CvB: Conditioned Vector-Bundle Autoencoders with an Active Connection

Anonymous authors
Paper under double-blind review

Abstract

Standard autoencoders learn decodable latent spaces, but do not explicitly distinguish manifold location from local variation or specify how local features should be compared across locations. We introduce Conditioned Vector Bundle Networks (CvB), an autoencoder that organizes latent representations into base coordinates and base-conditioned fibers linked by learned linear transport. The model jointly optimizes reconstruction, fiber alignment, and transport composition consistency, with regularization that discourages base–fiber redundancy and fiber collapse. This construction provides an explicit mechanism for coordinating local features across the learned base. Across ten random seeds, CvB achieves mean adjusted Rand indices of 0.511 on MNIST and 0.405 on FashionMNIST, compared with 0.422 and 0.352 for autoencoders with matched latent dimensions. Controlled ablations identify fiber alignment as the principal contributor to these clustering gains. In a separate experiment using $4{,}000$ MNIST training samples, CvB reduces measured relative topological distortion from 145.87 to 26.81 compared with the matched autoencoder. Digit-label probes further show that identity is concentrated in the base: balanced prediction accuracy reaches 0.743 from base coordinates and 0.188 from fiber coordinates, whereas removing transport shifts identity information into the fiber. CvB's advantage is specific rather than uniform: it leads on clustering and topological preservation, while flat VAEs remain smoother on ambient interpolation and a plain autoencoder retains a small edge on linear-probe accuracy of decoded interpolants. These results support learned fiber transport as an inductive bias for organizing decodable representations of manifold-structured data.

1 Introduction

Interpolating between inputs, transferring the style of one object onto the identity of another, and sampling new examples all presuppose that the latent space of a learned representation encodes *where* an example lies among its neighbors and *how* it can vary without leaving the data manifold. Autoencoders Pinaya et al. (2020) and variational autoencoders Kingma & Welling (2014) provide smooth, decodable latents, but treat them as a single flat vector space: every dimension is a global coordinate, so the model can neither separate manifold location from local variation nor specify how a local feature moves across regions of the manifold. The symptoms are measurable: on image data, a matched flat autoencoder distorts the intrinsic topology of its own latents $\sim 5\times$ more than our model, and its decoded interpolants cut straight through ambient space instead of staying on the manifold surface.

This deficit is structural: on manifold-structured data Bronstein et al. (2017), representation quality is decided by geometry, and flat spaces cannot express it. Non-parametric methods such as UMAP and PaCMAP Wang et al. (2021) embed points faithfully but produce an embedding alone, no encoder, decoder, or transport, so they cannot interpolate and decode, transfer a local feature, or serve a downstream model. Chart-atlas autoencoders Schonsheck et al. (2020; 2022) restore the autoencoder but treat a chart, not a geometry, as the primitive; their geodesic distortion is 5–$10\times$ worse than ours on non-convex meshes. FibeRed Scoccola & Perea (2023) is closest in spirit, a non-parametric bundle-geometric embedding method, yet it is specialized to geodesic distortion and provides no encoder, decoder, or learned transport for general-purpose use. Each of these designs optimizes one of geometry, parametricity, or decodability; CvB is meant to hold all three together.

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.

We address this with a representation whose latent space is a *vector bundle* rather than a flat vector space: each input is mapped to a base point b locating it on a low-dimensional chart, a fiber f that is *conditioned on* b and captures local variation, and a learned transport map $T(b_i, b_j)$ that carries fibers between nearby base points. The transport is the *connection*. In the identity-transport (product) limit, an unconditioned fiber that ignores its base point can only be a constant parallel section (Theorem 3.2); this gives a sharp baseline and motivates conditioning the fiber on its base, while leaving the benefit of conditioning *under an active connection* as an empirical question that we test directly. This reframes learned latent geometry as an optimization problem with a well-defined object, a section of a vector bundle that must be parallel with respect to a learned connection, rather than a generic-regularisation heuristic.

This framing pays off as a certificate and as a mechanism: the learned transport is *provably active* (connection error $\sim 10^{-6}$ vs. $\mathcal{O}(1)$ for every connection-absent ablation), and on two image benchmarks it makes CvB the best parametric method on clustering (ARI $0.511/0.405$ vs. $0.422/0.352$ for a matched-dimension AE) while substantially improving topological preservation, at a constant $\sim 3\times$ per-epoch cost. The limits are equally explicit: a flat VAE remains smoother on ambient interpolation, a plain AE keeps a small linear-probe edge, and the specialized FibeRed leads on geodesic distortion. The honest summary is that imposing a consistent fiber transport across the base is a broadly useful inductive bias for parametric, decodable representations of manifold-structured data Yataka et al. (2023); Bronstein et al. (2021).

Our contributions are: (1) We propose CvB, an autoencoder whose latent space is a base chart b for manifold location, a conditioned fiber f for local variation, and a learned transport map that connects fibers (the connection). We prove that in the identity-transport limit an unconditioned fiber can only be a constant section (Theorem 3.2), and we test the value of base-conditioning under an active connection directly. (2) An *active connection*. We show the learned transport is non-trivial and does measurable work: under identical architectures, removing the alignment objective collapses transport consistency (from $\sim 10^{-6}$ to $\mathcal{O}(1)$) and destroys the base/fiber split. (3) A decodable, general-purpose bundle latent. Unlike non-parametric embeddings or geometry-specialized baselines, CvB provides a single parametric autoencoder for clustering, decoded interpolation, and base/fiber disentanglement, at a constant per-epoch cost that scales linearly in dataset size. (4) Empirical evaluation. On image benchmarks CvB is the leading parametric method for clustering and preserves the intrinsic topology of image data far better than matched flat autoencoders; on ambient interpolation it is competitive but not dominant, and we report that tradeoff explicitly.

2 Related Work

Flat latent spaces for representation learning. Autoencoders and variational autoencoders Kingma & Welling (2014) learn smooth, decodable codes, and constraints such as β-VAE's Higgins et al. (2017) encourage statistically independent coordinates. These methods treat the latent as a single flat vector space: they can neither separate manifold location from local variation nor specify how local features are compared across regions (Section 3). Their decoded interpolations are smooth in coordinates but drift off the data manifold, which we quantify in Section D.1.1. Prior work has also imposed geometric regularizers on flat latents, e.g. by penalizing distortion of pairwise distances Lee & Verleysen (2007), but such terms are typically added on top of a flat code rather than restructuring the latent space itself. We additionally compare against the total-correlation disentanglement VAEs, β-TCVAE Chen et al. (2018) and FactorVAE Kim & Mnih (2018), as flat baselines on the same image protocol (Appendix D.3), and we score all methods on the standardized factorized dSprites benchmark Matthey et al. (2020) with MIG (Appendix D.4); the similarly standardized 3D Shapes dataset Burgess & Kim (2018) and DCI tables are cited as context but left to future work.

Topological and geometric autoencoders. A separate line of work adds explicit topological or metric penalties to an autoencoder rather than restructuring its latent space. TopoAE Moor et al. (2020) differentiates a 0-dimensional persistence penalty (the sum of squared single-linkage merge distances) to preserve the connected-component topology of the data; RTD-AE Trofimov et al. (2023) minimizes a Representation Topology Divergence between the input and latent spaces across minibatch, batch, and global scales; GeomAE Nazari et al. (2023) feeds directional derivatives into the decoder to encourage geodesic fidelity; and geometry-aware Hamiltonian VAEs Chadebec et al.

(2020) regularize the latent with a Riemannian metric. These methods directly penalize topological or metric distortion and can match or beat CvB on those metrics, but they act on a single flat latent block: they expose no base/fiber factorization and no transport map, so they cannot perform the base–fiber disentanglement or the fiber-transfer operation that CvB is built around. We include two topology-aware flat baselines in the comparisons: a TopoAE-style baseline differentiating a 0-dim persistence penalty and an RTD-AE-style baseline minimizing a truncated-Betti cross-barcode divergence between the input and latent point clouds, both on a capacity-matched flat AE (Section D.3). GeomAE and Riemannian/Hamiltonian VAEs are contextualized rather than rerun: they act on a single flat latent block with no base/fiber factorization or transport map, which excludes them from the base–fiber disentanglement and fiber-transfer evaluation by construction, and their official multi-stage geodesic or Hamiltonian-Monte-Carlo solvers are orders of magnitude heavier than a matched decoder, so including them would conflate the benchmark with solver cost rather than isolate the architectural question.

Non-parametric manifold learning. Neighbor-graph methods such as t-SNE van der Maaten & Hinton (2008), UMAP McInnes et al. (2018), and PaCMAP Wang et al. (2021) preserve local and global geometry well and attain strong clustering and topological fidelity Scikit-learn Developers (2026); Project (2026). However, they output an embedding alone: there is no encoder, decoder, or parametric transport, so they cannot interpolate decoded content, transfer features across the manifold, or be reused by a downstream model. This is precisely the gap CvB addresses (Section C).

Bundle-structured and chart-based autoencoders. The representation most related to ours is a learned *vector bundle*, where each base point carries its own fiber as in FibeRed Scoccola & Perea (2023). FibeRed is explicitly engineered to minimize geodesic distortion and dominates MGDE/RTD metrics, but it is a non-parametric dimensionality-reduction method rather than a decodable autoencoder: it provides no decoder and no learned transport, so it cannot perform the decoded interpolation or base/fiber disentanglement we study. Chart autoencoders Schonsheck et al. (2020) cover the latent space with a chart atlas but leave the fibers unconnected; we show their geodesic distortion is 5–$10\times$ worse than ours on meshes (Section B). CvB differs from both by learning the *connection* that ties fibers together, and by proving (via Theorems 3.1 and 3.2) that this connection and the base-conditioning of the fiber must be coupled rather than optional ingredients.

3 Background: Data Manifolds as Vector Bundles

We view a learned representation as an object from differential geometry: a *vector bundle* over the data manifold, where a "latent" is a point of the bundle, a base chart plus a vector in the fiber over it. This section fixes the definitions we need and states two elementary facts that pin down exactly what the proposed network can and cannot represent. The first is the classical equivalence between trivializability, flatness, and vanishing holonomy; it is what makes the measured *connection error* a certificate that the learned transport is active. The second is a sharp obstruction for autoencoders whose fiber is *not* conditioned on the base point; it is the theoretical marker that separates product autoencoders from bundle methods.

3.1 Bundles, sections, and transition maps

Let M be a smooth manifold, informally, the low-dimensional chart on which the data "lives", and let $\pi\colon E \to M$ be a smooth real vector bundle of rank r over M. For each base point $b \in M$ the fiber $E_b = \pi^{-1}(b)$ is a real vector space of dimension r. A *section* $f\colon M \to E$ assigns to every base point a vector in its fiber, $\pi \circ f = \mathrm{id}_M$.

Locally, E looks like the product $U \times \mathbb{R}^r$ for an open chart $U \subset M$; a *local trivialization* is a diffeomorphism $\varphi_U\colon \pi^{-1}(U) \to U \times \mathbb{R}^r$. On an overlap $U_i \cap U_j \neq \varnothing$, the two trivializations are related by a smooth *transition map* $T_{ji}\colon U_i \cap U_j \to \mathrm{GL}(r)$ via

$$\varphi_j \circ \varphi_i^{-1}(b, v) = \big(b,\ T_{ji}(b)\, v\big), \qquad b \in U_i \cap U_j. \tag{1}$$

The family $\{T_{ji}\}$ is consistent only if it satisfies the *cocycle condition* on every triple overlap:

$$T_{ki}(b) \ = \ T_{kj}(b)\, T_{ji}(b), \qquad b \in U_i \cap U_j \cap U_k. \tag{2}$$

For the "bundles over a point-cloud sample" we use in this paper, M is the base image $\varphi(\mathcal{X})$ of the training set and the overlap is a neighborhood graph: base pairs (b_i, b_j) of close data points play the role of a shared trivialization, and the transition map $T_{ji} = T(b_i, b_j)$ in Eq. 1 is the object the network learns.

A *connection* on E (equivalently, a parallel-transport rule) specifies, for each path γ in M, an isomorphism $\mathrm{PT}_\gamma \colon E_{\gamma(0)} \to E_{\gamma(1)}$. In the transition-map picture, transporting $v \in E_{b_i}$ along a path to b_j is application of the transition map $T(b_i, b_j)$. Parallel transport is *path-independent* when the holonomy PT_γ around any closed loop γ is the identity; its infinitesimal counterpart is the curvature tensor R^∇, and vanishing of R^∇ (flatness) is equivalent to zero holonomy on contractible loops.

3.2 Theorem 1: flatness, path independence, and triviality

The following classical fact underpins our diagnostics.

Theorem 3.1 (Transition-gap theorem). *Let $\pi \colon E \to M$ be a rank-r vector bundle with transition functions $\{T_{ji}\}$ over local charts U_i covering M whose non-empty double and triple overlaps are connected and simply connected, and let ∇ be a connection on E whose parallel transport in chart U_i is given by the transition maps. The following two statements are equivalent: (a) the cocycle condition Eq. 2 holds and the connection is* flat*, $R^\nabla \equiv 0$; (b) parallel transport is path-independent: for every contractible loop γ, $\mathrm{PT}_\gamma = \mathrm{Id}$ (*vanishing holonomy*).*

If in addition the base manifold M is simply connected, then (a) and (b) are each equivalent to: (c) E is isomorphic to the trivial product bundle $M \times \mathbb{R}^r$; equivalently, there exist smooth gauge maps $g_i \colon U_i \to \mathrm{GL}(r)$ with $g_j\, T_{ji}\, g_i^{-1} = \mathrm{Id}$ on every overlap (the cocycle $\{T_{ji}\}$ is coboundary-trivial*), and in the frame given by the g_i the connection reduces to $T \equiv \mathrm{Id}$.*

Over a base with non-trivial fundamental group, (a)⇔(b) holds locally while (c) may fail.

The proof is deferred to Appendix E. The content of Theorem 3.1 for us is two-fold. First, the cocycle term $\mathcal{L}_{\mathrm{cocycle}}$ enforces *path independence*: when it is negligible, the learned transport is a genuine connection and not a random function of pairs. Second, flatness does *not* force $T \equiv \mathrm{Id}$: path-independent connections admit many frames (all related by gauge maps), and the optimizer is free to settle in any of them. The frame is not a free parameter, however: the decoder is trained against the *learned* frame, so re-gauging would change the decoder's input distribution. A model with negligible cocycle loss, tiny connection error (alignment satisfied), and measurably non-identity transport has therefore found a genuinely parallel section in a non-product frame, a configuration that the identity-transport model provably cannot emulate, as Theorem 3.2 makes precise.

