STAY ON THE MANIFOLD: HIERARCHICAL BAYESIAN TRAJECTORY-FAMILY TRANSFER FOR SPARSE-TARGET RECONSTRUCTION AND SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

Where should reconstructions and new samples lie when a few observations reveal only fragments of a trajectory? We propose hierarchical Bayesian trajectory principal-manifold sampling (HB-TPM), which learns a family of smooth curves from ordered source tasks and adapts it to sparse targets. For unordered vectors, alternating progression assignments and coefficient updates estimate trajectory structure between and beyond the observed shots under the learned source family. Observation-aware extensions connect the trajectory prior to photon arrivals and image-encoded cardiac cycles. Our analysis characterizes the coefficient directions supplied by the prior and bounds sampling error conditional on curve recovery. Across 30 held-out vector tasks, HB-TPM reduces mean sliced-Wasserstein distance by 21.0–41.8% relative to a source-adapted VAE. Prior-strength and sampling ablations identify coefficient shrinkage as the main supported mechanism. On 100 simulated photon records, adaptation lowers mean relative trajectory RMSE by 50.2% versus target-only fitting and 23.6% versus source-mean prediction. Across three training seeds and five folds on 98 cardiac sequences, HB-TPM with latent-MSE training achieves the best mean reconstruction among the tested configurations. Together, the results demonstrate trajectory-family transfer for sparse-target reconstruction and trajectory-centered sampling when source and target tasks share comparable progression structure.

1 INTRODUCTION

Sparse-target reconstruction and sampling face a structural question: where should missing states and additional samples lie when the observed shots cover only fragments of the relevant variation? Metric-based methods transfer representations or comparison rules (Vinyals et al., 2016; Snell et al., 2017), while interpolation, feature synthesis, and generative adaptation construct new examples (Chawla et al., 2002; Zhang et al., 2018; Hariharan & Girshick, 2017; Ojha et al., 2021). Linear interpolation stays within the observed convex hull, and ambient perturbations can spread away from curved high-density regions. Recovering missing structure is therefore central to estimating a useful trajectory and sampling distribution.

Learned representations organize high-dimensional observations in simpler geometries (Bengio et al., 2013; Pope et al., 2021). We study tasks with a smooth progression, such as an experimental time course or a cardiac cycle. Each task follows its own curve within the representation space. Sparse observations leave many possible continuations, even under smoothness constraints. Related, well-observed tasks offer population-level information about which continuations are plausible.

Functional-data analysis recovers sparse individual trajectories from population structure (Yao et al., 2005), and hierarchical meta-learning formalizes transfer through shared priors (Grant et al., 2018). We transfer a *family of curves*: sources share a broad shape while retaining task-specific coefficients, and target observations adapt this family to a new trajectory (Figure 1). Comparable source progression coordinates make the between-task structure estimable.

Hierarchical Bayesian trajectory principal-manifold sampling (HB-TPM) implements this transfer with a finite-basis empirical-Bayes model. Sources determine a coefficient mean and covariance. Alternating latent progression assignments and Gaussian coefficient updates adapt the curve to

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

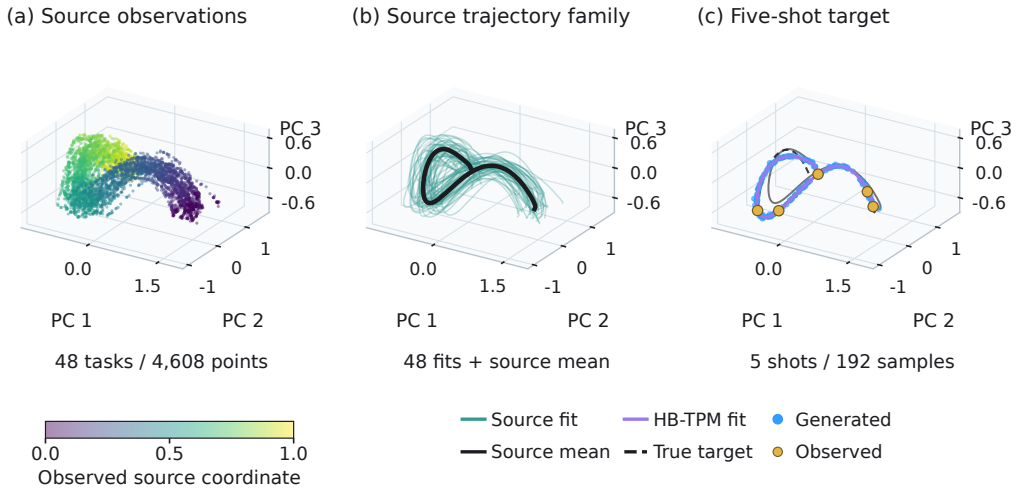


Figure 1: Learning a family before adapting a target. (a) All 48 synthetic source tasks, colored by observed progression. (b) Their fitted curves and coefficient-mean curve (black). (c) Five unordered observations (gold), HB-TPM's fitted target (purple) and 192 generated samples (blue), compared with the source mean (solid) and true target (dashed). All panels share a source-only PCA map retaining 89.1% of source variance. Estimation remains eight-dimensional. This is the first held-out target, not a selected winning case. Source spread and generated points are not credible regions.

unordered target vectors. The prior stabilizes directions that the target leaves weakly observed, and the fitted curve supports sampling beyond the observed shots.

Observation-aware extensions pair the same transfer principle with an event quasi-likelihood for diffusion inference and a learned image encoder for periodic cardiac reconstruction. These extensions use observed acquisition time or cycle phase and test how the observation model shapes target adaptation.

Our contributions are (i) a trajectory-transfer procedure combining coefficient priors, latent target coordinates, and curve-based sampling, with two observation-aware extensions, (ii) an analysis of prior-determined coefficient directions and sampling stability, and (iii) vector, photon, and real-ultrasound evaluations that distinguish curve recovery, target adaptation, and uncertainty calibration. The vector and photon studies establish recovery gains within transferable families. The three-seed cardiac comparison identifies latent-MSE HB-TPM as the best mean reconstruction configuration among those tested.

2 Related Work

Geometry, principal curves, and interpolation. Principal curves describe self-consistent population structure (Hastie & Stuetzle, 1989), and local principal curves follow density geometry (Ozertem & Erdogmus, 2011). HB-TPM uses finite-basis task-progression curves. Statistical manifold estimation and intrinsic-dimension analysis formalize the broader geometric question (Fefferman et al., 2016; Levina & Bickel, 2004). SMOTE, mixup, and manifold mixup interpolate observations or hidden states (Chawla et al., 2002; Zhang et al., 2018; Verma et al., 2019), while decoder geometry and adversarial regularization shape latent paths (Arvanitidis et al., 2018; Berthelot et al., 2019). Our focus is the distribution of task curves that guides sparse-target completion.

