = 128. \tag{3}$$

3.2 MODALITY-SPECIFIC PROJECTION

We use separate bias-free linear projections

$$W_v, W_t \in \mathbb{R}^{d \times D}. \tag{4}$$

The projected features are

$$\bar{\mathbf{z}}_i^v = W_v \mathbf{v}_i, \qquad \bar{\mathbf{z}}_i^t = W_t \mathbf{t}_i. \tag{5}$$

After ℓ_2 normalization,

$$\mathbf{z}_i^v = \frac{\bar{\mathbf{z}}_i^v}{\|\bar{\mathbf{z}}_i^v\|_2}, \qquad \mathbf{z}_i^t = \frac{\bar{\mathbf{z}}_i^t}{\|\bar{\mathbf{z}}_i^t\|_2}. \tag{6}$$

For $512 \to 128$ compression, the two matrices contain

$$2(512)(128) = 131{,}072 \tag{7}$$

parameters, or approximately 0.5 MB in FP32.

3.3 PCA INITIALIZATION

PCA is fitted only on training embeddings. We concatenate frozen training image and text representations,

$$X_{\text{train}} = \begin{bmatrix} V_{\text{train}} \\ T_{\text{train}} \end{bmatrix}, \tag{8}$$

and use the first d PCA components as initialization:

$$W_v^{(0)} = W_t^{(0)} = W_{\text{PCA}}. \tag{9}$$

After initialization, W_v and W_t are optimized independently.

3.4 ALIGNMENT OBJECTIVE

Each training batch contains B distinct images and one sampled caption for each image.

The projected student similarity matrix is

$$A_{ij}^S = (\mathbf{z}_i^v)^\top \mathbf{z}_j^t. \tag{10}$$

With learnable scale $s = \exp(\alpha)$,

$$L_{ij}^S = sA_{ij}^S, \tag{11}$$

where s is capped at 100 and initialized from temperature 0.07.

We use symmetric contrastive learning:

$$\mathcal{L}_{\text{align}} = \frac{1}{2}\left(\mathcal{L}_{I\to T} + \mathcal{L}_{T\to I}\right). \tag{12}$$

3.5 CROSS-MODAL SIMILARITY DISTILLATION

Teacher similarities are computed directly from the normalized original representations:

$$A_{ij}^T = \mathbf{v}_i^\top \mathbf{t}_j. \tag{13}$$

Teacher and student distillation logits are

$$\widetilde{L}_{ij}^T = \frac{A_{ij}^T}{\tau_{\text{KD}}}, \qquad \widetilde{L}_{ij}^S = \frac{A_{ij}^S}{\tau_{\text{KD}}}, \tag{14}$$

with

$$\tau_{\text{KD}} = 0.07. \tag{15}$$

We form row-wise image-to-text distributions and corresponding distributions from the transposed matrices for text-to-image retrieval.

The distillation objective is

$$\mathcal{L}_{\text{KD}} = \frac{1}{2}\left[D_{\text{KL}}(P_{I\to T}^T \| P_{I\to T}^S) + D_{\text{KL}}(P_{T\to I}^T \| P_{T\to I}^S)\right]. \tag{16}$$

The complete objective is

$$\boxed{\mathcal{L} = \mathcal{L}_{\text{align}} + \lambda_{\text{KD}}\mathcal{L}_{\text{KD}}} \tag{17}$$

with $\lambda_{\text{KD}} = 1$.

3.6 MATRYOSHKA BASELINE

Our MRL baseline uses the same raw 128D projections but optimizes the prefixes

$$\mathcal{D} = \{32, 64, 128\}. \tag{18}$$

Each prefix is independently normalized and receives a contrastive loss:

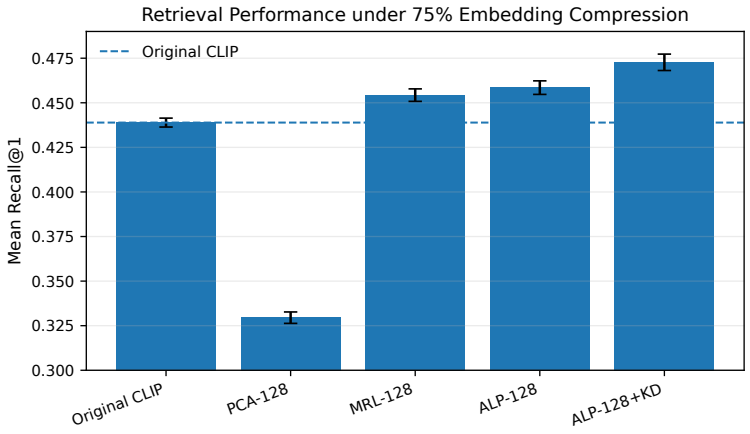

Figure 2: Main retrieval comparison. Error bars denote standard deviation across three image-level splits.

$$\mathcal{L}_{\text{MRL}} = \frac{1}{|\mathcal{D}|} \sum_{m \in \mathcal{D}} \mathcal{L}_{\text{align}}^{(m)}. \tag{19}$$

The MRL and ALP projections therefore use identical trainable parameter counts.