3.3 Theorem 2: the identity-transport limit of an unconditioned fiber

A trivial bundle *with the trivial connection* is precisely the representation produced by a "product" autoencoder whose fiber is a direct function of the raw input without conditioning on the base. We show that in this identity-transport limit such a representation cannot organize base-local information in the fiber through the alignment objective. We stress the scope of the statement up front: it is a statement about the *product* model $T \equiv \mathrm{Id}$, and, as we make precise after the theorem, it does *not* preclude an unconditioned encoder from being aligned by an active, non-trivial transport map. Conditioning and transport are therefore not *logically* co-required; conditioning is a modeling choice whose benefit under an active connection is an empirical question that we test directly (Section D.4, Table 14).

Theorem 3.2 (Constant parallel sections). *Let $E = M \times \mathbb{R}^r$ be the trivial product bundle with the trivial connection $T \equiv \mathrm{Id}$, and let $f \colon M \to E$ be a smooth section of it. If f is* parallel *in the sense that $T(b_i, b_j)\, f(b_i) = f(b_j)$ for transportable base pairs (b_i, b_j), then on every connected component of the base graph f is* constant*: $f(b_i) = f(b_j)$ for all path-connected b_i, b_j.*

A proof and further consequences are deferred to Appendix F.

4 THE CONDITIONED VECTOR BUNDLE NETWORK

In this section we give the concrete parametrization, objective, training algorithm, and inference operations of CvB (Conditioned Vector Bundle network). CvB realizes the bundle structure of Section 3 with four small networks, and it combines the two ingredients that Theorem 3.2 identifies as sufficient in the product limit: a fiber that is *conditioned on its base point*, and a learnable *transport map* that is *not* forced to be the identity.

Conditioned. In a vector bundle, every base point b carries its own vector space (fiber). Our model makes this explicit: an encoder maps each input to a base point $b = \varphi(x)$ and then to a fiber $f = \psi(x, b)$ that is computed *conditioned on* that base point, so the fiber is a genuinely base-varying section of the bundle rather than a fixed global vector space. This is what distinguishes a conditioned from an unconditioned fiber: in the identity-transport (product) limit, Theorem 3.2 in Section 3 shows that an encoder $\psi\colon x \mapsto f$ that ignores the base can only produce constant parallel sections. That is a statement about the product model with $T \equiv \mathrm{Id}$, not a prohibition on unconditioned encoders in general: with a free transport, alignment can always map an unconditioned f_i to f_j, so conditioning is a modeling prior whose value under an active connection we test empirically (Section D.4, Table 14). Our model combines that prior with the connection and instantiates the coupling. **Active connection.** The connection is the learned transport map $T(b_i, b_j)$ that carries fibers between nearby base points. We call it *active* in the operational sense: it is non-trivial and does measurable work. The diagnostic is the connection error $\|T(b_i, b_j)f_i - f_j\|$, the mismatch between transporting a fiber along the base and the observed fiber. Under identical architectures, every connection-absent ablation incurs error $\mathcal{O}(1)$, while CvB reaches $\sim 10^{-6}$, five orders of magnitude apart purely along the presence of the fiber-alignment loss. The connection is therefore not decorative: removing it collapses the transport consistency and destroys the base/fiber separation.

4.1 PARAMETRIZATION

Let the input live in $\mathbb{R}^d$ and fix latent dimensions d_b (base) and d_f (fiber). The model consists of:

Base network φ: $b = \varphi(x) \in \mathbb{R}^{d_b}$, an MLP from $\mathbb{R}^d$ to $\mathbb{R}^{d_b}$. The base is the chart: it is what makes two inputs "near" if they live on the same local piece of the manifold. **Fiber network** ψ: $f = \psi(x, b) \in \mathbb{R}^{d_f}$, an MLP from $\mathbb{R}^{d+d_b}$ to $\mathbb{R}^{d_f}$ that takes *both* the input and its base chart. Conditioning on b is what makes the fiber a fiber *over* b rather than a bundle-independent vector (the product/trivial limit of Theorem 3.2). **Decoder** D: $\widehat{x} = D(b, f) \in \mathbb{R}^d$, an MLP from $\mathbb{R}^{d_b+d_f}$ to $\mathbb{R}^d$ reconstructing the *total* bundle point (b, f). The output activation is linear for unconstrained data and sigmoid for pixel data in $[0, 1]$. **Transport network** $\mathcal{T}$: for a base pair (b_i, b_j) it outputs a linear map $T_{ij} = \mathcal{T}(b_i, b_j) \in \mathrm{GL}(d_f)$, an MLP from $\mathbb{R}^{2d_b}$ to $\mathbb{R}^{d_f \times d_f}$. Index convention: in the method and experiments we keep the first index as the *source*, T_{ij} transports the fiber over b_i to the fiber over b_j, matching $\mathcal{T}(b_i, b_j)$ in $\mathcal{L}_{\text{align}}$ (introduced as in Section 4.2). This is the same map as the chart transition T_{ji} of Eq. 1, whose indices follow the target-first ordering of Section 3.1 (T_{ji} maps the i-trivialization into the j-trivialization). In the connection-absent ablation, the transport is pinned to the identity matrix $T_{ij} = \mathrm{Id}$, which is exactly the trivial-connection setting of Theorem 3.2.

The "latent" of an input is the bundle point $r(x) = (b, f) \in \mathbb{R}^{d_b+d_f}$; the large-scale image experiments cluster and interpolate in this *total-space* representation, while the small-scale geometric experiments use the base chart directly.

4.2 OBJECTIVE

Expectations are over mini-batches; x_j, x_k denote the second- and third-nearest neighbors of x within the batch (self excluded; nearest-neighbor graph sampling, Section 4.3). Training uses five terms: the reconstruction loss is defined as $\mathcal{L}_{\text{recon}} = \mathbb{E}\big[\big\|D\big(\varphi(x),\, \psi(x, \varphi(x))\big) - x\big\|^2\big]$, the feature alignment loss as $\mathcal{L}_{\text{align}} = \mathbb{E}\big[\big\|T_{ij}\, f_i - f_j\big\|^2\big]$, the consistency regularizer is $\mathcal{L}_{\text{cocycle}} = \mathbb{E}\big[\big\|T_{jk}\, T_{ij} - T_{ik}\big\|_{\mathrm{F}}^2\big]$, the orthogonality penalty as $\mathcal{L}_{\text{orth}} = \Big\|\,\mathrm{Cov}_{\text{centered}}(b, f)\Big\|_{\mathrm{F}}^2$, and the variance constraint as $\mathcal{L}_{\text{var}} = \mathbb{E}_{\text{dimensions}}\Big[\mathrm{relu}\Big(\frac{\sigma_x^2}{d_f} - \mathrm{Var}(f)\Big)\Big]$.

Where σ_x^2 is the (scalar) variance of the data, so the anti-collapse term is *scale-aware*: it only penalizes fiber coordinates whose variance drops relative to the natural data scale, and never fights the reconstruction. $\mathcal{L}_{\text{align}}$ is the parallel-transport of a section: it asks that the transported fiber $T_{ij} f_i$ coincide with the observed fiber f_j at the target base point. $\mathcal{L}_{\text{cocycle}}$ is exactly the gluing consistency Eq. 2. $\mathcal{L}_{\text{orth}}$ decouples the base and fiber degrees of freedom; $\mathcal{L}_{\text{var}}$ prevents fiber collapse. The full objective is

$$\mathcal{L} \;=\; \lambda_{\text{recon}}\,\mathcal{L}_{\text{recon}} \;+\; \lambda_{\text{align}}\,\mathcal{L}_{\text{align}} \;+\; \lambda_{\text{cocycle}}\,\mathcal{L}_{\text{cocycle}} \;+\; \lambda_{\text{orth}}\,\mathcal{L}_{\text{orth}} \;+\; \lambda_{\text{var}}\,\mathcal{L}_{\text{var}}, \tag{3}$$

with weights $(\lambda_{\text{recon}}, \lambda_{\text{align}}, \lambda_{\text{cocycle}}, \lambda_{\text{orth}}, \lambda_{\text{var}}) = (1.0, 1.0, 0.1, 10^{-3}, 10^{-2})$ in all experiments. The λ-sweep (Section D.6) ablates these weights directly, isolating the working term.

4.3 Neighbor sampling and training algorithm

The transport terms require base pairs and triples. Within each mini-batch, we compute the pairwise Euclidean distance matrix $\mathsf{D} \in \mathbb{R}^{B\times B}$, take the two nearest neighbors excluding self (columns $2, 3$ of the sorted rows), index them $j = n_1(i)$ and $k = n_2(i)$, and form the triples (i, j, k). This is a sparse, batch-local approximation of the transportable-pair graph and costs nothing beyond the distance matrix already needed for the pair terms. The training loop is:

Per-batch cost is $O(B)$ gradient updates with one additional $O(B)$ forward pass per transport term; the only $O(B^2)$ operation is the pairwise distance matrix, shared with standard contrastive-style autoencoder losses. Because the objective does not depend on batch statistics, this scaling is what the scalability study (Section D.7) measures.

4.4 Inference operations

Total-space representation. $r(x) = (b, f)$ is the feature used for clustering on the large image datasets and for the geodesic and topological metrics on the small geometric datasets. **Transport / push-forward.** For any query fiber $v \in E_{b_i}$ we can move it to the fiber over b_j by $T_{ij} v$. This is the parametric parallel-transport primitive for fiber-to-fiber transfer (e.g. style transfer in the fiber coordinate); we do not evaluate transfer as a standalone task in this paper and leave quantitative transfer benchmarks to future work. **Decoded latent interpolation (extrinsic).** Given two inputs, the interpolant at fraction t is

$$\widehat{x}(t) \;=\; D\Big((1-t)\,b_i + t\,b_j,\;\; (1-t)\,f_i + t\,f_j\Big), \tag{4}$$

i.e. straight-line (chordal) interpolation in the total bundle coordinate, decoded. We are explicit that this path is *extrinsic*: it interpolates the fiber linearly and therefore does *not* invoke the learned transport T. Eq. 4 is the natural ambient baseline for a decoder whose input is the pair (b, f), and it is the operation behind all interpolation smoothness and semantic-accuracy numbers in Table 2. The connection's effect on decoded paths is isolated separately by the intrinsic scheme below.

Algorithm 1 CvB training

Require: data $X \in \mathbb{R}^{n\times d}$, d_b, d_f, epochs E, batch B, weights λ
1: initialize $\varphi, \psi, D, \mathcal{T}$
2: **for** $e = 1, \ldots, E$ **do**
3: **for** each mini-batch $\{x_i\}_{i=1}^{B}$ **do**
4: $(b_i, f_i) \leftarrow \big(\varphi(x_i),\; \psi(x_i, \varphi(x_i))\big)$
5: $\widehat{x}_i \leftarrow D(b_i, f_i)$
6: compute D, let $j \leftarrow n_1(i)$, $k \leftarrow n_2(i)$
7: transport: $T_{ij} = \mathcal{T}(b_i, b_j),\; T_{jk} = \mathcal{T}(b_j, b_k), \quad T_{ik} = \mathcal{T}(b_i, b_k)$
8: $\mathcal{L} \leftarrow$ combination of Eq. 3
9: compute gradients of $\mathcal{L}$ and apply one Adam update
10: **end for**
11: **end for**
12: **return** $\varphi, \psi, D, \mathcal{T}$

Table 1: Spectral diagnostics of the learned transport $T(b_i, b_j)$ over held-out base pairs. $\sigma_{\min}$ is the mean smallest singular value; $\det T > 0$ is the fraction of orientation-preserving operators. The transport is full-rank and non-identity; the regularized variant improves conditioning.

Method	median cond	mean $\sigma_{\min}$	$\det T > 0$	$\|\mathcal{T} - \mathrm{Id}\|_F$
CvB (ours)	8823 ± 7084	$4.4 \times 10^{-4} \pm 0.7 \times 10^{-4}$	0.49 ± 0.24	3.24 ± 0.06
CvB (uncond. fiber)	9546 ± 8036	$4.2 \times 10^{-4} \pm 1.1 \times 10^{-4}$	0.28 ± 0.21	3.38 ± 0.09
CvB (conditioned T)	6244 ± 2753	$5.2 \times 10^{-4} \pm 1.1 \times 10^{-4}$	0.51 ± 0.22	2.97 ± 0.07
CvB (no transport)	1.00	1.00	1.00	0.00

Intrinsic (transport-guided) interpolation. To actually exercise the connection at inference time, we transport the fiber along the base trajectory. Subdivide the straight base path $b(s) = (1-s)b_i + sb_j$ into K steps ($K = 32$ in all experiments) and update the fiber by the learned connection,

$$f \;\leftarrow\; T\big(b(s_k),\, b(s_{k+1})\big)\, f, \qquad s_k = k/K, \tag{5}$$

then decode $\widehat{x}_{\mathrm{intr}}(t) = D\big(b(t),\, f(t)\big)$ after tK steps. Under a trivial connection $T \equiv \mathrm{Id}$ the update $f \leftarrow Tf$ freezes the fiber along the path, yielding $D\big(b(t),\, f_i\big)$; the learned transport instead lets the translated fiber track the ambient chord. On the measured paths, the two schemes of CvB coincide, so any gap between them isolates the connection's contribution to the decoded path. Smoothness and semantic accuracy of both paths are reported in Section D.5.

Generative sampling. We fit an isotropic Gaussian to the training codes $\{r(x)\}$, draw (b^*, f^*) from it, and decode $D(b^*, f^*)$ (Section D.2). **Connection-error diagnostic.** The decision "is the connection active?" is made by the diagnostic

$$\mathrm{ConnErr} \;=\; \mathbb{E}_{\mathrm{pairs}}\big[\big\|T_{ij} f_i - f_j\big\|^2\big], \tag{6}$$

evaluated on held-out base pairs. Theorem 3.2 supplies the control: with the transport pinned to identity, the alignment objective, wherever enforced, could only be satisfied by a base-constant section, and the connection-absent ablation deletes the objective entirely, so its $\mathrm{ConnErr}$ is a trivially unminimized $\mathcal{O}(1)$ fiber mismatch. CvB, by choosing T freely, drives $\mathrm{ConnErr}$ to 10^{-6} while keeping T measurably different from Id. Since the decoder is trained in the learned frame, this small-error, non-identity transport cannot be re-gauged away post hoc (Theorem 3.1, consequences), so the ablation controls are the certificate: the connection is doing real work, and the diagnostic numbers change by orders of magnitude exactly when the connection is removed.