Functional data and Bayesian transfer. PACE estimates sparse functions from population structure (Yao et al., 2005; Wang et al., 2016). Gaussian processes (Rasmussen & Williams, 2006) and conditional, latent, attentive, and convolutional neural processes (Garnelo et al., 2018a;b; Kim et al., 2019; Gordon et al., 2020) model related prediction problems. Transformer and diffusion processes enrich their joint predictions (Nguyen & Grover, 2022; Dutordoir et al., 2023; Dou et al., 2025; Hassan et al., 2026). Hierarchical Bayes also provides a perspective on fast adaptation (Finn et al.,

2017; Grant et al., 2018). HB-TPM combines finite-basis Gaussian conditioning with latent progression assignment and curve-based generation. Its distinctive setting is sparse-target adaptation when vector progression coordinates are unobserved; fixed-coordinate functional reconstruction does not address this assignment problem. Appendix J details these connections and the information-matched comparisons they motivate.

Few-shot generative adaptation. Feature hallucination and delta encoders transfer within-class variation (Hariharan & Girshick, 2017; Schwartz et al., 2018), while distribution calibration borrows base-class statistics (Yang et al., 2021). Generative adaptation transfers richer source structure (Ojha et al., 2021). Diffusion, latent diffusion, diffusion autoencoders, and flow matching provide flexible generators (Ho et al., 2020; Rombach et al., 2022; Preechakul et al., 2022; Lipman et al., 2023). Recent few-shot work includes DataDream, HypDAE, geodesic-surface transfer, and unified diffusion distillation/adaptation (Kim et al., 2024; Li et al., 2025; Han et al., 2026; Bahram et al., 2026). Generative transport time need not be physical task progression, and geometric synthesis alone is not ordered trajectory-family inference. Our source-adapted VAE (Kingma & Welling, 2014) is a pooled-source vector baseline, not an evaluation against these image generators.

Inference from photon arrivals. Photon-level Bayesian inference links Brownian diffusion to fluorescence observations (McHale et al., 2004). We study transfer of time-varying diffusion coefficients across records, using a finite-state event quasi-likelihood and empirical Bayes across sources. The observation approximation and its numerical settings are specified in Appendix E.

Cardiac temporal reconstruction. EchoNet-Dynamic studies video-based cardiac function (Ouyang et al., 2020), CAMUS supports cardiac segmentation (Leclerc et al., 2019), and TED provides full-cycle annotations (Painchaud et al., 2022). We use TED to evaluate hidden-frame mask reconstruction from sparse images across the cardiac cycle. Its dense annotations support comparisons of prior adaptation and observation losses within a shared architectural family.

3 TRAJECTORY-FAMILY TRANSFER

HB-TPM represents each task by smooth-basis coefficients. Ordered sources determine which coefficient configurations are plausible, and sparse target observations update that information before sampling. Figure 2 summarizes the workflow. We first develop the vector model, then replace its observation model for photon and cardiac data. Algorithm 1 in Appendix B gives the updates. Fixed-representation and non-hierarchical specializations appear in Appendix H.

3.1 OBSERVATIONS AND WORKING MODEL

Let source task j provide $\{(t_{ji}, \boldsymbol{z}_{ji})\}_{i=1}^{n_j}$, where $t_{ji} \in [0,1]$ is an observed progression coordinate and $\boldsymbol{z}_{ji} \in \mathbb{R}^D$. A target task provides $\mathbf{Z} = [\boldsymbol{z}_1^\top; \ldots; \boldsymbol{z}_K^\top] \in \mathbb{R}^{K\times D}$ without its coordinates. We assume that a common basis $\boldsymbol{\phi} : [0,1] \to \mathbb{R}^p$ describes the tasks through

$$\boldsymbol{\gamma}_{\mathbf{W}}(t) = \mathbf{W}^\top \boldsymbol{\phi}(t), \qquad \mathbf{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_D] \in \mathbb{R}^{p\times D}. \tag{1}$$

Bold upper-case symbols denote matrices and bold lower-case symbols denote vectors. Scalar basis entries are $\phi_b(t)$. The implementation uses the degree-nine Bernstein basis ($p = 10$), rather than the simulator's trigonometric basis.

For each ambient coordinate d, the working hierarchy is

$$\boldsymbol{w}_d \sim \mathcal{N}(\boldsymbol{w}_{0d}, \boldsymbol{\Lambda}_d), \qquad \boldsymbol{z}_{:d} \mid \boldsymbol{w}_d, \boldsymbol{t} \sim \mathcal{N}(\mathbf{F}_{\boldsymbol{t}}\boldsymbol{w}_d, \sigma^2\mathbf{I}_K), \quad (\mathbf{F}_{\boldsymbol{t}})_{ib} = \phi_b(t_i). \tag{2}$$

Coordinates are independent under this prior and each $\boldsymbol{\Lambda}_d$ is diagonal. Hyperparameters are estimated from sources and held fixed during target inference, giving an empirical-Bayes model. The isotropic likelihood in Eq. (2) is a working likelihood; the benchmark simulator instead generates normal-direction perturbations.

3.2 SOURCE ESTIMATION

For each source, form its basis matrix $\mathbf{F}_j$ and observation matrix $\mathbf{Z}_j$. Ridge fits yield

$$\widehat{\mathbf{W}}_j = (\mathbf{F}_j^\top \mathbf{F}_j + \lambda_s \mathbf{I}_p)^{-1} \mathbf{F}_j^\top \mathbf{Z}_j. \tag{3}$$