4 Experiments

4.1 Experimental Setup

Our primary dataset contains 18,783 COCO images and 93,950 captions. All train, validation, and test partitions are constructed at the image level to avoid caption leakage.

We report results over three split seeds:

$$\{42, 123, 2026\}. \tag{20}$$

Training uses AdamW with learning rate 3×10^{-4}, weight decay 10^{-4}, and batch size 512. We use cosine learning-rate annealing, maximum gradient norm 1.0, and early stopping with patience 12 based on validation Mean Recall@1.

4.2 Evaluation Metrics

We report image-to-text and text-to-image Recall@1, Recall@5, and Recall@10.

Our main aggregate metric is

$$\text{Mean R@1} = \frac{\text{I2T R@1} + \text{T2I R@1}}{2}. \tag{21}$$

We define retention as

$$\text{Retention} = \frac{\text{Mean R@1}_{\text{compressed}}}{\text{Mean R@1}_{\text{original}}}. \tag{22}$$

4.3 Main Retrieval Results

PCA substantially degrades the original cross-modal representation, reducing Mean Recall@1 from 0.4389 to 0.3295.

Table 1: Main retrieval results. Results are mean ± standard deviation across three random seeds.

Method	Dim	I2T R@1	I2T R@5	I2T R@10	T2I R@1	T2I R@5	T2I R@10	Mean R@1
Original CLIP	512	0.5417 ± 0.0047	0.7801 ± 0.0052	0.8628 ± 0.0071	0.3361 ± 0.0036	0.5938 ± 0.0074	0.7052 ± 0.0068	0.4389 ± 0.0025
PCA-128	128	0.4221 ± 0.0069	0.6946 ± 0.0050	0.7995 ± 0.0020	0.2370 ± 0.0025	0.5007 ± 0.0034	0.6281 ± 0.0063	0.3295 ± 0.0032
MRL-128	128	0.5207 ± 0.0038	0.7856 ± 0.0075	0.8684 ± 0.0047	0.3880 ± 0.0034	0.6729 ± 0.0051	0.7806 ± 0.0007	0.4543 ± 0.0035
ALP-128	128	0.5256 ± 0.0048	0.7852 ± 0.0048	0.8689 ± 0.0049	0.3914 ± 0.0031	0.6763 ± 0.0043	$\mathbf{0.7835 \pm 0.0016}$	0.4585 ± 0.0038
ALP-128+KD	**128**	$\mathbf{0.5521 \pm 0.0066}$	$\mathbf{0.8053 \pm 0.0045}$	$\mathbf{0.8784 \pm 0.0065}$	$\mathbf{0.3932 \pm 0.0029}$	$\mathbf{0.6776 \pm 0.0054}$	0.7823 ± 0.0042	$\mathbf{0.4727 \pm 0.0046}$

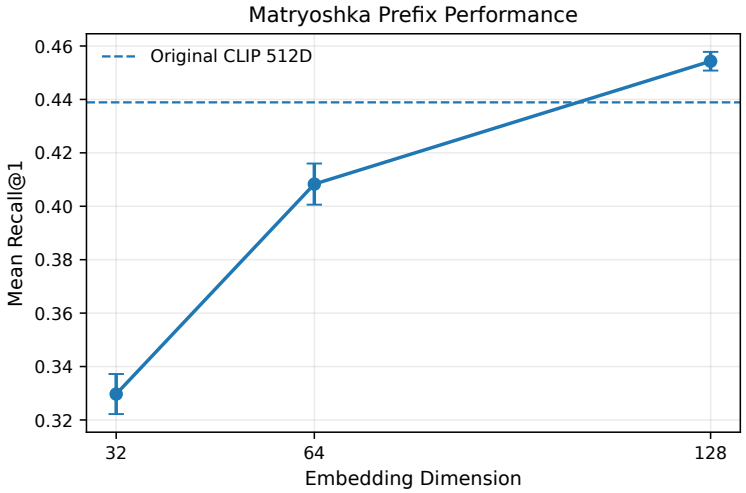

Figure 3: Retrieval performance of nested Matryoshka prefixes. Performance improves monotonically as the available dimensionality increases.