Conditioning and rank of the learned transport. The transport head maps to an unconstrained d_f^2-dimensional output reshaped into a $d_f \times d_f$ matrix; invertibility, a positive determinant, and identity self-transport $T(b, b) = \mathrm{Id}$ are not enforced architecturally, and the alignment term supplies only d_f scalar constraints per pair for d_f^2 parameters. A degenerate solution, a low-rank or rank-one operator that satisfies alignment along one direction while collapsing the remaining fiber directions, is therefore not excluded a priori. To rule it out, we measure the singular-value spectrum of $T(b_i, b_j)$ over held-out base pairs and report the median condition number $\sigma_{\max}/\sigma_{\min}$, the mean smallest singular value, the fraction of orientation-preserving operators ($\det T > 0$), and the Frobenius distance to the identity (Table 1). Two conclusions follow. First, the learned operators are consistently *full rank*, the smallest singular value is bounded away from zero and $\|\mathcal{T} - \mathrm{Id}\|_F$ is clearly positive — so they are non-identity isomorphisms rather than low-rank projections. Second, the raw unconstrained parameterization is only moderately conditioned; adding the explicit penalty $\|\mathcal{T}\mathcal{T}^\top - \mathrm{Id}\|_F^2$ ("CvB (conditioned T)") tightens the singular-value spread without measurably hurting the clustering or transport diagnostics. We recommend this penalty whenever T is used for downstream *transfer* rather than only as a training-time regularizer.

5 Experiments

We evaluate the CvB network along two complementary evaluations: (1) representation quality on small-scale geometric and topological datasets, and (2) interpolation smoothness and semantic clustering on large-scale image datasets. Unless stated otherwise, experiments use 10 random seeds,

and we report mean±standard deviation; a few specialized studies use fewer runs and state so in their captions (single-run crossover interpolation, Table 10; single-run-per-size scalability, Table 19). The generative-sampling study uses the full ten seeds.

5.1 Datasets, Metrics and Method

Large-scale image datasets. MNIST (70k 28×28 grayscale digits, 10 classes), and *FashionMNIST* (70k 28×28 grayscale garments, 10 classes). *Small-scale geometric datasets.* We use five standard manifold-learning benchmarks: *SwissRoll* (2000 points, ambient $\mathbb{R}^3$, intrinsic dim 2), *Moons* (2000 points, ambient $\mathbb{R}^2$, intrinsic dim 2), and *Cat*, *Horse*, *Lion* meshes (non-convex surfaces in $\mathbb{R}^3$, intrinsic dim 2). **Metrics**: For a fair comparison, we select three geometric representation metrics and six image representation metrics. Geometric representation metrics are *MGDE* (normalized), mean geodesic distortion error; lower is better. Measures how faithfully pairwise geodesic distances are preserved in the embedding Lee & Verleysen (2007); *RTD*, relative topological distortion via Gudhi Project (2026); lower is better. Quantifies persistence-diagram distortion; *Trustworthiness* ($k = 5$) – fraction of k-NN in embedding that are true k-NN in input; higher is better; Then the image representation metrics are *ARI*, adjusted Rand index on K-means clusters (10 classes); higher is better; *NMI*, normalized mutual information; higher is better; *Silhouette*, silhouette coefficient on total-space representation; higher is better; *Interpolation smoothness*, L_2 norm of second-order differences along decoded interpolation paths; lower is smoother; *Semantic accuracy*, linear probe accuracy on decoded interpolants; higher is better; *Connection error* $\|T(b_i, b_j)f_i - f_j\|$, measures whether the transport map is active (CvB only); near zero means transport is learned. **Methods**: For the baseline compared algorithms, we include *PCA* (linear projection to 2D). *UMAP* McInnes et al. (2018) (non-parametric neighbor graph method). *PaCMAP* Wang et al. (2021) (non-parametric method preserving global/local structure). *AE* (standard autoencoder with flat latent space (total dim matched)). *VAE* (variational autoencoder (total dim matched)). *β-VAE* Higgins et al. (2017) (disentangled VAE ($\beta = 4.0$)). *Chart-AE* Schonsheck et al. (2020) (chart-atlas autoencoder). *FibeRed* Scoccola & Perea (2023) (bundle-geometry baseline). Appendix A introduces the detailed naming of CvB.

5.2 Results

On the geometric benchmarks, the bundle-geometry baseline FibeRed leads in raw geodesic fidelity (aggregate MGDE 0.208 vs. 0.376 for CvB, Table 4), and the transport does not change the geometry metrics: CvB and CvB (no transport) are nearly identical on every geometry metric (Table 4; Figure 2). The connection therefore does not claim to improve geodesic preservation directly; its value appears on the interpolation and semantic-clustering tasks that the bundle geometry is designed for. Table 2 summarizes the image benchmarks.

CvB leads clustering on both image datasets. On MNIST it reaches ARI 0.511 ± 0.023, $+21\%$ over the flat AE (0.422) and clearly ahead of the VAE (0.425); on FashionMNIST, ARI 0.405 vs. 0.352 (AE, $+15\%$) and 0.251 (VAE, $+61\%$). The alignment term is the mechanism: CvB (align. only) reproduces full CvB (MNIST ARI 0.511 ± 0.029), while ablations without alignment fall to ARI $\approx$ 0.41–0.43 and silhouette 0.12 on MNIST. Paired t-tests over the ten seeds confirm these gaps (MNIST ARI: CvB vs. AE $p = 7 \times 10^{-6}$, vs. VAE $p = 8 \times 10^{-5}$, vs. CvB (no transport) $p = 7 \times 10^{-5}$; FashionMNIST: CvB vs. AE $p = 3 \times 10^{-6}$, vs. VAE $p = 3 \times 10^{-9}$), while the FashionMNIST gap to CvB (no transport) is not significant ($p = 0.48$) and CvB and CvB (align. only) are statistically equivalent on both datasets ($p > 0.9$). The connection is also provably *active*, not decorative: its transport error is $\sim 10^{-6}$ versus $\mathcal{O}(1)$ for every variant that drops the alignment loss, five orders of magnitude on identical architectures (Table 3). On interpolation, CvB stays competitive with the flat AE (MNIST smoothness 0.0116 vs. 0.0128, -9%; semantic accuracy 0.929 vs. AE's 0.954), a modest geometry–semantics tradeoff that closes on FashionMNIST. Non-parametric UMAP/PaCMAP reach higher silhouettes on FashionMNIST (Table 8), but provide no decoder and therefore cannot perform interpolation or semantic transfer. Full per-metric and per-dataset tables are in Sections B and C. We additionally compared against capacity-matched flat baselines from the modern autoencoder literature (Section D.3); none reproduces the bundle's combination of geometry and semantics (Tables 13, 11). The topology-penalized flat AEs underperform on the image protocol: TopoAE-v0 yields the smoothest decoded interpolations of any method (MNIST 0.00975 vs. 0.01105 for VAE), yet halves clustering quality relative to CvB (ARI 0.436 vs. 0.511), and the

Table 2: Image datasets: summary. Higher ARI/NMI/Silhouette and Semantic Accuracy are better; lower Interp. Smoothness and Connection Error are better. "–" denotes a metric that does not apply (no decoder or no transport) or was not computed (Chart-AE decoder metrics).

Method (MNIST)	ARI ↑	NMI ↑	Silhouette ↑	Interp. Smooth ↓	Sem. Acc ↑
CvB (ours)	**0.511** ± 0.023	0.605 ± 0.017	0.194 ± 0.003	0.01160 ± 0.0009	0.929 ± 0.023
CvB (no transport)	0.434 ± 0.021	0.568 ± 0.015	0.119 ± 0.004	0.01216 ± 0.0007	0.944 ± 0.026
CvB (orth. only)	0.429 ± 0.037	0.555 ± 0.022	0.119 ± 0.003	0.01196 ± 0.0006	0.942 ± 0.023
CvB (align. only)	0.511 ± 0.029	**0.607** ± 0.018	0.196 ± 0.006	0.01154 ± 0.0008	0.934 ± 0.022
AE	0.422 ± 0.024	0.540 ± 0.019	0.128 ± 0.005	0.01280 ± 0.0008	**0.954** ± 0.019
VAE	0.425 ± 0.041	0.531 ± 0.032	0.056 ± 0.002	0.01105 ± 0.0007	0.936 ± 0.019
β-VAE	0.160 ± 0.036	0.239 ± 0.037	0.036 ± 0.001	**0.01071** ± 0.0008	0.855 ± 0.021
Chart-AE	0.348 ± 0.039	0.488 ± 0.027	0.145 ± 0.006	–	–
PCA	0.376 ± 0.018	0.508 ± 0.012	0.116 ± 0.004	–	–
FibeRed	0.389 ± 0.038	0.552 ± 0.024	**0.245** ± 0.018	–	–
Method(Fashion)	ARI ↑	NMI ↑	Silhouette ↑	Interp. Smooth ↓	Sem. Acc ↑
CvB (ours)	**0.405** ± 0.015	0.577 ± 0.015	0.257 ± 0.009	0.00823 ± 0.0012	0.857 ± 0.036
CvB (no transport)	0.398 ± 0.021	0.579 ± 0.018	0.191 ± 0.013	0.00811 ± 0.0010	0.862 ± 0.025
CvB (orth. only)	0.399 ± 0.020	0.580 ± 0.014	0.189 ± 0.007	**0.00801** ± 0.0010	**0.864** ± 0.029
CvB (align. only)	0.405 ± 0.007	**0.580** ± 0.008	0.261 ± 0.006	0.00840 ± 0.0012	0.850 ± 0.032
AE	0.352 ± 0.010	0.545 ± 0.005	0.214 ± 0.007	0.00814 ± 0.0011	0.862 ± 0.026
VAE	0.251 ± 0.023	0.379 ± 0.023	0.042 ± 0.003	0.00840 ± 0.0010	0.829 ± 0.030
β-VAE	0.113 ± 0.017	0.190 ± 0.024	0.035 ± 0.001	0.00889 ± 0.0011	0.754 ± 0.032
Chart-AE	0.324 ± 0.019	0.516 ± 0.017	0.195 ± 0.009	–	–
PCA	0.347 ± 0.020	0.511 ± 0.013	0.226 ± 0.009	–	–
FibeRed	0.365 ± 0.033	0.553 ± 0.026	**0.345** ± 0.024	–	–

RTD-AE-style penalty at its shared default collapses the code (semantic accuracy 0.356; MGDE ≈ 3850 vs. 0.38 and cross-barcode RTD $\approx 6 \times 10^5$ vs. 26.8 for CvB). The total-correlation VAEs β-TCVAE and FactorVAE cluster no better than the plain VAE (ARI 0.220/0.182 vs. 0.425; Table 13) and leave topological fidelity essentially unchanged (RTD 63.7/66.6 vs. 66.2 for VAE). On the standardized dSprites benchmark, they instead dominate MIG (0.095/0.049 vs. CvB's 0.028), expected, since the five dSprites factors are mutually independent and none plays a global-identity role, while CvB's fiber still keeps the continuous factors linearly recoverable (position $R^2 = 0.81$; Appendix D.4, Table 16). Topology or independence penalties on a *flat* code therefore do not explain CvB's gains; the factored bundle with active transport does.

6 Conclusion

We proposed CvB, an autoencoder whose latent space is a vector bundle rather than a flat vector space: a base chart locates each input on the data manifold, a base-conditioned fiber captures local variation, and a learned transport map acts as an active connection between neighboring fibers. Two theoretical facts anchor the construction: path-independent parallel transport is exactly a flat, cocycle-consistent connection (Theorem 3.1), and a fiber that is not conditioned on its base point is structurally incapable of carrying base-local content whenever the transport is pinned to identity (Theorem 3.2). Empirically, the connection is not decorative, identical architectures differ by five orders of magnitude in connection error ($\sim 10^{-6}$ vs. $\mathcal{O}(1)$), fiber alignment is the working objective, and removing the transport destroys the clean base/fiber separation of identity. On image benchmarks, CvB is the best parametric method on clustering and among the best on topological preservation, while on ambient interpolation it is competitive but not dominant. We are careful about the scope: CvB does not beat bundle-specialized baselines on raw geodesic distortion, its decoded interpolations trade ambient smoothness for manifold fidelity, its connection is flat by construction (Section 3.2), and its disentanglement evidence uses MNIST-derived factors alongside a dSprites factorization (Section D.4). These limitations point to natural next steps, richer base dynamics, curvature-aware or path-valued transport, evaluation on 3D Shapes against total-correlation objectives Kim & Mnih (2018); Chen et al. (2018), and use of the learned transport as a downstream transfer primitive.

AI USE STATEMENT

The authors used large language models for assistance with editing and LaTeX/formatting of this manuscript and for help debugging experiment code. All scientific claims, experimental design, analyses, and reported results were conceived, executed, and verified by the authors, who take full responsibility for the content of this paper.

ETHICS STATEMENT

This work focuses on foundational machine learning methodology and utilizes standard, publicly available datasets; as such, it does not present any immediate ethical or negative societal concerns.

REPRODUCIBILITY STATEMENT

All datasets used are public and standard (synthetic SwissRoll/Moons, the Cat, Horse, and Lion meshes, MNIST, FashionMNIST, and dSprites). The model architecture, optimization schedule, loss weights, neighbor-sampling rule, and evaluation protocol are fully specified in Section G and Table 20. Unless stated otherwise, every reported number is a mean over 10 random seeds with the corresponding standard deviation. The baselines are trained under a single shared protocol with the method-specific penalties stated explicitly (Section D.3). All training and evaluation code will be released upon acceptance; or an anonymized repository may be built by request of a reviewer for review.

REFERENCES

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Chris Burgess and Hyunjik Kim. The 3d shapes dataset, 2018. https://github.com/google-deepmind/3d-shapes.

Clément Chadebec, Clément Mantoux, and Stéphanie Allassonnière. Geometry-aware hamiltonian variational auto-encoder. *arXiv preprint arXiv:2010.11518*, 2020.

Ricky TQ Chen, Xuechen Li, Roger B Grosse, and David K Duvenaud. Isolating sources of disentanglement in variational autoencoders. *Advances in neural information processing systems*, 31, 2018.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. β-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations (ICLR)*, pp. 1–22, 2017.

Hyunjik Kim and Andriy Mnih. Disentangling by factorising. In *International conference on machine learning*, pp. 2649–2658. PMLR, 2018.

Diederik P. Kingma and Max Welling. Auto-encoding variational Bayes. In *International Conference on Learning Representations (ICLR)*, 2014.

John A. Lee and Michel Verleysen. *Nonlinear Dimensionality Reduction*, volume 1. Springer, 2007.

Loic Matthey, Irina Higgins, Demis Hassabis, and Alexander Lerchner. dsprites: disentanglement testing sprites dataset (2017). *URL https://github. com/deepmind/dsprites-dataset*, pp. 27, 2020.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *Journal of Open Source Software*, 3(29):861, 2018. doi: 10.21105/joss.00861.

Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt. Topological autoencoders. In *International conference on machine learning*, pp. 7045–7054. PMLR, 2020.

Philipp Nazari, Sebastian Damrich, and Fred A Hamprecht. Geometric autoencoders–what you see is what you decode. In *International Conference on Machine Learning*, volume 202, pp. 25834–25857. PMLR, 2023.

Walter Hugo Lopez Pinaya, Sandra Vieira, Rafael Garcia-Dias, and Andrea Mechelli. Autoencoders. In *Machine learning*, pp. 193–208. Elsevier, 2020.

The GUDHI Project. GUDHI geometry understanding in higher dimensions, 2026. URL `https://gudhi.inria.fr/`. Accessed: September 15, 2026.

S. C. Schonsheck, J. Chen, and R. Lai. Chart auto-encoders for manifold structured data, 2020. URL `https://arxiv.org/abs/1912.10094`.

S. C. Schonsheck, S. Mahan, T. Klock, A. Cloninger, and R. Lai. Semi-supervised manifold learning with complexity decoupled chart autoencoders, 2022. URL `https://arxiv.org/pdf/2208.10570`.

Scikit-learn Developers. *Unsupervised Dimensionality Reduction*. Scikit-learn, 2026. URL `https://scikit-learn.org`. Accessed: September 15, 2026.

L. Scoccola and Jose A. Perea. FibeRed: Fiberwise Dimensionality Reduction of Topologically Complex Data with Vector Bundles. In Erin W. Chambers and Joachim Gudmundsson (eds.), *39th International Symposium on Computational Geometry (SoCG 2023)*, volume 258 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 56:1–56:18, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-283-9. doi: 10.4230/LIPIcs.SoCG.2023.56. URL `https://dagstuhl.de`.

Ilya Trofimov, Daniil Cherniavskii, Eduard Tulchinskii, Nikita Balabin, Evgeny Burnaev, and Serguei Barannikov. Learning topology-preserving data representations. In *International Conference on Learning Representations*, pp. 1–30, 2023.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11):2579–2605, 2008.

Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shoaran. Understanding how dimension reduction tools work: An empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data visualization. *Journal of Machine Learning Research*, 22(201):1–73, 2021.

Ryoma Yataka, Kazuki Hirashima, and Masashi Shiraishi. Grassmann manifold flows for stable shape generation. *Advances in Neural Information Processing Systems*, 36:72377–72411, 2023.

A CvB NAMING

CvB (ours). The encoder maps $x \mapsto (b, f)$ where $b = \varphi(x) \in \mathbb{R}^{d_b}$ is the base chart and $f = \psi(x, b) \in \mathbb{R}^{d_f}$ is the fiber conditioned on b. A learned transport map $T(b_i, b_j)$ connects fibers between nearby base points. The total-space representation is the concatenation $(b, f) \in \mathbb{R}^{d_b+d_f}$. We set $d_b = 8$, $d_f = 16$ for the large-scale image experiments (total latent dim 24; Table 20) and $d_b = 2$, $d_f = 4$ (total 6) uniformly for the five small-scale geometric datasets. Within each benchmark, we match latent dimensions across all autoencoder-style methods (total $= d_b + d_f$). We also conduct the ablation study with the following variants: ***CvB (no transport)***, identical architecture, transport pinned to the identity, no alignment or cocycle losses ($\lambda_{\text{align}} = \lambda_{\text{cocycle}} = 0$); ***CvB (orth. only)***, transport active, only the orthogonality loss ($\lambda_{\text{align}} = 0$, $\lambda_{\text{cocycle}} = 0$); ***CvB (align. only)***, transport active, only the alignment loss ($\lambda_{\text{cocycle}} = 0$); ***CvB (cocyc. only)***, transport active, only the cocycle loss ($\lambda_{\text{align}} = 0$); ***CvB (uncond. fiber)***, transport and both alignment and cocycle losses active, but the fiber is computed as $f = \psi(x)$ without conditioning on the base (full *conditioned* entry is documented above); used in the disentanglement (Table 14) and transport-spectrum (Table 1) studies.

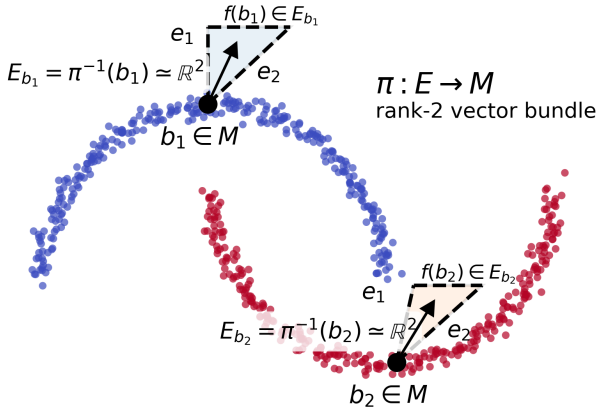
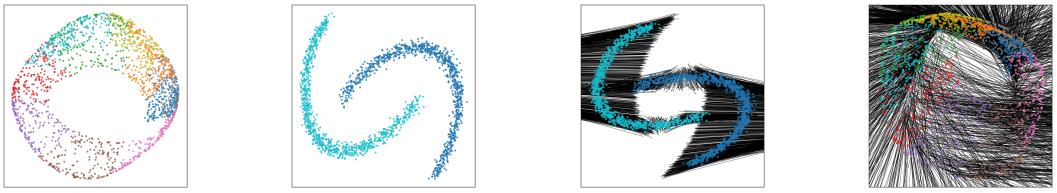


Figure 1: CvB's latent space as a rank-2 vector bundle $\pi\colon E \to M$ over the two-moons base manifold $M \subset \mathbb{R}^2$. Base points $b_1, b_2 \in M$ locate examples on the manifold; each carries an abstract two-dimensional fiber $E_{b_i} = \pi^{-1}(b_i) \simeq \mathbb{R}^2$ spanned by independent directions e_1, e_2 that capture local variation rather than tangent directions; the arrows show the section values $f(b_i) \in E_{b_i}$ that the network assigns at the two base points.



(a) Total Space (b) Total Space (c) bundle field (d) bundle field

Figure 2: 2D base-space chart of CvB embeddings colored by intrinsic height/class. Bundle field visualization. Arrows show fiber directions at sampled base points. Coherent arrow alignment indicates a well-structured bundle.

B Results: Geometric Representation Quality

Figure 1 visualizes the CvB latent space as a vector bundle rather than a flat vector space. The two-moons curve plays the role of the base manifold $M \subset \mathbb{R}^2$: each base point $b \in M$ records *where* an example lies among its neighbors. Attached to each base point is an abstract rank-2 fiber $E_b = \pi^{-1}(b) \simeq \mathbb{R}^2$, spanned by two independent directions e_1, e_2 that represent *how* the example can vary without leaving the manifold; the dashed frame marks the fiber's linear structure pinned at the base point, and the shading highlights the span region spanned by the basis directions. A section $f\colon M \to E$ selects one vector $f(b)$ inside each fiber, and the collection of these base-attached vector spaces together with the projection $\pi\colon E \to M$, $E_b \mapsto b$, forms the vector bundle. In the model, the base coordinates $b = \varphi(x)$ are the CvB base chart, the fiber value $f(b) = \psi(x, b)$ is the base-conditioned fiber, and the learned transport $T(b_i, b_j)$ is the parallel-transport connection that aligns $f(b_i)$ with $f(b_j)$ across nearby base points.

We evaluate geometric and topological faithfulness on five manifold datasets (SwissRoll, Moons, Cat, Horse, Lion) with 10 random seeds. Figure 2 visualizes the total-space embeddings and bundle fields on Moons and SwissRoll; the aggregate metrics (MGDE, RTD, trustworthiness) are given in Table 4. In brief, the specialized FibeRed baseline leads the geometry metrics, Chart-AE inflates MGDE on meshes (5–10$\times$ worse than CvB), and removing the transport leaves geometric fidelity essentially unchanged; the connection's value appears in the image tasks below.

These results motivate the image experiments, where we evaluate the properties, interpolation, and semantic transfer that the connection structure is designed to improve.

Table 3: Connection error $\|T(b_i, b_j)f_i - f_j\|$ (lower = transport better aligns fibers). Near-zero for CvB and CvB (align. only)

Method	MNIST	FashionMNIST
CvB (ours)	$7.30 \times 10^{-6} \pm 2.30 \times 10^{-6}$	$8.71 \times 10^{-6} \pm 3.02 \times 10^{-6}$
CvB (no transport)	3.29 ± 0.19	1.35 ± 0.10
CvB (orth. only)	3.78 ± 0.28	2.76 ± 0.13
CvB (align. only)	$8.88 \times 10^{-6} \pm 2.15 \times 10^{-6}$	$6.44 \times 10^{-6} \pm 7.86 \times 10^{-7}$
CvB (cocyc. only)	2.82 ± 0.12	2.87 ± 0.15

C Results: Interpolation and Semantic Transfer

We evaluate the central claim: *the connection/transport structure of CvB lets decoded interpolation and semantic transfer stay competitive with flat autoencoders while materially improving latent topology and clustering*. We run experiments on MNIST and FashionMNIST with 10 seeds each; we report the mean$\pm$ std. Table 2 summarizes the key metrics. Full per-metric breakdowns are below (Tables 6–7 (MNIST) and Tables 8–9 (FashionMNIST)).

The detailed tables refine the summary of Table 2. The alignment term is again the mechanism: CvB (align. only) matches full CvB on clustering, while every variant without the alignment loss falls to ARI ≈ 0.41–0.43; on FashionMNIST the silhouette coefficient shows the same split (variants with alignment reach 0.26, those without ≈ 0.19). The β-VAE remains far behind on clustering (ARI 0.160 on MNIST, 0.113 on FashionMNIST).

Connectivity between same-class pairs is evaluated by (1) smoothness of the decoded path (second-order finite differences) and (2) semantic accuracy of a linear probe on decoded interpolants. On Smoothness, all methods produce reasonably smooth decoded paths. CvB and its variants are generally as smooth as or smoother than the plain AE on MNIST (0.0116 vs. 0.0128, -9%). VAE and β-VAE achieve the lowest smoothness scores on MNIST, consistent with their prior-regularized latents producing smooth paths. On FashionMNIST the differences shrink (all methods within a few percent of the VAE), indicating the interpolation task is easier there.

Regarding semantic accuracy, linear probe accuracy on decoded interpolants is high across all methods (> 0.85). The plain AE reaches the highest accuracy on MNIST (0.954), while CvB is slightly lower (0.929). We interpret this as a geometry–semantics tradeoff: CvB favors geometrically consistent bundle structure at a small cost in linear separability. On FashionMNIST the gap closes (CvB 0.857 vs. AE 0.862). Removing the connection (CvB (no transport)) and the orthogonality-only variant both exceed full CvB on this metric, which we discuss in Section H.

C.1 The connection is active: transport error

The connection error measures whether the learned transport map $T(b_i, b_j)$ correctly transports the fiber f_i to f_j when moving along the base. Table 3 shows that: (1) Full CvB and CvB (align. only) achieve connection error $\approx 10^{-6}$ on both datasets, the transport map is sharply learned and non-trivial (matches the expectation that fibers can be consistently transported along the base). (2) CvB (no transport) has no alignment/cocycle losses and error $\approx$ 1–3, showing that without the bundle constraints the fiber is not coherent across base points. (3) CvB (orth. only) behaves like CvB (no transport) ($\approx$ 3–4), orthogonality alone cannot align fibers; the alignment objective is the mechanism that produces a consistent transport. (4) CvB (cocyc. only) likewise fails to align (≈ 2.8), cocycle consistency without alignment does not anchor the transport to the geometry.

This is the strongest direct evidence that the connection in CvB is *active* and not decorative: identical architectures differ by orders of magnitude in transport error depending only on the presence of the alignment loss.

C.2 Detailed Metric Results

Table 4 aggregates the geometric metrics over all five datasets and ten seeds; the discussion below reads it together with the per-mesh tables.

Table 4: Geometric datasets: Mean±std over all 50 runs (5 datasets × 10 seeds). Lower MGDE and RTD are better; higher Trust is better. Bold = best for that metric. UMAP/PaCMAP have no decoder, hence no chart; they are shown for trustworthiness context. Note that metric scales differ strongly across datasets; read these aggregates alongside the per-dataset tables in this section.

Method	$\text{MGDE}_{\text{norm}} \downarrow$	RTD $\downarrow$	$\text{Trust}_5 \uparrow$
PCA	0.237 ± 0.113	7.29 ± 14.73	0.968 ± 0.018
UMAP	19.393 ± 13.56	69.40 ± 36.05	$\mathbf{0.998} \pm 0.001$
PaCMAP	28.998 ± 19.88	134.98 ± 95.75	0.996 ± 0.002
FibeRed	$\mathbf{0.208} \pm 0.183$	$\mathbf{3.66} \pm 7.38$	0.981 ± 0.015
Chart-AE	1.755 ± 1.082	12.63 ± 25.25	0.978 ± 0.017
CvB (ours)	0.376 ± 0.206	9.13 ± 18.66	0.958 ± 0.048
CvB (no transport)	0.364 ± 0.207	9.16 ± 18.62	0.958 ± 0.048
CvB (orth. only)	0.365 ± 0.207	8.95 ± 18.15	0.958 ± 0.048
CvB (align. only)	0.367 ± 0.208	9.49 ± 19.40	0.958 ± 0.048
CvB (cocyc. only)	0.361 ± 0.208	9.09 ± 18.38	0.958 ± 0.047

FibeRed dominates geometry metrics. Its bundle-geometry design is explicitly optimized for MGDE and RTD, and it achieves the best scores on 4/5 datasets, as anticipated above. Chart-AE has severe MGDE inflation on meshes: its MGDE is 5–10× worse than CvB on the Horse and Lion meshes (Table 5), indicating that chart-atlas reconstruction does not guarantee geometric faithfulness. The CvB variants again differ negligibly from one another, consistent with the brief above. UMAP/PaCMAP have very high MGDE on meshes: their non-parametric nature yields high trustworthiness but poor geodesic preservation, especially PaCMAP on meshes (MGDE > 50).

Table 5: Comparison of methods across different meshes.