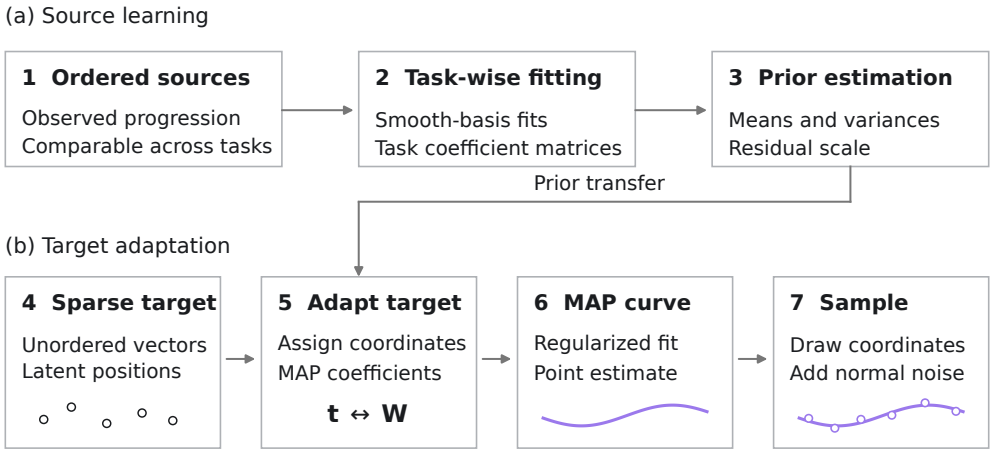


Figure 2: Vector HB-TPM workflow. Ordered sources supply a diagonal coefficient prior and residual scale (1–3). Alternating projection assignments and conditional MAP updates fit the target curve (4–6), which is then perturbed for sampling (7). Assignments may coincide. The curve is a point estimate, not a posterior draw. The sketches depict operations rather than experimental results.

The task-average coefficients and their entrywise sample variances define the prior, with a positive variance floor. Figure 1(b) shows the fitted source curves behind these estimates. A clipped source-residual heuristic supplies σ. Source fitting error is not removed from these plug-in estimates, and the radial scale need not calibrate an isotropic likelihood. Appendix B specifies the settings.

3.3 Sparse target adaptation

With a uniform prior on target grid positions, adaptation minimizes

$$\mathcal{J}(\mathbf{W}, \boldsymbol{t}) = \frac{\|\mathbf{Z} - \mathbf{F}_{\boldsymbol{t}}\mathbf{W}\|_F^2}{2\sigma^2} + \frac{1}{2}\sum_{d=1}^{D}(\boldsymbol{w}_d - \boldsymbol{w}_{0d})^\top \boldsymbol{\Lambda}_d^{-1}(\boldsymbol{w}_d - \boldsymbol{w}_{0d}). \tag{4}$$

Initialize $\mathbf{W}$ at the source mean. Each observation is independently assigned to its closest point on a uniform grid $\mathcal{G} \subset [0, 1]$:

$$\widehat{t}_i \leftarrow \arg\min_{t \in \mathcal{G}} \|\boldsymbol{z}_i - \mathbf{W}^\top \boldsymbol{\phi}(t)\|^2. \tag{5}$$

Indices may coincide. This is latent coordinate assignment rather than recovery of a permutation. Conditional on these assignments, update each coefficient column by

$$\begin{aligned} \boldsymbol{\Sigma}_d &= (\boldsymbol{\Lambda}_d^{-1} + \sigma^{-2}\mathbf{F}_{\widehat{\boldsymbol{t}}}^\top \mathbf{F}_{\widehat{\boldsymbol{t}}})^{-1}, \\ \widehat{\boldsymbol{w}}_d &= \boldsymbol{\Sigma}_d\big(\boldsymbol{\Lambda}_d^{-1}\boldsymbol{w}_{0d} + \sigma^{-2}\mathbf{F}_{\widehat{\boldsymbol{t}}}^\top \boldsymbol{z}_{:d}\big). \end{aligned} \tag{6}$$

The implementation solves linear systems using 401 grid points and at most 20 alternations, stopping when assignments stabilize. Each block minimizes its part of the objective. The joint optimization remains nonconvex, so the result is a local alternating fit.

3.4 Sampling from the fitted curve

Draw $u \sim \mathrm{Unif}[0, 1]$ and form the unit tangent $\boldsymbol{v}(u) = \widehat{\boldsymbol{\gamma}}'(u)/\|\widehat{\boldsymbol{\gamma}}'(u)\|$. For $D \geq 2$ and nonzero tangent and projected direction, let

$$\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_D), \quad \boldsymbol{r}(u) = \frac{(\mathbf{I}_D - \boldsymbol{v}(u)\boldsymbol{v}(u)^\top)\boldsymbol{g}}{\|(\mathbf{I}_D - \boldsymbol{v}(u)\boldsymbol{v}(u)^\top)\boldsymbol{g}\|}, \quad \widetilde{\boldsymbol{z}} = \widehat{\boldsymbol{\gamma}}(u) + a\boldsymbol{r}(u), \tag{7}$$

where $a \sim \mathcal{N}(0, \sigma^2)$ independently. Code uses a 10^{-12} denominator floor for degenerate directions. Equation (7) therefore produces trajectory-centered samples with Gaussian normal displacement and uniform progression sampling. A nonuniform target progression density is not estimated from the K shots.

3.5 OBSERVATION-AWARE TRANSFER TO PHOTON RECORDS

The same coefficient-prior construction can be paired with an indirect observation model. In the photon experiment, let $\mathcal{T}_j$ denote the observed arrival times in a record of duration T and set

$$\log D_j(t) = (1 - t/T)w_{j0} + (t/T)w_{j1}, \qquad \boldsymbol{w}_j = (w_{j0}, w_{j1})^\top \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}). \tag{8}$$

This family uses observed clock time. Successful source fits under the event quasi-likelihood $\ell_{\mathrm{ev}}(\boldsymbol{w}; \mathcal{T}_j)$ determine a mean and covariance, with an approximate inverse-Hessian correction and regularization (Appendix E). This is not unbiased recovery of the latent source distribution.

For a target record, observation-aware HB-TPM (OA-HB-TPM) computes

$$\widehat{\boldsymbol{w}}_c = \operatorname*{arg\,min}_{\boldsymbol{w} \in \mathcal{B}} \left\{ -\ell_{\mathrm{ev}}(\boldsymbol{w}; \mathcal{T}_c) + \tfrac{1}{2}(\boldsymbol{w} - \widehat{\boldsymbol{\mu}})^\top \widehat{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{w} - \widehat{\boldsymbol{\mu}}) \right\}, \tag{9}$$

where $\mathcal{B} = [\log 10, \log 220]^2$. A finite-state Markov-modulated Poisson surrogate and bounded optimization replace the conjugate vector update. Occupancy is estimated from counts and calibrated optics. Neither fitting stage receives true diffusion coefficients, positions, or molecular identities. We evaluate $D(t)$ and marginal endpoint intervals, not generated photon streams. Appendix E specifies the surrogate and its approximations.