In contrast, learned low-dimensional alignment recovers this loss. ALP achieves 0.4585 Mean Recall@1, while ALP+KD reaches $\mathbf{0.4727 \pm 0.0046}$, corresponding to 107.7% retention relative to the original representation.

4.4 Comparison with Matryoshka Learning

ALP+KD exceeds the parameter-matched MRL baseline by 0.0184 Mean Recall@1.

The difference is positive for every evaluated split: $+0.0183$, $+0.0169$, and $+0.0197$ for seeds 42, 123, and 2026, respectively.

The MRL baseline reaches Mean R@1 values of 0.3297, 0.4083, and 0.4543 at 32D, 64D, and 128D, respectively. This monotonic pattern indicates that the nested baseline is functioning as intended rather than failing because of optimization instability.

4.5 Effect of Similarity Distillation

Removing KD reduces Mean Recall@1 from 0.4727 to 0.4585. Thus, similarity distillation contributes an absolute $+\mathbf{0.0142}$ Mean Recall@1 beyond low-dimensional alignment alone.

The largest improvement appears in I2T Recall@1, which increases from 0.5256 to 0.5521.

4.6 Cross-Dataset and Cross-Backbone Generalization

We freeze the projection trained on COCO and directly evaluate it on unseen retrieval distributions.

On Flickr8K, original CLIP achieves Mean Recall@1 of 0.6243 and the compressed ALP representation achieves 0.6203, corresponding to 99.36% retention.

For Visual Genome whole-image retrieval, retention is 92.74%. For region-crop-to-phrase retrieval, ALP retains 94.22%.

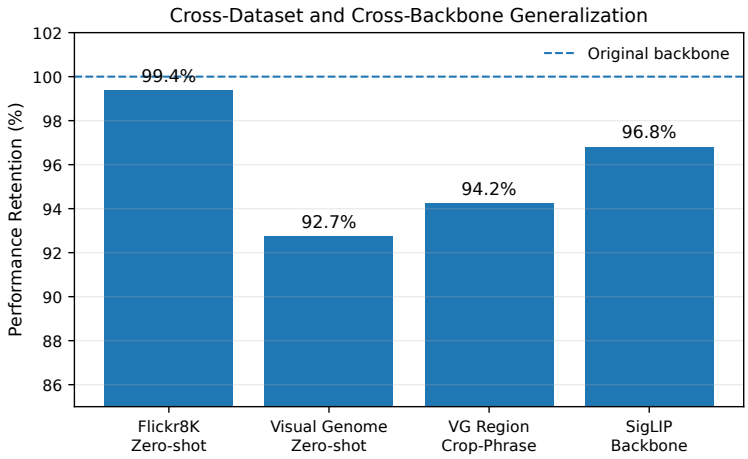

Figure 4: Retrieval-performance retention under cross-dataset and cross-backbone evaluation.

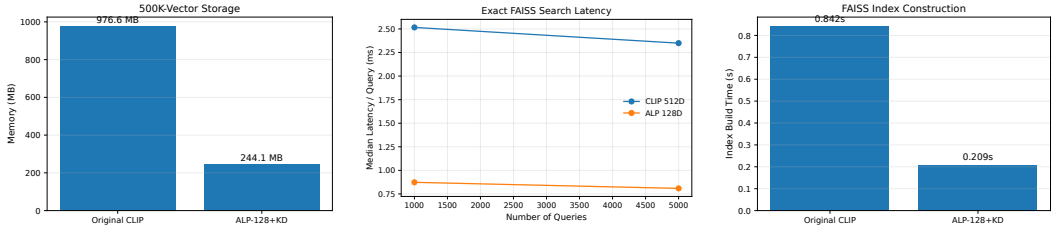

Figure 5: System efficiency from low-dimensional embeddings. Compression reduces vector storage by 75%, exact search latency by approximately $2.9\times$, and index-construction time by $4.02\times$.

We additionally apply the method to SigLIP representations. Compressing SigLIP from 768 dimensions to 128 dimensions yields 0.5726 Mean Recall@1 compared with 0.5915 for the original representation, corresponding to 96.8% retention under an 83.3% dimensionality reduction.

4.7 Large-Scale Retrieval Efficiency

We benchmark exact FAISS (Johnson et al., 2021) `IndexFlatIP` retrieval over 500,000 FP32 database vectors. Measurements use four CPU threads and five repetitions.

The 512-dimensional database requires 976.56 MB, whereas the 128-dimensional database requires only 244.14 MB.

Median index-construction time falls from 0.8417 s to 0.2095 s, a $\mathbf{4.02}\times$ speedup.

For 5,000 queries, median per-query latency falls from 2.3493 ms to 0.8087 ms, corresponding to a $\mathbf{2.905}\times$ speedup.