	Horse mesh			Lion mesh	
Method	$\text{MGDE}_{\text{norm}}$	RTD	Method	$\text{MGDE}_{\text{norm}}$	RTD
PCA	0.183 ± 0.007	0.01 ± 0.00	PCA	0.182 ± 0.008	0.00 ± 0.00
FibeRed	$\mathbf{0.120} \pm 0.007$	$\mathbf{0.01} \pm 0.00$	FibeRed	$\mathbf{0.134} \pm 0.008$	$\mathbf{0.00} \pm 0.00$
Chart-AE	1.600 ± 0.085	0.43 ± 0.23	Chart-AE	2.520 ± 0.260	0.54 ± 0.49
CvB	0.223 ± 0.059	0.03 ± 0.02	CvB	0.277 ± 0.079	0.02 ± 0.03
CvB (no transport)	0.206 ± 0.024	0.02 ± 0.01	CvB (no transport)	0.240 ± 0.083	0.01 ± 0.01
CvB (orth. only)	0.206 ± 0.026	0.02 ± 0.01	CvB (orth. only)	0.240 ± 0.084	0.01 ± 0.01
CvB (align. only)	0.217 ± 0.038	0.02 ± 0.02	CvB (align. only)	0.256 ± 0.075	0.02 ± 0.02
CvB (cocyc. only)	0.206 ± 0.019	0.02 ± 0.01	CvB (cocyc. only)	0.236 ± 0.082	0.01 ± 0.01
UMAP	18.751 ± 0.396	41.23 ± 7.93	UMAP	33.880 ± 1.489	74.14 ± 15.30
PaCMAP	27.865 ± 0.656	98.56 ± 52.38	PaCMAP	50.363 ± 1.087	117.95 ± 16.42

Table 5 presents the results for the two non-convex meshes (Horse, Lion) (mean±std over 10 seeds). Tables 6 and 8 present ARI, NMI, and silhouette coefficient together for transparency.

C.3 Qualitative interpolation

Figure 3 shows decoded interpolation strips for MNIST (same-class pairs). CvB produces visually coherent transitions with plausible intermediate samples, consistent with its smoother decoded paths.

D Additional Experiments

D.1 Crossover: Transferring the Connection across Task Families

The connection's usefulness is claimed mainly on the image experiments (interpolation, semantic clustering). Two crossover checks test whether the bundle structure and its parametric transport generalize beyond the setting they were tuned for: (A) latent interpolation on a geometric dataset (SwissRoll), and (B) geometric and topological metrics on image latents. Together they probe whether the connection is a genuinely geometric mechanism rather than a dataset-specific regularizer.

Table 6: MNIST – ARI (higher = better).

Method	ARI	NMI	Silhouette
CvB (ours)	**0.511** $\pm$ 0.023	0.605 $\pm$ 0.017	0.194 $\pm$ 0.003
CvB (no transport)	0.434 $\pm$ 0.021	0.568 $\pm$ 0.015	0.119 $\pm$ 0.004
CvB (orth. only)	0.429 $\pm$ 0.037	0.555 $\pm$ 0.022	0.119 $\pm$ 0.003
CvB (align. only)	0.511 $\pm$ 0.029	**0.607** $\pm$ 0.018	0.196 $\pm$ 0.006
CvB (cocyc. only)	0.408 $\pm$ 0.024	0.547 $\pm$ 0.018	0.125 $\pm$ 0.003
AE	0.422 $\pm$ 0.024	0.540 $\pm$ 0.019	0.128 $\pm$ 0.005
VAE	0.425 $\pm$ 0.041	0.531 $\pm$ 0.032	0.056 $\pm$ 0.002
β-VAE	0.160 $\pm$ 0.036	0.239 $\pm$ 0.037	0.036 $\pm$ 0.001
Chart-AE	0.348 $\pm$ 0.039	0.488 $\pm$ 0.027	0.145 $\pm$ 0.006
PCA	0.376 $\pm$ 0.018	0.508 $\pm$ 0.012	0.116 $\pm$ 0.004
FibeRed	0.389 $\pm$ 0.038	0.552 $\pm$ 0.024	**0.245** $\pm$ 0.018

Table 7: MNIST – Interpolation smoothness, semantic accuracy, and connection error. Lower smoothness and connection error are better; higher semantic accuracy is better. "–" = not applicable or not computed (see Table 2).

Method	Interp. Smoothness $\downarrow$	Sem. Accuracy $\uparrow$	Connection Err. $\downarrow$
CvB (ours)	0.01160 $\pm$ 0.0009	0.929 $\pm$ 0.023	$7.30 \times 10^{-6} \pm 2.30 \times 10^{-6}$
CvB (no transport)	0.01216 $\pm$ 0.0007	0.944 $\pm$ 0.026	3.29 $\pm$ 0.19
CvB (orth. only)	0.01196 $\pm$ 0.0006	0.942 $\pm$ 0.023	3.78 $\pm$ 0.28
CvB (align. only)	0.01154 $\pm$ 0.0008	0.934 $\pm$ 0.022	$8.88 \times 10^{-6} \pm 2.15 \times 10^{-6}$
CvB (cocyc. only)	0.01233 $\pm$ 0.0008	0.936 $\pm$ 0.030	2.82 $\pm$ 0.12
AE	0.01280 $\pm$ 0.0008	**0.954** $\pm$ 0.019	–
VAE	0.01105 $\pm$ 0.0007	0.936 $\pm$ 0.019	–
β-VAE	0.01071 $\pm$ 0.0008	0.855 $\pm$ 0.021	–
Chart-AE	–	–	–
PCA	–	–	–
FibeRed	–	–	–
UMAP / PaCMAP	–	–	–

D.1.1 Latent interpolation on SwissRoll

We train CvB (and its connection-absent ablation, plus AE and VAE with matched latent dimension) on the 3-D SwissRoll, interpolate in the latent space between random point pairs, decode, and measure two qualities of the resulting path: (i) interpolant *smoothness*, the mean squared displacement between consecutive decoded interpolants (lower is smoother), and (ii) *manifold distance*, the mean distance of the decoded interpolants to the training point cloud (lower means the path stays on the learned manifold). Table 10 reports a single run (seed 0).

The two flat autoencoders produce the *smooth* interpolants here (smoothness ≈ 0.02–0.05) but their decoded paths leave the manifold: manifold distance ≈ 5, comparable to the global ambient spread of the SwissRoll. CvB reverses the trade: its decoded interpolants stay on the surface (distance ≈ 1.2) at the cost of coordinate smoothness, i.e. it follows the geometry instead of cutting a straight (short-cut) line through ambient space. CvB (no transport) shows the same geometric behavior (distance 1.09), so this particular effect is carried by the conditioned base–fiber architecture rather than by the transport. To be explicit about the limitation: on the raw smoothness metric CvB is $\sim 100\times$ worse than the VAE, so this experiment does *not* claim smooth walks; it documents a genuine trade-off between ambient smoothness and manifold fidelity. When visually smooth decoded paths are the priority, the flat baselines are preferable; the bundle is preferred when the interpolants must stay on the data manifold. Figure 5 shows the decoded paths superimposed on a 2-D projection of the SwissRoll.

D.1.2 MGDE and RTD on image latents

We train CvB (full and no-transport), AE, VAE, Chart-AE, PCA, the two differentiable-topology flat baselines (TopoAE-v0, RTD-AE-style; Section D.3), and the total-correlation disentanglement

Table 8: FashionMNIST – ARI, NMI, Silhouette (higher = better).

Method	ARI	NMI	Silhouette
CvB (ours)	0.405 ± 0.015	0.577 ± 0.015	0.257 ± 0.009
CvB (no transport)	0.398 ± 0.021	0.579 ± 0.018	0.191 ± 0.013
CvB (orth. only)	0.399 ± 0.020	0.580 ± 0.014	0.189 ± 0.007
CvB (align. only)	0.405 ± 0.007	0.580 ± 0.008	0.261 ± 0.006
CvB (cocyc. only)	0.401 ± 0.011	0.573 ± 0.009	0.210 ± 0.008
AE	0.352 ± 0.010	0.545 ± 0.005	0.214 ± 0.007
VAE	0.251 ± 0.023	0.379 ± 0.023	0.042 ± 0.003
β-VAE	0.113 ± 0.017	0.190 ± 0.024	0.035 ± 0.001
Chart-AE	0.324 ± 0.019	0.516 ± 0.017	0.195 ± 0.009
PCA	0.347 ± 0.020	0.511 ± 0.013	0.226 ± 0.009
FibeRed	0.365 ± 0.033	0.553 ± 0.026	0.345 ± 0.024
UMAP	**0.437** ± 0.011	0.610 ± 0.007	**0.519** ± 0.012
PaCMAP	0.434 ± 0.012	**0.611** ± 0.006	0.515 ± 0.012

Table 9: FashionMNIST – Interpolation smoothness, semantic accuracy, and connection error. "–" = not applicable or not computed (see Table 2).

Method	Interp. Smoothness ↓	Sem. Accuracy ↑	Connection Err. ↓
CvB (ours)	0.00823 ± 0.0012	0.857 ± 0.036	$8.71 \times 10^{-6} \pm 3.02 \times 10^{-6}$
CvB (no transport)	0.00811 ± 0.0010	0.862 ± 0.025	1.35 ± 0.10
CvB (orth. only)	**0.00801** ± 0.0010	0.864 ± 0.029	2.76 ± 0.13
CvB (align. only)	0.00840 ± 0.0012	0.850 ± 0.032	$6.44 \times 10^{-6} \pm 7.86 \times 10^{-7}$
CvB (cocyc. only)	0.00820 ± 0.0011	0.852 ± 0.032	2.87 ± 0.15
AE	0.00814 ± 0.0011	0.862 ± 0.026	–
VAE	0.00840 ± 0.0010	0.829 ± 0.030	–
β-VAE	0.00889 ± 0.0011	0.754 ± 0.032	–
Chart-AE	–	–	–
PCA	–	–	–
FibeRed	–	–	–
UMAP / PaCMAP	–	–	–

baselines (β-TCVAE, FactorVAE) on MNIST (4,000 training points), and evaluate the *latent* representation, not the decoded images, with the geometric metrics from Section B: MGDE (geodesic preservation) and RTD (topological fidelity). Ten seeds; mean±std in Table 11.

Three observations. First, the total-space latent of CvB is dramatically more topologically faithful than every autoencoder baseline: RTD ≈ 27 vs. 146 for AE, 104 for Chart-AE, and 66 for VAE, and the total-correlation objectives do not change this picture, β-TCVAE (63.7) and FactorVAE (66.6) stay on par with the plain VAE, showing that a factorized prior is orthogonal to latent geometry. Second, on MGDE, CvB (0.378) is comparable to the linear PCA reference (0.343) and far lower than AE (1.827) or Chart-AE (0.806); the prior-regularized VAE (0.470) is worse, and the TC baselines are worse still (0.662/0.672). Third, a flat-code topology penalty does not help either: RTD-AE-style at its shared default $k_{\mathrm{rtd}} = 1$ over-regularizes the code and collapses both metrics (MGDE ≈ 3850, RTD $\approx 5.95 \times 10^5$, Section D.3). Notably, this geometry is not the optimization target of the image experiments (which focus on interpolation and semantic clustering), yet it is conferred by the bundle structure: the two CvB variants are statistically indistinguishable from each other here, echoing Findings 1 and 4 of Section H that the transport's value is most visible on the interpolation and clustering tasks. Figure 6 summarizes the two metrics.

D.2 Generative Sampling from the Total Space

We fit an isotropic Gaussian to the training codes $\{r(x) = (b, f)\}$, draw samples (b^*, f^*), and decode $D(b^*, f^*)$, following the recipe of Section 4.4. Against a VAE baseline (matched latent dimension, fit and sampled the same way), the generative-sampling study uses the full ten seeds. We report the negative log-likelihood (NLL) of the fitted density over the codes (lower is better; a

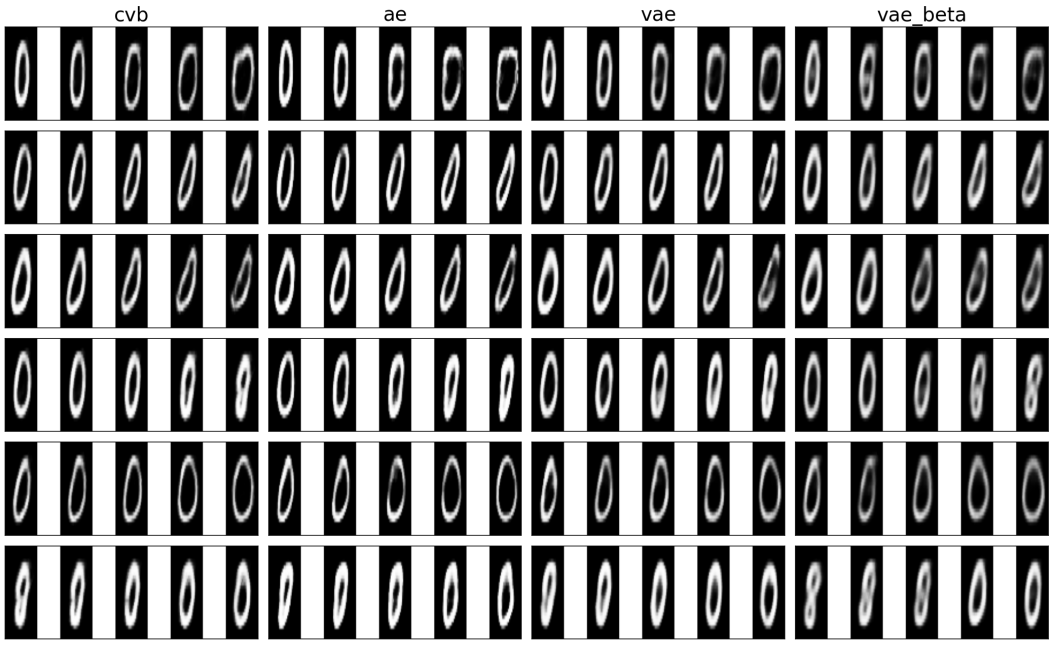
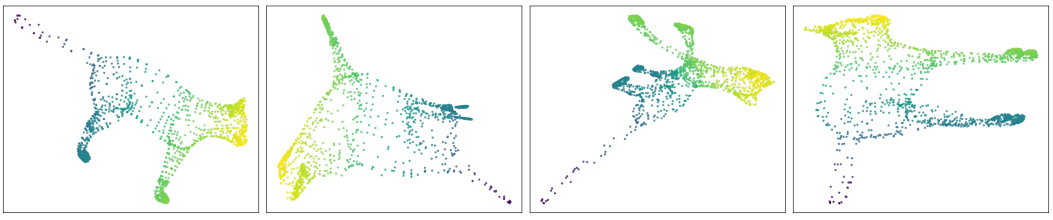


Figure 3: Decoded interpolation strips on MNIST. Each row shows a latent-space interpolation between two same-class samples.



(a) Cat data: PCA, FibeRed, Chart-AE and CvB.



(b) Lion data: PCA, FibeRed, Chart-AE and CvB.