3.6 IMAGE-ENCODED PERIODIC TRAJECTORIES

For cardiac images, an encoder maps each observed frame to a 16-dimensional mean and diagonal variance. A three-harmonic Fourier basis maps a Gaussian coefficient matrix to the full latent cycle. A convolutional decoder predicts left-ventricular mask probabilities. Target inference is heteroskedastic Gaussian conditioning on these encodings, with an image-derived circular phase offset. Unlike the vector plug-in fit, the encoder, decoder, phase head, and coefficient prior are trained jointly on source-patient masks. Thus "observation-aware" here denotes a learned latent observation model, not a likelihood for raw ultrasound pixels. Training combines segmentation, latent observation, coefficient-KL, cycle, and temporal losses. Appendix F defines the objective, baseline variants, and numerical safeguards.

4 IDENTIFIABILITY, SHRINKAGE, AND SAMPLING STABILITY

Three statements connect the model to its recovery objective. The first identifies what finite observations leave unresolved, the second characterizes the conditional coefficient update, and the third bounds sampling error given an accurate curve. Their proofs use standard perturbation, Gaussian conditioning, and coupling arguments (Appendix A).

Proposition 1 (Finite observations do not identify a smooth curve)**.** *Given any C^2 curve $\boldsymbol{\gamma} : [0,1] \to \mathbb{R}^D$, $D \geq 1$, and finitely many coordinates $t_1, \ldots, t_K \in [0,1]$, there exists a C^2 curve with a different image that agrees with $\boldsymbol{\gamma}$ in a neighborhood of every t_i.*

This unrestricted-function result shows why source-based structure is needed when a target provides only finitely many observations.

Proposition 2 (Conditional Gaussian shrinkage)**.** *Fix the coordinates, source hyperparameters, and one output dimension in the correctly specified model* (2)*, with $\boldsymbol{\Lambda} \succ 0$ and $\sigma > 0$. The posterior covariance and mean are Eq.* (6)*, and the mean is the unique conditional MAP. For a fixed probability measure q on $[0,1]$ and $\mathbf{B}_q = \int \boldsymbol{\phi}(t)\boldsymbol{\phi}(t)^\top dq(t)$, its posterior integrated squared-error risk is*

$$\mathbb{E}\left[\int \{\boldsymbol{\phi}(t)^\top(\widehat{\boldsymbol{w}} - \boldsymbol{w})\}^2 dq(t) \,\middle|\, \boldsymbol{z} \right] = \operatorname{tr}(\mathbf{B}_q \boldsymbol{\Sigma}). \tag{10}$$

If $\mathbf{F}$ has full column rank, then $\boldsymbol{\Sigma} \prec \sigma^2(\mathbf{F}^\top\mathbf{F})^{-1}$. If it is rank deficient, the positive-definite prior still makes the update well posed.

The mechanism is especially transparent in prior-whitened coordinates. If $\mathbf{A} = \sigma^{-1}\mathbf{F}\boldsymbol{\Lambda}^{1/2}$ has singular values s_b, posterior variances along its right-singular directions are $(1 + s_b^2)^{-1}$. Null directions retain unit prior variance and zero posterior mean displacement from the prior center. Thus when $K < p$, at least $p - K$ coefficient directions remain entirely prior-determined conditional on the indices.

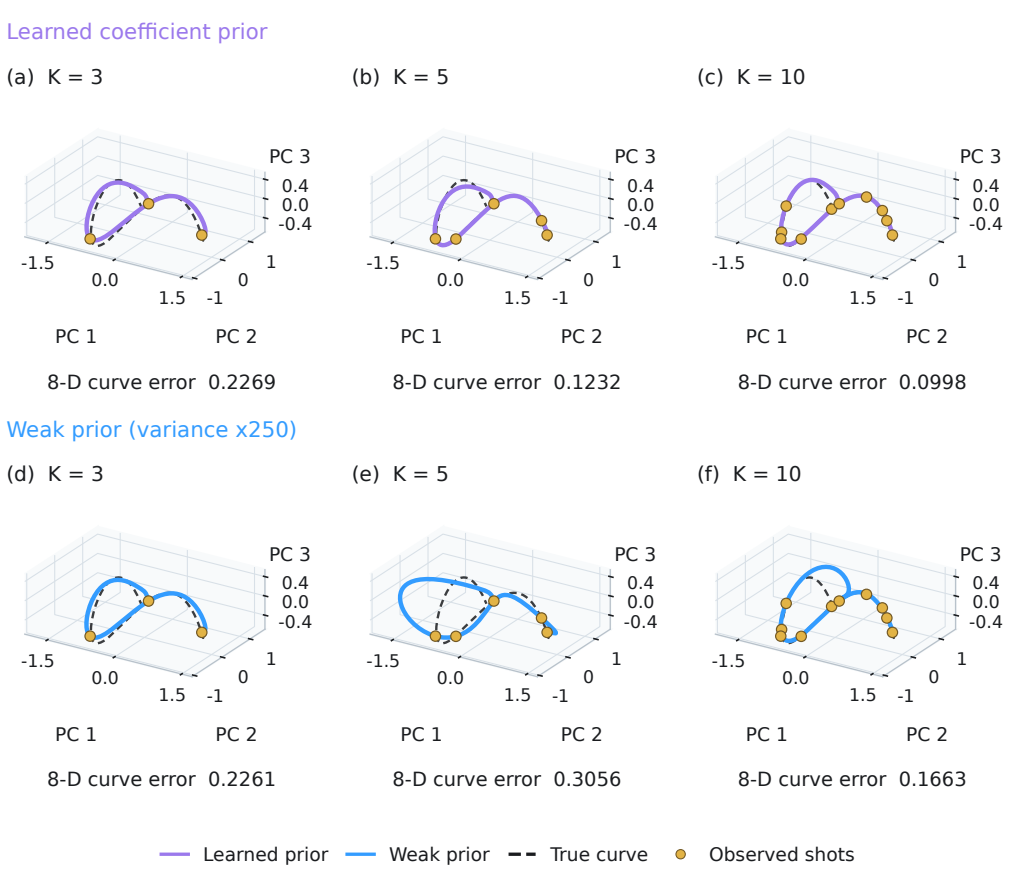


Figure 3: Observations and prior strength in the first held-out target. Columns use nested $K = 3, 5, 10$ shots (gold), with the learned prior above and its variance-inflated ($\times 250$) ablation below. Dashed curves are truth. Panels share a source-only PCA map and limits. Annotations give symmetric curve error in eight dimensions. The learned prior improves recovery at $K = 5, 10$, while the weak prior has slightly lower error at $K = 3$. Table 1 reports aggregate distribution recovery.