Search throughput consequently increases from 425.66 to 1236.55 queries/s.

The measured projection overhead is only 0.000069 ms/query, which is negligible compared with retrieval latency.

5 Discussion

A notable result is that the compressed 128D representation can outperform the original 512D CLIP embedding on the primary evaluation distribution. We interpret this as evidence that the projection performs not only compression but also retrieval-specific metric adaptation.

The original VLM representation is trained to support broad transfer rather than a particular downstream retrieval distribution. A learned low-dimensional map may therefore suppress dimensions that are irrelevant to image-text discrimination while emphasizing retrieval-relevant structure.

The poor PCA result supports this interpretation. Variance alone does not identify the directions most important for cross-modal ranking.

The additional gain from similarity distillation suggests that positive-pair contrastive supervision is also incomplete. Teacher similarity distributions encode relative semantic relationships among non-matching examples, providing a richer regularization signal for the compressed space.

6 LIMITATIONS

Our primary experiments concern global frozen image-text embeddings rather than token-level representations used internally by multimodal generative models.

Furthermore, the primary COCO result includes supervised adaptation of the projection on the training split. Accordingly, performance above the original CLIP representation should be interpreted as retrieval-specific adaptation under the evaluated distribution, not as evidence that the compressed representation universally dominates the pretrained embedding.

Although our cross-dataset experiments provide evidence of transfer, broader evaluation across additional retrieval datasets and modern VLM families would further strengthen the generality claim.

Finally, our systems experiments isolate dimensionality using exact inner-product retrieval. Approximate nearest-neighbor search, product quantization, and reduced numeric precision are complementary efficiency techniques that may provide additional gains.

7 CONCLUSION

We presented Alignment-Preserving Low-Dimensional Projection, a lightweight framework for compressing frozen vision-language embeddings while maintaining the cross-modal geometry required for retrieval.

ALP+KD combines modality-specific projections with similarity-distribution distillation from the original high-dimensional representation. On the primary CLIP experiment, the method reduces dimensionality from 512 to 128 while improving Mean Recall@1 from 0.4389 to 0.4727 across three random seeds.

The representation also provides substantial systems benefits: 75% lower raw vector storage, $4.02\times$ faster index construction, and approximately $2.9\times$ faster exact similarity search.

These results suggest that the embedding dimensionality chosen during VLM pretraining need not remain fixed during deployment. A small retrieval-aware projection can provide a substantially more efficient representation while preserving most, and under the primary evaluation setting even improving, cross-modal retrieval accuracy.

REFERENCES

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Gabriel Ilharco, Mitchell Wortsman, Ross Wightman, Cade Gordon, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, and Ludwig Schmidt. Openclip, 7 2021. URL `https://doi.org/10.5281/zenodo.5143773`.

Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc V. Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 4904–4916, 2021.

Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547, 2021. doi: 10.1109/TBDATA.2019.2921572.

Ian T. Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016. doi: 10.1098/rsta.2015.0202.

Aditya Kusupati, Gantavya Bhatt, Aniket Rege, Matthew Wallingford, Aditya Sinha, Vivek Ramanujan, William Howard-Snyder, Kaifeng Chen, Sham Kakade, Prateek Jain, and Ali Farhadi. Matryoshka representation learning. In *Advances in Neural Information Processing Systems*, volume 35, pp. 30233–30249, 2022.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 8748–8763, 2021.

Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. Laion-5b: An open large-scale dataset for training next generation image-text models. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Frederick Tung and Greg Mori. Similarity-preserving knowledge distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1365–1374, 2019.

Yueqi Wang, Zhenrui Yue, Huimin Zeng, Dong Wang, and Julian McAuley. Train once, deploy anywhere: Matryoshka representation learning for multimodal recommendation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 13461–13472, 2024. doi: 10.18653/v1/2024.findings-emnlp.786.

Xiaohua Zhai, Xiao Wang, Basil Mustafa, Andreas Steiner, Daniel Keysers, Alexander Kolesnikov, and Lucas Beyer. Lit: Zero-shot transfer with locked-image text tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18123–18133, 2022.

Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11975–11986, 2023.

AI USE STATEMENT

Generative AI tools were used to assist with code development and debugging, LaTeX formatting, language editing, and improving the clarity and organization of the manuscript. They were also used as an interactive aid during exploratory analysis and interpretation of experimental results.

All experimental designs, methodological decisions, data processing procedures, reported numerical results, and scientific claims were reviewed and verified by the authors. AI-generated code and manuscript edits were inspected and, where applicable, tested against the underlying experimental outputs.

The authors take full responsibility for the final content of this work, including all text, claims, analyses, and artifacts produced with the assistance of generative AI tools.