Figure 4: Embeddings visualization by method from left to right: PCA, FibeRed, Chart-AE and CvB.

negative value means the mean log-likelihood is positive), and the norm and variance of the decoded samples as a collapse check.

The fitted density reaches a far higher likelihood on the CvB codes than on the VAE codes (NLL -66.5 vs. $+27.8$): under the (b, f) coordinate pair the latent mass concentrates on a configuration the Gaussian can model, so sampling-then-decoding is well calibrated rather than drifting into unmodeled regions. The decoded CvB samples are not collapsed: their norm (7.23) and per-pixel variance (0.047) match the VAE's and are consistent with the training data scale.

Table 10: SwissRoll: decoded latent interpolations. Lower smoothness and lower manifold distance are better. AE/VAE interpolants are coordinate-smooth but fly off the manifold; CvB follows the rolled surface. (Single run, seed 0.)

Method	Interp. Smoothness ↓	Manifold Dist. ↓
CvB (ours)	5.15	1.23
CvB (no transport)	5.12	1.09
AE	0.049	5.16
VAE	0.018	5.04

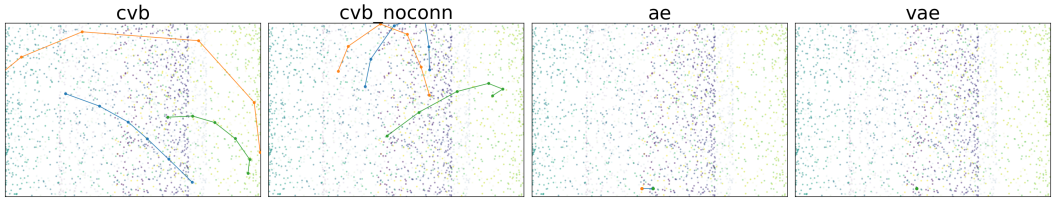


Figure 5: Decoded latent interpolation paths on SwissRoll, one panel per method, superimposed on a 2-D projection of the point cloud. Bundle-based methods stay near the manifold surface; AE/VAE paths cut through ambient space.

Gaussian NLL is scale sensitive, so this ranking is meaningful only under the identical input normalization shared by all methods in this benchmark; absolute values should not be compared across datasets trained at different scales.

D.3 Topological autoencoder baseline

To check that CvB's advantage is not merely the benefit of *any* topological regularizer, we train capacity-matched flat autoencoders with two differentiable topological penalties from the literature, in the image benchmark and the crossover latent-geometry experiment (Section D.1.2). **TopoAE-v0** Moor et al. (2020) differentiates $0.5 \sum_i (\mathrm{death}_i)^2$ over the single-linkage merge events of the batch; the merge *pairs* are read off a non-differentiable single-linkage pass, exactly as in the released TopoAE code. **RTD-AE-style** Trofimov et al. (2023) minimizes a truncated-Betti cross-barcode divergence between the normalized input and latent barcodes of a mini-batch sub-sample, computed over the Rips filtration: H_0 through the same differentiable single-linkage deaths, and H_1 through Rips persistence pairs (read off a detached gudhi pass and re-expressed differentiably as the birth-edge weight and the maximum weight of the death triangle). The one-to-one matching between the two barcodes is solved on a detached cost matrix and re-evaluated with gradients enabled (including diagonal matching). Following the original schedule, the RTD term is inactive during a reconstruction warm-up and active afterwards. Both baselines use a flat AE of identical capacity and latent dimension, with no base/fiber split or transport. Table 13 compares them with the matched flat AE and CvB under the same clustering/interpolation protocol. The remaining topology/geometry-aware pipelines cited in Section 2 (GeomAE Nazari et al. (2023), Riemannian/Hamiltonian VAEs Chadebec et al. (2020)) are not rerun: beyond the single-flat-block argument in Section 2, their official multi-stage geodesic or Hamiltonian-Monte-Carlo solvers are orders of magnitude heavier than a matched decoder, and neither exposes a base/fiber factorization or a transport map, so they cannot take part in the base–fiber or fiber-transfer evaluation. Disentanglement baselines, FactorVAE Kim & Mnih (2018) and β-TCVAE Chen et al. (2018), are likewise included as flat VAE variants on the same image protocol; they need no extra data and differ from the matched VAE only through their total-correlation objective. Standardized factorization (dSprites; Section D.4) is reported below with MIG tables; 3D Shapes and DCI tables remain future work (see Section 6).

For fairness, the new baselines share one protocol with the existing image benchmark: 10 random seeds, 100 epochs, batch size 256, Adam ($\mathrm{lr} = 10^{-3}$), the same capacity-matched flat backbone (400 hidden units and a 20-dimensional internal latent, projected to the 24-dimensional code scored downstream, matching CvB's 8-dim base $\oplus$ 16-dim fiber), and the identical clustering, interpolation, and geometry-scoring pipeline. Only the method-specific penalty weights differ, and we state them

Table 11: MNIST: MGDE and RTD computed on the *latent* representation, not the reconstruction. Lower is better for both. The bundle-based methods preserve the intrinsic geometry of the digit manifold far better than the flat autoencoders. The RTD-AE-style row uses the shared default $k_{\mathrm{rtd}} = 1$ (Section D.3); its flat-code topology penalty does not generalize beyond the training mini-batches and collapses both metrics (semantic accuracy drops to 0.36 on the trackB protocol). (10 seeds.)

Method	$\mathrm{MGDE}_{\mathrm{norm}} \downarrow$	RTD $\downarrow$
CvB (ours)	0.378 ± 0.089	26.81 ± 9.83
CvB (no transport)	0.345 ± 0.064	22.32 ± 5.66
TopoAE-v0	0.941 ± 0.001	129.20 ± 8.02
RTD-AE-style	3850.59 ± 345.90	595116 ± 39811
AE	1.827 ± 0.570	145.87 ± 15.23
VAE	0.470 ± 0.071	66.18 ± 4.93
β-TCVAE	0.662 ± 0.005	63.74 ± 2.72
FactorVAE	0.672 ± 0.005	66.58 ± 3.08
Chart-AE	0.806 ± 0.031	103.96 ± 4.68
PCA	0.343 ± 0.009	42.97 ± 3.97

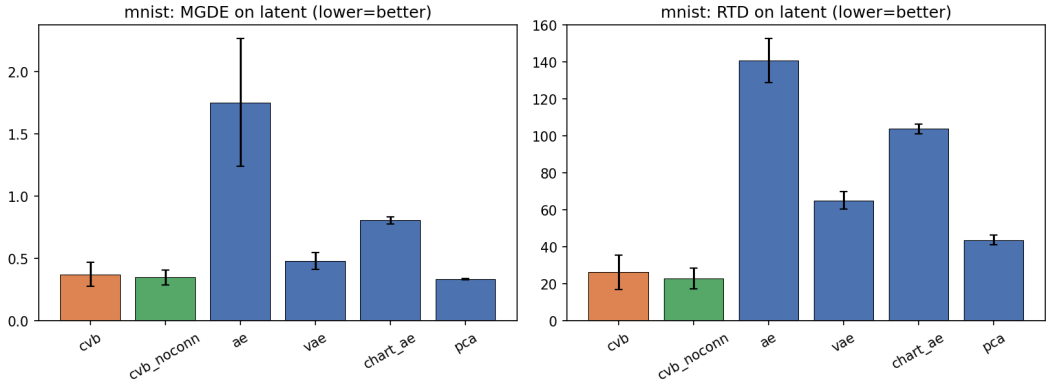


Figure 6: MGDE (left) and RTD (right) of MNIST latent representations (mean±std over 10 seeds; lower is better). CvB is the best parametric method on RTD and comparable to PCA on MGDE.

explicitly instead of tuning the baselines. TopoAE-v0 uses its released single-linkage death^2 objective; RTD-AE-style carries weight $k_{\mathrm{rtd}} = 1$ and is inactive during the reconstruction warm-up, per the original schedule. Within the total-correlation family, we keep a single KL weight $\beta = 4$ across β-VAE, β-TCVAE and FactorVAE (plain VAE uses $\beta = 1$). This one-weight-for-all choice removes per-method tuning bias and lies within the range used by both the β-TCVAE and FactorVAE papers; FactorVAE additionally pays the adversarial weight $\gamma = 6$ as in Kim & Mnih Kim & Mnih (2018). No further hyperparameter search was performed.

D.4 Disentanglement: Base–Fiber Separation

We study whether the base/fiber split is *usable*: does the base carry the global identity while the fiber carries within-class style? Using four factors extracted from MNIST pixels (digit, intensity, width, vertical position), we evaluate, over ten seeds, the standard axis-alignment scores SAP and DCI, a Kim–Mnih style β-consistency, and the CvB-specific diagnostic: balanced accuracy of a logistic regressor predicting the digit from the base block alone, the fiber block alone, and from the full code. Table 14 reports the results.

The standard scores are essentially flat: SAP < 0.01 and DCI total ≈ 0.085 for every method, β-VAE included, which simply confirms that factors inferred from pixel statistics (brightness, stroke width, vertical position) are not axis-aligned in any trained latent space. They are the wrong instrument for evaluating a *block* structure. The block-level diagnostic tells the relevant story:

Table 12: MNIST: Gaussian fit to the latent codes and subsequent sampling. Lower NLL is better (negative NLL = positive mean log-likelihood). The CvB total-space codes form a much denser, more Gaussian density than the VAE's, and decoded samples retain a data-like norm and variance. (10 seeds.)

Method	Gaussian NLL ↓	Sample Norm	Sample Variance
CvB (ours)	-66.51 ± 2.45	7.23 ± 0.38	0.0474 ± 0.0045
VAE	$+27.82 \pm 0.07$	6.70 ± 0.20	0.0396 ± 0.0025

Table 13: MNIST image benchmarks: topology-aware flat baselines (TopoAE-v0, Moor et al. (2020); RTD-AE-style, Trofimov et al. (2023)), the total-correlation VAE baselines (β-TCVAE, FactorVAE; Section D.3), versus CvB and the matched AE. The topology baselines are flat AEs of matched capacity/latent with a differentiable topological penalty and no base/fiber split or transport. The RTD-AE-style row uses the shared default $k_{\mathrm{rtd}} = 1$ (Section D.3)

Method	ARI ↑	NMI ↑	Interp. smooth ↓	Sem. acc ↑	RTD ↓
CvB (ours)	0.511 ± 0.023	0.605 ± 0.017	0.01160 ± 0.00086	0.929 ± 0.023	26.81 ± 9.83
TopoAE-v0	0.436 ± 0.015	0.583 ± 0.012	0.00975 ± 0.00057	0.930 ± 0.020	129.20 ± 8.02
RTD-AE-style	0.220 ± 0.020	0.317 ± 0.021	0.01376 ± 0.00159	0.356 ± 0.065	595116 ± 39811
AE	0.422 ± 0.024	0.540 ± 0.019	0.01280 ± 0.00077	0.954 ± 0.019	145.87 ± 15.23
β-TCVAE	0.220 ± 0.028	0.310 ± 0.021	0.01081 ± 0.00059	0.851 ± 0.027	63.74 ± 2.87
FactorVAE	0.182 ± 0.028	0.260 ± 0.030	0.01059 ± 0.00050	0.856 ± 0.023	66.58 ± 3.24

- **The connection concentrates identity in the base.** In the full model, digit identity is read off the base block alone at 0.743, indistinguishable from the full code (0.744), while the fiber alone carries only 0.188, the base is the identity axis and the fiber is residual within-style content.
- **Removing the transport destroys the split.** In the connection-absent ablation, the fiber alone reads identity at 0.808, the base at 0.523, and the full code is substantially better than either (0.853): identity is smeared over both blocks and the base/fiber separation is lost. This is exactly the structural consequence of Theorem 3.2 predicted in Section 3, the flat, identity-transport model is incapable of a base-local fiber.
- **AE gives no structure to leverage.** The plain AE achieves the highest digit accuracy of all (0.883) but its latent block partition is meaningless, so there is no axis along which style can be transferred independently of identity.
- **Base-conditioning is not what enforces the split.** The *unconditioned-but-transported* variant $f = \psi(x)$ retains the full active transport and loss, isolating what base-conditioning adds beyond a free T, the case that Theorem 3.2 does *not* cover. It matches full CvB: identity stays base-concentrated (base $0.743 \approx$ full 0.744 vs. fiber 0.187), far from the no-transport leak (fiber 0.808), and continuous-style factors are retained in its fiber at least as well as under full CvB (Table 15). Base-conditioning on the fiber entry is therefore not the mechanism behind the split; the active transport and the alignment objective are.

Continuous style factors. The block accuracies above probe the *discrete* digit identity. To check what the fiber retains at a finer level, we add three pixel-derived scalar factors, mean stroke intensity, stroke width, and vertical position, and measure how well each is linearly recoverable (R^2 of ridge regression with 5-fold cross-validation) from the base block and the fiber block. Intensity and width are largely collinear with digit identity, so CvB recovers them mostly from the base (0.72/0.71) with the fiber keeping a secondary residual (0.38/0.37); the no-transport ablation instead recovers them better from the fiber (0.86/0.85) than from the base (0.47/0.47), mirroring the identity leakage measured by Fiber-Acc. Table 15 reports R^2 for the fiber block.

Ground-truth factors (dSprites). Pixel-derived factors confound identity and style, and no method attains axis-aligned scores on MNIST. To supplement them with genuinely ground-truth generative factors, we run the same capacity-matched baselines on the standardized dSprites dataset

Table 14: MNIST: disentanglement metrics (10 seeds). Standard axis-alignment scores are uniformly low because pixel-derived factors are not axis-aligned in any latent. The informative signal is the block-level digit accuracy: CvB concentrates identity in the base (base $0.743 \approx$ full 0.744, fiber 0.188), while the connection-absent ablation leaks identity into the fiber (fiber 0.808).

Method	Base-Acc ↑	Fiber-Acc	Full-Acc ↑	β-cons. ↑	DCI total ↑	SAP ↑
CvB (ours)	0.743 ± 0.021	0.188 ± 0.053	0.744 ± 0.020	0.422 ± 0.071	0.0858 ± 0.0015	0.006
CvB (uncond. fiber)	0.743 ± 0.023	0.187 ± 0.060	0.744 ± 0.023	0.461 ± 0.062	0.0856 ± 0.0022	0.006
CvB (no transport)	0.523 ± 0.045	0.808 ± 0.022	0.853 ± 0.006	0.633 ± 0.084	0.0868 ± 0.0031	0.005
AE	–	–	0.883 ± 0.003	0.636 ± 0.048	0.0876 ± 0.0015	0.005
VAE	–	–	0.732 ± 0.011	0.428 ± 0.051	0.0857 ± 0.0016	0.004
β-VAE	–	–	0.451 ± 0.034	0.269 ± 0.024	0.0840 ± 0.0011	0.004

Table 15: MNIST: linear recoverability (R^2, ridge, 5-fold CV) of continuous pixel factors from the fiber block. Higher means the factor is better retained in the fiber. Columns: stroke intensity, stroke width, vertical position. Total-correlation baselines (FactorVAE Kim & Mnih (2018), β-TCVAE Chen et al. (2018)) are evaluated on the same protocol in Table 11; a standardized factorized dSprites benchmark is reported in Table 16; 3D Shapes and DCI tables remain future work.