Proposition 3 (Sampling stability conditional on curve error)**.** *Fix continuous curves* $\widehat{\boldsymbol{\gamma}}, \boldsymbol{\gamma} : [0,1] \to \mathbb{R}^D$ *with* $\sup_t \|\widehat{\boldsymbol{\gamma}}(t) - \boldsymbol{\gamma}(t)\| \le \epsilon$*. Couple* $\widetilde{\boldsymbol{z}} = \widehat{\boldsymbol{\gamma}}(U) + \boldsymbol{\eta}$ *to* $\boldsymbol{z}^\star = \boldsymbol{\gamma}(U)$ *using the same* $U \in [0,1]$*. If* $\mathbb{E}\|\boldsymbol{\eta}\| \le \tau$*, then for every* L*-Lipschitz* $f : \mathbb{R}^D \to \mathbb{R}$*,*

$$|\mathbb{E}f(\widetilde{\boldsymbol{z}}) - \mathbb{E}f(\boldsymbol{z}^\star)| \le L(\epsilon + \tau). \tag{11}$$

For every $\rho > \epsilon$*,* $\Pr\{\mathrm{dist}(\widetilde{\boldsymbol{z}}, \Gamma) > \rho\} \le \Pr\{\|\boldsymbol{\eta}\| > \rho - \epsilon\}$*, where* Γ *is the true curve image. If* Γ *has reach* $r_0 > \rho$*, points within this tube have a unique nearest point on* Γ*.*

Here reach is the supremum of tube radii for which nearest-point projection is unique (Federer, 1959). For Eq. (7), $\tau = \sigma\sqrt{2/\pi}$ away from degeneracies and the tail is bounded by $2\bar{\Phi}((\rho - \epsilon)/\sigma)$. The resulting stability bound is conditional on the curve approximation error ϵ.

5 Experiments

The experiments test recovery of missing variation, coefficient shrinkage, and adaptation under indirect observations. Tables 1–3 use task-specific comparators for vector distributions, diffusion curves, and cardiac masks. Their shared roles are target fitting, source transfer, and component ablation. Table 4 summarizes datasets and information access. Earlier non-hierarchical image and spiral studies appear in Appendix I.

Table 1: **Synthetic 8-D trajectories: distribution recovery.** Sliced 2-Wasserstein distance (mean $\pm$ standard error over 30 targets, lower is better). Bold marks the lowest displayed means, not significance. Budgets are nested and uncertainty conditions on one source cohort. Only HB-TPM variants use source coordinates. Both VAEs use source standardization.

Method	$K = 3$	$K = 5$	$K = 10$
Gaussian shrinkage	0.2687 ± 0.0137	0.2089 ± 0.0085	0.1786 ± 0.0056
Pairwise interpolation	0.2837 ± 0.0157	0.2010 ± 0.0117	0.1601 ± 0.0090
Naive TPM	0.2910 ± 0.0144	0.1945 ± 0.0121	0.1539 ± 0.0075
Target-trained VAE	0.2762 ± 0.0143	0.2004 ± 0.0100	0.1783 ± 0.0077
Source-adapted VAE	0.1494 ± 0.0037	0.1340 ± 0.0033	0.1270 ± 0.0033
HB-TPM (normal)	0.1180 ± 0.0045	0.0934 ± 0.0047	0.0739 ± 0.0031
HB-TPM (weak prior)	0.1426 ± 0.0165	0.1729 ± 0.0306	0.1273 ± 0.0144
HB-TPM (isotropic)	$\mathbf{0.1175 \pm 0.0049}$	$\mathbf{0.0926 \pm 0.0038}$	$\mathbf{0.0636 \pm 0.0034}$

5.1 Vector benchmark: protocol and information

An eight-dimensional smooth trajectory family supplies 48 source tasks with 96 ordered observations each and 30 held-out targets with nested $K \in \{3, 5, 10\}$ unordered shots. Each target has 192 fresh oracle observations for evaluation. Target coordinates, simulator coefficients, and true curves are excluded from fitting. Truth-based projections and support metrics are evaluation-only (Appendix B).

Means and standard errors describe 30 targets conditional on one source cohort and base seed. The 90 task–budget comparisons are not independent tasks.

Baselines are Gaussian shrinkage, arbitrary-pair interpolation, PCA-ordered naive TPM, and target-trained and source-adapted VAEs. Both VAEs use source standardization. Naive TPM uses the source residual scale. Only HB-TPM variants use source progression coordinates. Thus comparisons combine source-information and modeling differences, rather than isolating architecture under identical information.

The primary metric is empirical sliced 2-Wasserstein distance (SWD) over 192 generated samples and 64 shared projection directions (Bonneel et al., 2015). Secondary metrics are energy distance, generated-to-curve and curve-to-generated distances, and symmetric curve error. All are lower-is-better. The curve distances are nearest-neighbor diagnostics, not probabilistic precision or credible-band coverage. Appendix C provides definitions and all-baseline visual comparisons.

5.2 Distribution recovery and mechanism checks

The default HB-TPM sampler achieves mean SWD of 0.1180, 0.0934, and 0.0739 at $K = 3, 5, 10$ (Table 1), reductions of 21.0%, 30.2%, and 41.8% relative to the source-adapted VAE. It has lower SWD on 28, 28, and 30 of the 30 paired targets, respectively. These comparisons support distribution recovery in this family, subject to the source-information difference described above.

Local fidelity and recovered extent separate interpolation from trajectory transfer. At $K = 3$, pairwise interpolation yields lower generated-to-curve distance (0.1411 versus 0.2584), while HB-TPM yields lower curve-to-generated distance (0.2579 versus 0.6517) and SWD (0.1180 versus 0.2837). Staying near a small observed segment can give good one-sided proximity while leaving much of the trajectory uncovered. Figure 6 visualizes this distinction, complementing the full baseline comparison in Figure 5. Joint assessment of fidelity and extent follows the broader rationale for quality-and-coverage evaluation (Kynkäänniemi et al., 2019).

Prior strength and the gaps between observations. Multiplying coefficient variances by 250, with the source mean and other fitting choices unchanged, increases mean curve error from $0.2588/0.1717/0.0845$ to $0.3047/0.2957/0.1623$ at the three budgets. SWD also worsens. Figure 3 shows what this ablation changes in the first target: the weak-prior curve at $K = 5$ deviates between observed points, whereas the learned prior limits that deviation. At $K = 3$ the weak prior is slightly better, so this example is not a uniform dominance claim. The aggregate evidence supports shrinkage within this family, not every prior strength or an unregularized estimator.

Table 2: **Synthetic photon arrivals: diffusion recovery.** 100 paired targets. Mean relative curve RMSE is lower-is-better. Brackets are 95% whole-record bootstrap intervals conditional on one source cohort. All finite estimates, including nine target-only boundary fits, are retained. Coverage averages 200 recorded indicators for nominal 95% marginal endpoint intervals, not whole curves. The source mean uses no target photons. Success and coverage do not apply to this point predictor. Bold marks the lowest mean error, most successful fits, and coverage closest to 95%, not statistical significance.

Method	Mean error	95% CI	Successful fits	Coverage
Target-only event QMLE	0.4378	[0.3875, 0.4927]	91/100	82%
Source mean (post-hoc)	0.2855	[0.2568, 0.3163]	n/a	—
OA-HB-TPM	**0.2180**	[0.1928, 0.2450]	**100/100**	**91%**

Normal versus isotropic perturbations. The isotropic ablation keeps the curve and matches expected squared noise radius using covariance $(\sigma^2/D)\mathbf{I}_D$. Its mean SWD is lower at every budget, most clearly at $K = 10$ (0.0636 versus 0.0739). Thus the stronger evidence is for curve estimation with shrinkage, not a benefit from normal-only perturbations. Separate sampling draws add Monte Carlo variation to this directional comparison.

An off-trajectory diagnostic. On full-dimensional Gaussian data, Gaussian shrinkage outperforms naive TPM in 82 of 90 comparisons. This warns against imposing a target-only curve on arbitrary distributions, but does not test HB-TPM under family shift: HB-TPM is absent from this diagnostic. The full protocol and results are in Appendix C.3.

5.3 Photon benchmark: transfer through an observation model

We simulate 40 source and 100 target records from one log-linear diffusion family. Thirty-nine successful source fits determine the prior. We compare three documented end-to-end estimators: target-only fitting uses no coefficient prior, the source mean uses no target photons, and OA-HB-TPM combines both. Target fits share events, bounds, calibrations, and quasi-likelihood, and each uses its specified initialization procedure. Table 2 retains all 100 targets, including nine target-only boundary fits. Bootstrap intervals condition on this source cohort.

OA-HB-TPM reduces mean error by 50.2% versus target-only fitting (0.4378 to 0.2180) and by 23.6% versus source-mean prediction (0.2855). The paired differences are -0.2198 (95% interval $[-0.2632, -0.1800]$) and -0.0675 ($[-0.1067, -0.0292]$). It improves 88/100 and 60/100 targets, respectively. The reduction versus target-only fitting remains 47.8% on the 91 jointly successful records. Adaptation therefore adds predictive value beyond the learned mean in this experiment. Calibration is a separate issue: nominal 95% endpoint coverage is 91% (versus 82%), and source endpoint correlation is estimated as 0.041 rather than 0.650. Appendix E reports these diagnostics and the surrogate-model limitations.

5.4 TED: sparse-frame cardiac reconstruction

Protocol and comparisons. TED contains 98 annotated cardiac cycles (Painchaud et al., 2022). Five patient-disjoint test folds have separate validation and source-training patients. Sequences use 24 distinct 112×112 frames, with $K \in \{2, 3, 5, 7, 10\}$ observed images at uniform, random, or clustered times. Hidden images and target masks are excluded from inference. Table 3 averages six configurations over three training seeds and adds an OA setting with observation-noise scale 4.0. Replicates are averaged within patient–budget–pattern cells, followed by equal weighting of cells and seeds. Repeated seeds retain the same 98 patients.

Weak-prior adaptation retains source-trained networks with a broad coefficient prior. *Fixed source coefficients* retains image-derived phase but fixes coefficients. *Latent-MSE training* replaces the observation loss while retaining Gaussian inference. Each system is trained separately.

Reconstruction and calibration. HB-TPM with latent-MSE training achieves the best reported mean on all three reconstruction metrics: Dice 0.7519, HD95 9.4843 pixels, and area RMSE 0.6357.

Table 3: **TED ultrasound: three-seed sparse-frame reconstruction.** 98 patients, five held-out folds. Replicates are averaged within patient–budget–pattern cells, then cells and training seeds receive equal weight. The last row uses observation-noise scale 4.0. Other rows retain the original protocol. Bold marks the best mean reconstruction score and coverage closest to nominal 95%, not statistical significance. Coverage is pointwise soft-area coverage.

Method	Dice ↑	HD95 (px) ↓	Area RMSE ↓	95% coverage
Weak-prior adaptation	0.7454	10.0020	0.6926	43.6%
Fixed source coefficients	0.7472	9.8771	0.6622	—
HB-TPM (latent-MSE training)	**0.7519**	**9.4843**	**0.6357**	**63.6%**
OA-HB-TPM (original)	0.7461	9.9467	0.6728	39.3%
OA-HB-TPM (isotropic prior)	0.7461	9.9472	0.6727	39.1%
OA-HB-TPM (no phase alignment)	0.7462	9.9459	0.6726	39.3%
OA-HB-TPM (noise scale 4.0)	0.7461	9.9404	0.6716	42.7%

The seven rows compare complete configurations, so the result supports the latent-MSE HB-TPM configuration as a whole rather than isolating any single training or prior component.

Observation-noise scale 4.0 increases OA area coverage from 39.3% to 42.7% and lowers area RMSE from 0.6728 to 0.6716, with unchanged Dice at reported precision. Latent-MSE reaches 63.6% coverage. All remain below nominal 95%. Noise scaling therefore yields a modest coverage improvement, not calibrated uncertainty.

6 Discussion and conclusion

HB-TPM makes missing-structure recovery explicit: ordered sources learn a trajectory family, sparse targets adapt it, and a sampler distributes mass around the recovered curve. Vector ablations support coefficient shrinkage as the central mechanism, and photon records demonstrate useful observation-aware adaptation. On TED, latent-MSE training is the strongest tested mean reconstruction configuration. The modest effect of noise scaling separates reconstruction quality from uncertainty calibration.