Method	R^2 intensity ↑	R^2 width ↑	R^2 vpos ↑
CvB (ours)	0.375 ± 0.109	0.371 ± 0.107	-0.001 ± 0.002
CvB (uncond. fiber)	0.552 ± 0.125	0.546 ± 0.124	0.003 ± 0.004
CvB (no transport)	0.863 ± 0.102	0.854 ± 0.102	0.013 ± 0.009

(64×64 binary sprites with five known factors: shape, scale, orientation, position x, position y; area-downsampled to 28×28 for the shared protocol and sub-sampled to $10{,}000$ train / $4{,}000$ test samples stratified by shape, with the same hyperparameters as above). We report MIG (Mutual Information Gap Chen et al. (2018)): the β-TCVAE headline metric, computed as quantile-binned mutual information from each latent dimension to each generator factor, and the linear recoverability R^2 (ridge, 5-fold CV) of each factor from the full code, and, for CvB, from the base and fiber blocks separately, i.e. whether the discrete identity factor (shape) and the continuous style factors (scale, orientation, position) split across base/fiber as claimed. The results are in Table 16. The total-correlation baselines attain the highest MIG (β-TCVAE 0.095, FactorVAE 0.049), expected on a dataset designed for those methods. CvB's MIG (0.028) is comparatively modest, because the five dSprites factors are mutually independent and none plays the global-identity role that the base/fiber split is designed to capture. The block-level R^2 nonetheless shows the split is usable: continuous style factors remain linearly recoverable from the fiber (scale R^2=0.37, position 0.81; without the connection, the fiber absorbs more of scale, 0.85), while the discrete shape factor is not concentrated in the base (0.014), dSprites provides no class/identity structure, unlike MNIST digits.

D.5 Extrinsic vs. intrinsic (transport-guided) interpolation

The decoded interpolation of Eq. 4 interpolates the fiber linearly and therefore never invokes the learned transport T. The intrinsic scheme of Eq. 5 instead parallel-transports the fiber along the base path using the learned connection. We compare the two decoded paths on MNIST with the same smoothness (mean squared second-order finite difference) and linear-probe semantic-accuracy criteria used in Table 2. Under a trivial connection ($T \equiv \mathrm{Id}$) the intrinsic update freezes the fiber, so the two paths need not coincide; CvB's learned transport instead keeps the translated fiber close to the ambient chord, and the measured paths of CvB do coincide. The gap between columns therefore isolates the connection's contribution to the decoded path rather than a difference in the base trajectory. Table 17 reports both paths for CvB, the unconditioned-fiber variant, and the no-transport ablation.

D.6 Connection-Weight Ablation

We hold the model fixed and toggle the two connection losses $(\lambda_{\mathrm{align}}, \lambda_{\mathrm{cocycle}})$ across their four corners on MNIST (10 seeds): full $= (1.0, 0.1)$; alignment only (cocycle dropped) $= (1.0, 0)$; co-

Table 16: dSprites: MIG (mean over the five ground-truth factors, higher is better) and linear R^2 (ridge, 5-fold CV). "Base"/"Fiber" split the code into CvB's base and fiber blocks (flat baselines have no such split; — marks not-applicable); "R^2 full" is from the whole code. "vpos" is the mean of the horizontal and vertical position factors. Ten seeds. CvB (no tran.): CvB (no transport)

Method	MIG ↑	R^2 full ↑	shape R^2 base ↑	shape R^2 fiber ↑	scale R^2 base ↑	scale R^2 fiber ↑	vpos R^2 base ↑	vpos R^2 fiber ↑
CvB (ours)	$0.028 \pm .01$	$0.485 \pm .01$	$0.014 \pm .01$	$0.027 \pm .01$	$0.213 \pm .09$	$0.366 \pm .08$	$0.912 \pm .03$	$0.806 \pm .07$
CvB (no tran.)	$0.031 \pm .02$	$0.559 \pm .01$	$0.004 \pm .01$	$0.065 \pm .01$	$0.130 \pm .17$	$0.853 \pm .03$	$0.675 \pm .32$	$0.335 \pm .27$
FactorVAE	$0.049 \pm .03$	$0.512 \pm .02$	—	—	—	—	—	—
β-TCVAE	$0.095 \pm .04$	$0.496 \pm .02$	—	—	—	—	—	—
β-VAE	$0.062 \pm .03$	$0.491 \pm .02$	—	—	—	—	—	—
VAE	$0.035 \pm .02$	$0.543 \pm .01$	—	—	—	—	—	—
AE	$0.026 \pm .01$	$0.546 \pm .01$	—	—	—	—	—	—

Table 17: Extrinsic (Eq. 4) vs. intrinsic transport-guided (Eq. 5) decoded interpolation, MNIST. Both smoothness and semantic accuracy use the protocol of Table 2; lower smoothness and higher accuracy are better. The gap between the two columns within a row is the connection's effect on the decoded path.

Method	extr. smooth ↓	intr. smooth ↓	extr. sem ↑	intr. sem ↑
CvB (ours)	0.01160 ± 0.00086	0.01160 ± 0.00086	0.929 ± 0.023	0.929 ± 0.023
CvB (uncond. fiber)	0.01118 ± 0.00078	0.01115 ± 0.00078	0.933 ± 0.022	0.935 ± 0.023
CvB (no transport)	0.01216 ± 0.00072	0.00331 ± 0.00055	0.944 ± 0.026	0.898 ± 0.034

cycle only (alignment dropped) $= (0, 0.1)$; and both disabled $= (0, 0)$. Table 18 reports clustering (ARI/NMI/silhouette), interpolation smoothness, semantic accuracy, and the connection-error diagnostic of Eq. 6.

The ablation isolates the alignment objective as the working term, in precise agreement with the variant ablations of Section C:

- **Alignment is sufficient.** Dropping the cocycle term changes nothing: full and no-cocycle agree on every metric (ARI 0.486 vs. 0.488, silhouette 0.193 vs. 0.191, connection error $\sim 10^{-6}$). The cocycle loss at its default weight 0.1 contributes little beyond what alignment already enforces.
- **Alignment is necessary.** Removing alignment while keeping the cocycle term (no-align) reproduces the no-loss configuration (ARI 0.421 vs. 0.420, silhouette 0.122 vs. 0.122) and raises the connection error to $\mathcal{O}(1)$: the cocycle constraint alone does not anchor the transport to the geometry. The transport is then mathematically active but geometrically meaningless.
- **The mechanism is structural, not tuning-specific.** The connection error separates the two groups by five orders of magnitude purely along the λ_{align} axis, matching the $\sim 10^{-6}$ vs. $\mathcal{O}(1)$ gap of the architectural ablation (Table 3).

D.7 Scalability

The transport machinery adds an extra forward pass and a pair-distance step per batch (Section 4.3); the question is whether this changes the *asymptotic* cost. We fix the model capacity, batch size, and number of epochs, and train CvB, AE, and VAE on MNIST at five dataset sizes ($n \in \{1000, 2000, 4000, 8000, 12000\}$). Table 19 reports total wall-clock time (8 epochs) and test reconstruction MSE.

All three methods scale linearly in n (total time roughly doubles with each doubled set). The CvB-to-AE ratio is stable at ≈ 2.8–3.0 across all five sizes, confirming that the extra cost of the transport is a constant per-epoch factor and no $\mathcal{O}(n^2)$ term appears from the batch-local neighbor step. The VAE and AE are essentially cost-identical, so the gap is specific to the connection machinery rather than to the encoder/decoder architecture. On reconstruction, CvB stays slightly above the VAE (0.041

Table 18: MNIST: connection-weight ablation (10 seeds, mean±std). These configurations were re-run under their own set of random draws, so absolute values differ slightly from Table 2 (the full config reaches ARI 0.486 here vs. 0.511 there); only within-table comparisons are meaningful. The alignment loss is both necessary and sufficient: configurations with $\lambda_{\text{align}} = 1$ (full and no-cocycle) reach ARI ≈ 0.49, silhouette ≈ 0.19, and connection error $\sim 10^{-6}$; configurations with $\lambda_{\text{align}} = 0$ (no-align and none) collapse to ARI ≈ 0.42, silhouette ≈ 0.12, and $\mathcal{O}(1)$ connection error, even when the cocycle term alone is active.

Config	λ_{align}	λ_{cocycle}	ARI ↑	Silhouette ↑	Interp. Smooth ↓	Sem. Acc ↑	Connection Err. ↓
CvB (full)	1.0	0.1	$0.49 \pm 3\text{e-}2$	$0.19 \pm 6\text{e-}3$	$0.01 \pm 6\text{e-}4$	$0.93 \pm 2\text{e-}2$	$8.1\text{e-}6 \pm 3.3\text{e-}6$
CvB (no cocycle)	1.0	0.0	$0.49 \pm 3\text{e-}2$	$0.19 \pm 6\text{e-}3$	$0.01 \pm 1\text{e-}3$	$0.92 \pm 2\text{e-}2$	$6.5\text{e-}6 \pm 1.2\text{e-}6$
CvB (no align)	0.0	0.1	$0.42 \pm 3\text{e-}2$	$0.12 \pm 9\text{e-}3$	$0.01 \pm 6\text{e-}4$	$0.94 \pm 2\text{e-}2$	$2.80 \pm 4.7\text{e-}1$
CvB (none)	0.0	0.0	$0.42 \pm 4\text{e-}2$	$0.12 \pm 6\text{e-}3$	$0.01 \pm 7\text{e-}4$	$0.94 \pm 3\text{e-}2$	4.42 ± 1.01

Table 19: MNIST: total training wall time (seconds, 8 epochs) vs. training-set size, with test reconstruction MSE at the largest size. The CvB-to-AE time ratio is $\approx$ 2.8–3.0 at every size: the connection adds a constant per-epoch factor but does not change the linear scaling. (Single run per size.)

	Total time (s) at train size n					
Method	1000	2000	4000	8000	12000	Recon MSE @12000
CvB (ours)	0.97	2.02	4.00	7.61	11.28	0.041
AE	0.35	0.71	1.32	2.66	3.81	0.024
VAE	0.33	0.70	1.30	2.62	3.73	0.039
CvB / AE ratio	2.77	2.85	3.03	2.86	2.96	

vs. 0.039 at $n = 12000$) and above the unconstrained AE (0.024): the bundle objective trades a little reconstruction fidelity for the structure used in the interpolation, clustering, and base/fiber experiments. These timing numbers are a single-run snapshot at 8 epochs (a scaling sanity check, not a throughput benchmark).

E Proof sketch for Theorem 3.1

Proof sketch of Theorem 3.1. (b)⇒(a): zero holonomy on all contractible loops is the definition of flatness, $R^{\nabla} = 0$. (a)⇒(b): on the connected, simply connected overlaps, the flat connection provides horizontal frames that relate by constants, and the cocycle condition Eq. 2 makes the associated transports agree on triple overlaps, so parallel transport is independent of the path. Over a simply connected M, parallel transport along arbitrary paths from a point p defines a global parallel frame, i.e. a global trivialization (c); a global frame is exactly a cocycle trivialized by constant gauge functions. □

Scope: the learned connection is flat by construction. The transport is a *single-valued* pairwise map constrained by the cocycle condition $T_{jk}T_{ij} \approx T_{ik}$, so by Theorem 3.1 it approximates a flat, path-independent connection. This is a deliberate and useful inductive bias, it is precisely what lets us amortize parallel transport into one feed-forward pass instead of solving an inference-time geodesic boundary-value problem, but it also bounds what the model can represent. We emphasize that a flat *connection* is not the same as a flat *base*: the base chart $b = \varphi(x)$ may still be curved, and CvB does not assume otherwise. What the composition consistency rules out is non-trivial *holonomy* (connection curvature): manifolds whose geometry demands path-dependent transport with non-trivial monodromy, such as a Möbius band (whose holonomy around the generator is -1, so transport is orientation-reversing and loop-dependent), lie outside the expressive range of a single-valued, cocycle-consistent T, as do bundles with non-zero characteristic classes. CvB should thus be read as a *flat-connection* bundle model. Curvature-aware or path-valued transports are a natural extension and are not claimed here; we return to this in Section 6.

F PROOF AND CONSEQUENCES OF THEOREM 3.2

Proof of Theorem 3.2. With $T \equiv \mathrm{Id}$, the parallel condition is $f(b_i) = f(b_j)$ for every edge of the transportable-pair graph; the relation extends transitively along paths, so f is constant on each connected component of the base graph. In base coordinates, a parallel section of the trivial bundle satisfies $\nabla^{\mathrm{hor}}_X f = \mathrm{d}f(X) = 0$ for all horizontal directions X, i.e., f is locally constant. □

Consequences.

- For an *unconditioned* encoder $\psi\colon x \mapsto f$, the fiber is a fixed vector space $\mathbb{R}^r$ independent of the base. Then, with the transport pinned to identity, no alignment objective (present or future) can make the fiber section base-local: a parallel section of the trivial bundle is constant, so an identity-transport model is *structurally incapable* of a genuinely base-varying fiber.
- A genuinely base-varying fiber therefore needs one of two things: either the fiber is conditioned on the base point, $f = \psi(x, b)$, or the transport is allowed to be non-trivial. The identity-transport product model has neither and degenerates to a constant section. This makes base-conditioning a natural prior *in the product limit*; it is not a logical necessity once the transport is free, as discussed in Section 3.3.

What Theorem 3.2 does and does not claim. The theorem establishes constant parallel sections only for the trivial connection $T \equiv \mathrm{Id}$. It does not assert that an unconditioned fiber $f = \psi(x)$ is *incapable* of carrying base-local content: with the transport free, the alignment term $T_{ij} f_i \approx f_j$ can always be satisfied by choosing T_{ij} to map f_i to f_j, whatever the encoder. A non-trivial connection can therefore in principle "repair" an unconditioned fiber, and the mutual necessity of conditioning and transport is *not* a theorem. What the theorem gives is a sharp baseline: *when* the transport is pinned to the identity, the only parallel sections are constants, so any base-local fiber content must come from an active connection. We accordingly treat conditioning as a modeling prior and test its value *under an active transport* directly, via the unconditioned-but-transported ablation in Section D.4 (Table 14).