Scope and next comparisons. The model assumes comparable source progression. Each simulation uses one source cohort, and TED uses one dataset with three training seeds. GP/PACE/neural-process comparisons, frozen-encoder tests, and external cohorts would extend the evidence (Appendices J and F.5). Patient-level rerun records and scale-selection metadata are needed to quantify uncertainty around the updated cardiac means.

Learning a distribution of plausible trajectories supplies structure for sparse-target recovery. Explicit coefficient inference makes the contributions of source information, target observations, and the sampling rule separately testable.

AI USE STATEMENT

Generative AI tools (ChatGPT/Codex) assisted with conceptual development, experiment design, implementation, mathematical formulations and proof writing, result interpretation, literature search, manuscript editing, and figure preparation. Synthetic data were generated by explicit numerical simulators. Vector and photon summaries were checked against experiment-level outputs, and the vector benchmark was reproduced. The three-seed cardiac table uses the reported aggregate results, with its verification scope specified in Appendix F. The authors are responsible for the correctness and originality of all content, including AI-assisted text, code, proofs, and figures.

ETHICS STATEMENT

The vector, photon, and spiral experiments are simulated. The cardiac study uses existing, publicly available, de-identified TED ultrasound sequences and annotations (Painchaud et al., 2022). No new participants were recruited. The supplementary package supplies derived evaluation records and code, not patient image arrays. Archived digit/face illustrations in Appendix I.1 are not used to infer personal attributes or support clinical claims. A misspecified source prior or image representation can bias predictions, and the observed interval undercoverage precludes reliance on the reported uncertainty for clinical decisions. The study does not evaluate clinical deployment or downstream task performance.

REPRODUCIBILITY STATEMENT

Algorithm 1 specifies vector inference. Sections 3.5 and 3.6 describe the observation-aware variants. Appendices A–F give proofs, settings, metrics, and diagnostics, and Appendix G summarizes the reproduction protocol. Vector and photon means and paired contrasts were recomputed from experiment-level results. Table 3 reports updated three-seed cardiac aggregate means, with the aggregation and evaluation protocol specified in Appendix F. Appendix F distinguishes these means from the earlier single-seed diagnostics and specifies the records needed for paired uncertainty and calibration checks. Appendix I separates the reproduced spiral pilot from archival image panels that cannot be independently regenerated.

REFERENCES

Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. Latent space oddity: on the curvature of deep generative models. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1710.11379.

Yara Bahram, Mélodie Desbos, Mohammadhadi Shateri, and Eric Granger. Uni-DAD: Unified distillation and adaptation of diffusion models for few-step few-shot image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026. URL https://arxiv.org/abs/2511.18281.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003. doi: 10.1162/089976603321780317.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013. doi: 10.1109/TPAMI.2013.50. URL https://arxiv.org/abs/1206.5538.

David Berthelot, Colin Raffel, Aurko Roy, and Ian Goodfellow. Understanding and improving interpolation in autoencoders via an adversarial regularizer. In *International Conference on Learning Representations*, 2019. URL https://iclr.cc/virtual/2019/poster/799.

Christopher M. Bishop, Markus Svensén, and Christopher K. I. Williams. GTM: The generative topographic mapping. *Neural Computation*, 10(1):215–234, 1998. doi: 10.1162/089976698300017953.

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015. doi: 10.1007/s10851-014-0506-3.

Wessel P. Bruinsma, Stratis Markou, James Requeima, Andrew Y. K. Foong, Tom R. Andersson, Anna Vaughan, Anthony Buonomo, J. Scott Hosking, and Richard E. Turner. Autoregressive conditional neural processes. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2303.14468.

Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16: 321–357, 2002. doi: 10.1613/jair.953.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.14687.

Hongkun Dou, Junzhe Lu, Zeyu Li, Xiaoqing Zhong, Wen Yao, Lijun Yang, and Yue Deng. Score-based neural processes. *IEEE Transactions on Neural Networks and Learning Systems*, 36(6): 11589–11601, 2025. doi: 10.1109/TNNLS.2025.3527223.

Vincent Dutordoir, Alan Saul, Zoubin Ghahramani, and Fergus Simpson. Neural diffusion processes. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 8990–9012, 2023. URL https://proceedings.mlr.press/v202/dutordoir23a.html.

Bradley Efron and Carl Morris. Stein's estimation rule and its competitors—an empirical Bayes approach. *Journal of the American Statistical Association*, 68(341):117–130, 1973. doi: 10.1080/01621459.1973.10481350.

Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3): 418–491, 1959. doi: 10.1090/S0002-9947-1959-0110078-1.

Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016. doi: 10.1090/jams/852. URL https://arxiv.org/abs/1310.0425.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1126–1135, 2017. URL https://proceedings.mlr.press/v70/finn17a.html.

Marta Garnelo, Dan Rosenbaum, Chris J. Maddison, Tiago Ramalho, David Saxton, Murray Shanahan, Yee Whye Teh, Danilo J. Rezende, and S. M. Ali Eslami. Conditional neural processes. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1704–1713, 2018a. URL https://proceedings.mlr.press/v80/garnelo18a.html.

Marta Garnelo, Jonathan Schwarz, Dan Rosenbaum, Fabio Viola, Danilo J. Rezende, S. M. Ali Eslami, and Yee Whye Teh. Neural processes. *arXiv preprint arXiv:1807.01622*, 2018b. URL https://arxiv.org/abs/1807.01622.

Jonathan Gordon, Wessel P. Bruinsma, Andrew Y. K. Foong, James Requeima, Yann Dubois, and Richard E. Turner. Convolutional conditional neural processes. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/1910.13556.

Erin Grant, Chelsea Finn, Sergey Levine, Trevor Darrell, and Thomas Griffiths. Recasting gradient-based meta-learning as hierarchical Bayes. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1801.08930.

Yuexing Han, Liheng Ruan, and Bing Wang. Few-shot image generation via information transfer from the built Geodesic surface. *Pattern Recognition*, 172:112293, 2026. doi: 10.1016/j.patcog.2025.112293. URL https://www.sciencedirect.com/science/article/abs/pii/S0031320325009549.

Bharath Hariharan and Ross Girshick. Low-shot visual recognition by shrinking and hallucinating features. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3018–3027, 2017. URL https://openaccess.thecvf.com/content_iccv_2017/html/Hariharan_Low-Shot_Visual_Recognition_ICCV_2017_paper.html.

Conor Hassan, Nasrulloh Loka, Cen-You Li, Daolang Huang, Paul E. Chang, Yang Yang, Francesco Silvestrin, Samuel Kaski, and Luigi Acerbi. Efficient autoregressive inference for transformer probabilistic models. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/8c5111676c19e5945473acf20ad7026f-Abstract-Conference.html.

Trevor Hastie and Werner Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84(406):502–516, 1989. doi: 10.1080/01621459.1989.10478797.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020. URL https://arxiv.org/abs/2006.11239.

Gavin Kerrigan, Giosue Migliorini, and Padhraic Smyth. Functional flow matching. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 3934–3942, 2024. URL https://proceedings.mlr.press/v238/kerrigan24a.html.

Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. Attentive neural processes. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1901.05761.

Jae Myung Kim, Jessica Bader, Stephan Alaniz, Cordelia Schmid, and Zeynep Akata. DataDream: Few-shot guided dataset generation. In *European Conference on Computer Vision*, 2024. URL https://www.ecva.net/papers/eccv_2024/papers_ECCV/html/8991_ECCV_2024_paper.php.

Diederik P. Kingma and Max Welling. Auto-encoding variational Bayes. In *International Conference on Learning Representations*, 2014. URL https://arxiv.org/abs/1312.6114.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/0234c510bc6d908b28c70ff313743079-Abstract.html.

Sarah Leclerc, Erik Smistad, João Pedrosa, Andreas Østvik, Frederic Cervenansky, Florian Espinosa, Torvald Espeland, Erik Andreas Rye Berg, Pierre-Marc Jodoin, Thomas Grenier, Carole Lartizien, Jan D'hooge, Lasse Lovstakken, and Olivier Bernard. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Transactions on Medical Imaging*, 38 (9):2198–2210, 2019. doi: 10.1109/TMI.2019.2900516. URL https://arxiv.org/abs/1908.06948.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791. URL https://leon.bottou.org/papers/lecun-98h.

Elizaveta Levina and Peter J. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems*, volume 17, 2004. URL https://proceedings.neurips.cc/paper/2004/hash/74934548253bcab8490ebd74afed7031-Abstract.html.

Lingxiao Li, Kaixuan Fan, Boqing Gong, and Xiangyu Yue. HypDAE: Hyperbolic diffusion autoencoders for hierarchical few-shot image generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17119–17128, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Li_HypDAE_Hyperbolic_Diffusion_Autoencoders_for_Hierarchical_Few-shot_Image_Generation_ICCV_2025_paper.html.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2210.02747.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, 2015. URL https://liuziwei7.github.io/projects/FaceAttributes.html.

Emile Mathieu, Vincent Dutordoir, Michael J. Hutchinson, Valentin De Bortoli, Yee Whye Teh, and Richard E. Turner. Geometric neural diffusion processes. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/a797c2d2e0c1fdabf4d1ab8cd0b465c6-Abstract.html.

Kevin McHale, Andrew J. Berglund, and Hideo Mabuchi. Bayesian estimation for species identification in single-molecule fluorescence microscopy. *Biophysical Journal*, 86(6):3409–3422, 2004. doi: 10.1529/biophysj.103.038414.

Tung Nguyen and Aditya Grover. Transformer neural processes: Uncertainty-aware meta learning via sequence modeling. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 16569–16594, 2022. URL https://proceedings.mlr.press/v162/nguyen22b.html.

Utkarsh Ojha, Yijun Li, Jingwan Lu, Alexei A. Efros, Yong Jae Lee, Eli Shechtman, and Richard Zhang. Few-shot image generation via cross-domain correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10743–10752, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Ojha_Few-Shot_Image_Generation_via_Cross-Domain_Correspondence_CVPR_2021_paper.html.

David Ouyang, Bryan He, Amirata Ghorbani, Neal Yuan, Joseph Ebinger, Curtis P. Langlotz, Paul A. Heidenreich, Robert A. Harrington, David H. Liang, Euan A. Ashley, and James Y. Zou. Video-based AI for beat-to-beat assessment of cardiac function. *Nature*, 580(7802):252–256, 2020. doi: 10.1038/s41586-020-2145-8. URL https://echonet.github.io/dynamic/.

Umut Ozertem and Deniz Erdogmus. Locally defined principal curves and surfaces. *Journal of Machine Learning Research*, 12(34):1249–1286, 2011. URL https://www.jmlr.org/papers/v12/ozertem11a.html.

Nathan Painchaud, Nicolas Duchateau, Olivier Bernard, and Pierre-Marc Jodoin. Echocardiography segmentation with enforced temporal consistency. *IEEE Transactions on Medical Imaging*, 41 (10):2867–2878, 2022. doi: 10.1109/TMI.2022.3173669. URL https://www.creatis.insa-lyon.fr/~bernard/publis/tmi_2022_nathan.pdf.

Phillip Pope, Chen Zhu, Ahmed Abdelkader, Micah Goldblum, and Tom Goldstein. The intrinsic dimension of images and its impact on learning. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2104.08894.

Konpat Preechakul, Nattanat Chatthee, Suttisak Wizadwongsa, and Supasorn Suwajanakorn. Diffusion autoencoders: Toward a meaningful and decodable representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10619–10629, 2022. URL https://arxiv.org/abs/2111.15640.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. URL https://gaussianprocess.org/gpml/.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022. URL https://arxiv.org/abs/2112.10752.

Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000. doi: 10.1126/science.290.5500.2323.

Eli Schwartz, Leonid Karlinsky, Joseph Shtok, Sivan Harary, Mattias Marder, Abhishek Kumar, Rogerio Feris, Raja Giryes, and Alex M. Bronstein. Delta-encoder: an effective sample synthesis method for few-shot object recognition. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/1714726c817af50457d810aae9d27a2e-Abstract.html.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/cb8da6767461f2812ae4290eac7cbc42-Abstract.html.

Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000. doi: 10.1126/science.290.5500.2319.

Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua Bengio. Manifold mixup: Better representations by interpolating hidden states. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 6438–6447, 2019. URL https://proceedings.mlr.press/v97/verma19a.html.

Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, volume 29, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/90e1357833654983612fb05e3ec9148c-Abstract.html.

Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016. doi: 10.1146/annurev-statistics-041715-033624.

Shuo Yang, Lu Liu, and Min Xu. Free lunch for few-shot learning: Distribution calibration. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2101.06395.

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005. doi: 10.1198/016214504000001745. URL https://utstat.utoronto.ca/fyao/2005-jasa.pdf.

Biao Zhang and Peter Wonka. Functional diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4723–4732, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Zhang_Functional_Diffusion_CVPR_2024_paper.html.

Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1710.09412.