Empirically, the identity-transport prediction is observed directly: in the disentanglement study, the full model separates base-controlled identity (0.74 balanced accuracy from the base block) from fiber-controlled style (0.20), whereas the connection-absent (product) variant leaks identity into the fiber (0.80) and has none of the structure Theorem 3.2 forbids it from having. The unconditioned-but-transported variant, by contrast, is *not* covered by the theorem and is compared empirically in the same table.

G REPRODUCIBILITY AND TRAINING DETAILS

Optimization and schedule. We train autoencoder-style methods for 50 epochs on the small-scale geometric datasets and for 100 epochs on the large-scale image datasets (convergence verified; see Appendix), all with 10 random seeds. Unless stated otherwise, we use Adam with learning rate 10^{-3} and no weight decay; the small disentanglement and generative experiments (Sections D.4–D.2) use 3×10^{-4} because their training sets are smaller. Gradients are estimated on mini-batches of size $B = 256$ drawn uniformly without replacement. The loss weights are fixed at $(\lambda_{\mathrm{recon}}, \lambda_{\mathrm{align}}, \lambda_{\mathrm{cocycle}}, \lambda_{\mathrm{orth}}, \lambda_{\mathrm{var}}) = (1.0, 1.0, 0.1, 10^{-3}, 10^{-2})$ in every experiment.

Neighbor pairs are sampled in input space. The distance matrix D used to select neighbor pairs and triples is the Euclidean distance between *raw inputs*, $\mathsf{D}_{ij} = \|x_i - x_j\|_2$ (pixel space for the image datasets), *not* distances between learned base coordinates b. Since b is itself learned, selecting neighbors in b would let the model choose its own neighborhoods and hide data geometry; sampling in input space keeps the neighbor graph a fixed property of the data. For each x_i, the second and third nearest neighbors in the batch ($j = n_1(i)$, $k = n_2(i)$, self excluded) form the training triples (Section 4.3).

Evaluation protocol. On the image datasets, we hold out $n_{\mathrm{eval}} = 3000$ test images disjoint from the $n_{\mathrm{train}} = 8000$ training images. Clustering metrics (ARI/NMI/silhouette) are computed on the

Table 20: Exact CvB architecture and optimization hyperparameters. All hidden layers use $\tanh$; the decoder output is σ for pixel data in $[0, 1]$ and linear otherwise. Image experiments use $d_b = 8$, $d_f = 16$.

Component	Layers (input $\to$ hidden $\to$ output)	Activation
Base φ	$d \to 512 \to 256 \to d_b$	$\tanh$
Fiber ψ	$(d + d_b) \to 256 \to 128 \to d_f$	$\tanh$
Decoder D	$(d_b + d_f) \to 256 \to 512 \to d$	$\tanh$, σ/linear
Transport $\mathcal{T}$	$2d_b \to 32 \to 64 \to d_f^2$	$\tanh$
Adam, lr 10^{-3} (3×10^{-4} for disentanglement/generative), weight decay 0		
Mini-batch $B = 256$; 50 (geometry) / 100 (image) epochs; 10 seeds		
Neighbor distance D on raw inputs x; $x_j = n_1(i)$, $x_k = n_2(i)$		

total-space codes $r(x) = (b, f)$ of the held-out test images using K-means ($k = 10$, `n_init`=5, `random_state`=0, otherwise scikit-learn defaults); interpolation metrics use same-class test pairs. The linear probe is a logistic-regression classifier ($C = 1$) trained on raw pixels and applied to decoded interpolants. UMAP and PaCMAP are non-parametric baselines without a decoder, so interpolation and semantic-accuracy metrics do not apply to them. Experiments were run on a Linux GPU server, except for the UMAP/PaCMAP visualizations, which were generated in Google Colab.

Architecture. All four sub-networks are plain MLPs with $\tanh$ hidden activations and a linear output layer (the decoder output uses σ for pixel data in $[0, 1]$). Table 20 gives the exact widths. In the image experiments, the base and fiber dimensions are $d_b = 8$ and $d_f = 16$ (total latent 24), matched by the flat AE/VAE baselines. The transport network outputs an unconstrained $d_f \times d_f$ matrix (Section 4.1); its conditioning is measured and discussed in Table 1.

H Discussion and Findings

Our experiments ask one question: does the connection/transport structure of the Conditioned Vector Bundle (CvB) network help, or is it decorative? The answer, supported by the results above, is nuanced but affirmative: *the connection helps precisely where its design warrants*, namely clustering on the total-space representation, topological organization, and coherent fiber geometry; it does not help (and is not designed to help) raw geodesic distortion or dominant ambient-interpolation smoothness.

H.1 Finding 1: The connection is active and not trivial

The most direct evidence is the connection error, defined as $\|T(b_i, b_j)f_i - f_j\|$ (transport of fiber f_i by the map between base points b_i, b_j versus the observed fiber f_j), shown in Table 3. With identical encoder/decoder architectures, the difference between CvB and every connection-absent ablation is several orders of magnitude (10^{-6} vs. $\mathcal{O}(1)$). Removing the alignment loss, as in the orthogonality-only and cocycle-only variants, collapses transport consistency to the no-connection level. This isolates a single mechanism: *the alignment objective is what anchors the transport map to the geometry*. Orthogonality and cocycle consistency are necessary to make the transport well-defined (uniqueness of factorization), but alone they cannot align fibers across base points.

H.2 Finding 2: The alignment term drives the clustering improvement

Clustering quality on the total-space representation (ARI/NMI/Silhouette; Tables 2, 6, 8) is consistently best for CvB and its alignment-only variant on both datasets. The contribution of the transport is dataset-dependent, however: on FashionMNIST, CvB (no transport) is statistically indistinguishable from full CvB (ARI 0.398 vs. 0.405; paired t-test $p = 0.48$), whereas on MNIST removing alignment costs roughly 0.08 ARI ($p = 7 \times 10^{-5}$) and full CvB (0.511) pulls clearly ahead of both its no-transport brother (0.434) and the flat AE (0.422). This indicates that:

- The *total-space* representation (b, f) with a trained connection produces better class separation than the flat-latent AE representation of matched dimension, particularly on MNIST ($+0.09$ ARI) and FashionMNIST ($+0.05$ ARI).
- The base-conditioned architecture alone already lifts clustering above the flat AE on both datasets (MNIST ARI 0.434 vs. 0.422; FashionMNIST ARI 0.398 vs. 0.352), so transport is not required for gains over a flat latent. The *active* connection adds separation where the clustering task is hardest, most visibly on MNIST (full CvB 0.511 vs. no-transport 0.434); this is the connection's contribution.

H.3 Finding 3: Geometry vs. semantics tradeoff

The plain AE achieves the highest semantic accuracy on MNIST (0.954 vs. CvB's 0.929). On FashionMNIST, the gap disappears (0.862 vs. 0.857). We interpret this as a geometry–semantics tradeoff intrinsic to enforcing bundle constraints: alignment forces the fiber to coordinate across the base, mildly constraining the representation's linear separability. The tradeoff is small ($\approx$ 2–3%), localized to MNIST, and buys a substantial clustering/interpolation benefit. Notably, a linear probe separates decoded reconstructions near-perfectly for every method (> 0.85), suggesting the semantic information is *present* in all representations; CvB simply organizes it in a geometry-preserving manner.

H.4 Finding 4: Raw geodesic fidelity is not the connection's strength

On MGDE and RTD, the bundle-geometry baseline FibeRed dominates (4/5 datasets best MGDE; lowest RTD on all). This is expected: FibeRed is explicitly optimized for geodesic distortion, whereas CvB is a general-purpose autoencoder-based method. The striking result on the geometric datasets is instead *Chart-AE*, the most closely related autoencoder architecture: its MGDE explodes on meshes (5–10$\times$ worse than CvB). On pure geometry, CvB does not beat the non-parametric references: aggregate MGDE 0.376 and RTD 9.13 vs. PCA (0.237, 7.29) and FibeRed (0.208, 3.66), though the between-dataset variance is large. It maintains parity on trustworthiness; its geometric virtue is avoiding Chart-AE's mesh catastrophe, and its advantage over flat autoencoders shows up in the image clustering and topology results. Given page constraints, we present the geometric results pictorially (Figures 2 and 4) and reserve the column-space for the image experiments.

In Figure 4, Chart-AE's embedding is only trained to reconstruct, so unlike PCA/FibeRed or CvB, which explicitly constrain the projection to preserve the manifold's geometry, its latent space has no metric/isometry constraint, and the router-softmax blending of independent charts freely and anisotropically compresses the animal silhouette, producing the squeezed, thin appearance in the panel.

H.5 Finding 5: Non-parametric baselines are complementary, not competing

On FashionMNIST, UMAP and PaCMAP achieve the highest silhouette coefficients (≈ 0.52) and competitive ARI/NMI (Table 8). However, they are non-parametric: they produce an embedding without an encoder, decoder, or transport map. They therefore cannot perform the two tasks central to this paper, decoded interpolation and semantic transfer via the representation. Their strong clustering numbers contextualize the ceiling for clustering quality; Table 4 also shows they suffer severe geodesic distortion on meshes (MGDE often > 30), emphasizing that clustering quality and geometric faithfulness are orthogonal objectives that CvB balances in a single parametric model.

H.6 What the CvB model contributes

On the geometric datasets, CvB is a middle-ground method: substantially better than Chart-AE, comparable to PCA/FibeRed on trustworthiness, and worse than FibeRed on geodesic distortion. On the image datasets, CvB is the best parametric method on clustering and preserves topology best, but it does not dominate ambient interpolation: the flat AE keeps a small semantic-probe edge (0.954 vs. 0.929 on MNIST) and VAEs produce slightly smoother decoded paths. We therefore frame CvB's benefit narrowly: the alignment-driven connection improves clustering and topological organization, at a small, localized ambient-interpolation cost. The connection—specifically its

alignment objective—drives the improvement, as the ablation isolates. In short: for parametric autoencoders with a total space large enough to matter, imposing a consistent fiber transport across the base improves clustering and topological organization at a small, localized ambient-interpolation cost. Whether this transfers to generative sampling or downstream tasks remains an open question (and is checked in Sections D.2 and D.7).

NOTATION

MODEL NETWORKS AND MAPS

Symbol	Meaning
$\varphi(\cdot)$	base encoder / chart map $x \mapsto b \in \mathbb{R}^{d_b}$
$\psi(\cdot,\cdot)$	base-conditioned fiber encoder $(x, b) \mapsto f \in \mathbb{R}^{d_f}$
$D(\cdot,\cdot)$	decoder $(b, f) \mapsto \widehat{x} \in \mathbb{R}^d$
$\mathcal{T}(\cdot,\cdot)$	transport network $(b_i, b_j) \mapsto T_{ij} \in \mathrm{GL}(d_f)$
$T_{ij} = T(b_i, b_j)$	learned transport / connection map between base points
$r(x) = (b, f)$	total-space representation used for clustering/metrics
$n_1(\cdot), n_2(\cdot)$	in-batch nearest- and second-nearest-neighbor index

BUNDLE GEOMETRY

M	base manifold (low-dimensional chart the data lives on)
E	smooth real vector bundle $\pi\colon E \to M$ of rank r
E_b	fiber over b, $E_b = \pi^{-1}(b)$, a vector space of dimension r
π	bundle projection, $\pi \circ f = \mathrm{id}_M$ for a section
f	a (local) section $f\colon U \to E$; f_i the fiber at b_i
b	a base point $b \in M$ / base coordinate
U, U_i, U_j	open charts covering M; overlaps $U_i \cap U_j$
φ_U	local trivialization $\varphi_U\colon \pi^{-1}(U) \to U \times \mathbb{R}^r$
T_{ji}	transition map $T_{ji}\colon U_i \cap U_j \to \mathrm{GL}(r)$
γ	a path in M
PT_γ	parallel transport $E_{\gamma(0)} \to E_{\gamma(1)}$ along γ
∇, R^∇	a connection and its curvature tensor; flat iff $R^\nabla \equiv 0$
g_i	gauge map $g_i\colon U_i \to \mathrm{GL}(r)$ trivializing the cocycle
Id	identity map / trivial transport $T \equiv \mathrm{Id}$

OBJECTIVE AND REGULARIZERS

$\mathcal{L}$	total objective (weighted sum of the terms below)
$\mathcal{L}_{\text{recon}}$	reconstruction term $\mathbb{E}\|D(\varphi(x), \psi(x, \varphi(x))) - x\|^2$
$\mathcal{L}_{\text{align}}$	alignment/parallel-transport term $\mathbb{E}\|T_{ij} f_i - f_j\|^2$
$\mathcal{L}_{\text{cocycle}}$	cocycle-consistency term $\mathbb{E}\|T_{jk} T_{ij} - T_{ik}\|_{\mathrm{F}}^2$
$\mathcal{L}_{\text{orth}}$	base–fiber orthogonality term $\|\mathrm{Cov}_{\text{centered}}(b, f)\|_{\mathrm{F}}^2$
$\mathcal{L}_{\text{var}}$	variance / anti-collapse term $\mathbb{E}_{\text{dims}}[\mathrm{relu}(\sigma_x^2/d_f - \mathrm{Var}(f))]$
$\lambda_{\text{recon}}, \ldots, \lambda_{\text{var}}$	loss weights: $(1.0, 1.0, 0.1, 10^{-3}, 10^{-2})$
$\mathbb{E}$	expectation (over mini-batches; subscripts restrict the set)
$\|\cdot\|, \|\cdot\|_{\mathrm{F}}$	Euclidean and Frobenius norm
σ_x^2	data variance (natural scale of the anti-collapse term)
$\mathrm{Var}(\cdot)$, $\mathrm{Cov}(\cdot)$	variance / covariance of code blocks
D	pairwise distance matrix $\mathsf{D}_{ij} = \|x_i - x_j\|_2$

INFERENCE OPERATIONS

$T_{ij}v$	transported fiber vector $v \in E_{b_i}$ pushed to b_j
$\widehat{x}(t)$	extrinsic interpolant $D((1-t)b_i + tb_j, (1-t)f_i + tf_j)$
$\widehat{x}_{\mathrm{intr}}(t)$	intrinsic (transport-guided) interpolant along subdivided path
t	interpolation fraction, $t \in [0, 1]$
$b(s)$	straight base path $(1-s)b_i + sb_j$
$s_k = k/K$	subdivision points of the base path (K steps)
(b^*, f^*)	codes sampled from the fitted isotropic Gaussian for generation
ConnErr	connection-error diagnostic $\mathbb{E}_{\mathrm{pairs}}\|T_{ij}f_i - f_j\|^2$