FROM FLOW MATCHING TO TYPE I ERROR GUARANTEES IN ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Anomaly detection asks whether a new observation is likely to be generated by a reference distribution that characterizes normal behavior. We address this problem by learning a transport, via flow matching, that maps the reference distribution to a standard Gaussian. Once trained, this map lets us test a new sample by checking how far its transported representation lies from the Gaussian: under the reference distribution, its squared norm follows a chi-squared statistic, yielding a threshold with an explicit type I error rate. We further derive upper bounds that link the training loss of the underlying model to the resulting detection guarantees, showing that a well-trained model yields a test whose behavior remains close to this ideal. These results are made explicit in two cases, Gaussian and spherically symmetric reference distributions, and validated numerically. We evaluate the resulting detector on three distinct applications: detecting images that do not belong to a reference set, detecting targets in radar signals against background clutter, and classifying imbalanced biological data. Across these settings, our method is competitive against existing detectors.

1 INTRODUCTION

This work deals with the problem of anomaly detection, which we define as follows: given a *reference* distribution $P_{\mathcal{H}_0}$ and an observation, is this observation likely to have been drawn from $P_{\mathcal{H}_0}$? If not, the observation is called an anomaly, an outlier, or abnormal data. An application of this problem is out-of-distribution detection in vision Hendrycks & Gimpel (2017): distinguishing images of common objects such as dogs, trucks, and horses (data from CIFAR-10) from images of house numbers (data from SVHN), when reference data are, say, the former. Another application is imbalanced data classification in biology. The last application addressed in this work is target detection in radar: the reference distribution is given by radar observations of the sea, and the appearance of a boat in the scene constitutes an anomaly to detect Maio & Greco (2015).

It has been a few decades since machine learning began addressing anomaly detection problems, as reviewed in Chandola et al. (2009); Pang et al. (2021). Since abnormal data are usually rare and unlabeled, methods are traditionally trained on reference data. Older machine learning methods often project data into a different space to define a detection score Tax & Duin (2004); Ruff et al. (2018). This idea has since been revisited with deep, often self-supervised, representations that use distances or activation statistics in feature space to separate anomalies (Sun et al., 2021; Djurisic et al., 2022). In parallel, generative models reconstruct data and derive a detection score from the likelihood or reconstruction error, with numerous variants (An & Cho, 2015; Akcay et al., 2019), following the early observation that a well-trained generative model should assign low likelihood to out-of-distribution inputs (Bishop, 1994). More recent works, such as Nalisnick et al. (2019) and Choi et al. (2018), have, however, shown that generative models can assign higher likelihood to out-of-distribution inputs than to training data.

A separate line of deep learning work is based on diffusion models. Well-known for generation, they work by adding noise to data and learning to reverse this process to recover the original distribution Ho et al. (2020). For anomaly detection, the resulting score can rely on the reconstruction error or on other quantities derived from the noising process to distinguish test samples from training data Graham et al. (2023a); Liu et al. (2023b); Shoushtari et al. (2026). Due to the computational

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.

cost of training such models on the reference distribution, recent work instead proposes detection algorithms based on a diffusion model pretrained on standard datasets Barkley et al. (2026).

However, this can remain problematic for domains where data differ substantially from natural images, such as radar or biological data, for which no such generic pretrained backbone exists. At the same time, flow matching (FM) has been proposed for generative modeling, matching diffusion models in generation quality with fewer integration steps. Concretely, FM is a deep learning approach that learns an ordinary differential equation (ODE) that transports between two empirically observed distributions; the learned quantity is the ODE's velocity field Lipman et al. (2023); Liu et al. (2023a).The most common approach is to estimate the optimal transport path toward a centered Gaussian distribution, whose properties are well known.

This work leverages the strength of FM for anomaly detection by analyzing the transport from the reference distribution $P_{\mathcal{H}_0}$ to a centered Gaussian distribution. Once the flow is trained to transport $P_{\mathcal{H}_0}$ to $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, a natural detection test is to check whether the transported sample still belongs to this Gaussian distribution. Concretely, we propose a detector that checks whether the squared norm of the transported sample follows a χ^2 distribution, which yields an explicit expression for the type I error of the detection test. In addition, this work derives bounds on the transport estimation error and on the resulting type I error, both depending on the training loss of the velocity field, guaranteeing accurate detection when the training loss is sufficiently small. Theoretical results are validated numerically. We finally apply the algorithm to three application domains mentioned earlier: out-of-distribution detection in images, radar target detection, and biological classification. The algorithm is competitive with classical methods in radar detection and biological classification, and achieves the highest average AUROC among methods trained solely on the reference distribution in images.

Related works. Literature on flow matching for anomaly detection remains scarce. In robotics, the norm of the noise estimate obtained from a learned flow has been used as a score for detecting policy failures or triggering intervention during robot manipulation Sun et al. (2026); Xu et al. (2025). Also, Li et al. (2025) proposes a detection method for tabular data inspired by flow matching, parameterizing the velocity field as an MLP that predicts, at each time t, a contraction vector toward a fixed target, avoiding the need to solve an ODE. Closer to our theoretical concerns, Pham (2026) studies the generalization of conditional flow matching to unseen conditioning variables and derives theoretical bounds, but this differs from our problem, as we consider a binary detection test on the data itself.

This work proposes three main contributions:

- Formulating anomaly detection as a transport problem, addressed through flow matching, and deriving a detector with an explicit type I error.
- Linking the Conditional Flow Matching (CFM) training loss to the detection statistic obtained from the reversed flow, deriving bounds on the transport estimation error and on the resulting type I error - with closed-form expressions of the transport map when $P_{\mathcal{H}_0}$ follows a spherical distribution - and validating these results numerically.
- Evaluating the detector across several application domains: out-of-distribution detection in images, biological classification, and radar target detection, where it remains competitive with existing methods.

2 Conditional Flow Matching

We call velocity field a time-dependent function $\mathbf{u} : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$, written $\mathbf{u}_t(\mathbf{x}) := \mathbf{u}(t, \mathbf{x})$. A velocity field $\mathbf{u}_t$ can be used to construct a time-dependent map, called a flow, $\psi : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$, written $\psi_t(\mathbf{x}) := \psi(t, \mathbf{x})$, defined via the ODE:

$$\frac{d}{dt}\psi_t(\mathbf{x}) = \mathbf{u}_t(\psi_t(\mathbf{x})), \tag{1}$$

$$\psi_0(\mathbf{x}) = \mathbf{x}. \tag{2}$$

At each time t, the flow induces a *probability path* p_t, the distribution of $\psi_t(\mathbf{x}_0)$ when $\mathbf{x}_0 \sim P_{\mathcal{H}_0}$; we say that $\mathbf{u}_t$ *generates* p_t.

Let $\mathbf{x}_1$ be a random variable distributed according to the reference distribution $P_{\mathcal{H}_0}$ (with access to data samples). The goal of flow matching is to infer a probability path p_t on $[0, 1]$ from a known source distribution p_0 (here $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$) to a target p_1 approximately equal in distribution to $P_{\mathcal{H}_0}$.

Given a target probability density path $p_t(\mathbf{x})$ together with a vector field $\mathbf{u}_t(\mathbf{x})$ generating it, the flow matching objective is defined as

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{t,\mathbf{x}_0}\left[\|\mathbf{v}_\theta(\psi_t(\mathbf{x}_0)) - \mathbf{u}_t(\psi_t(\mathbf{x}_0))\|^2\right], \qquad t \sim \mathcal{U}[0, 1]. \tag{3}$$

with $\mathbf{v}_\theta$ a neural network and θ its learnable parameters. $\mathcal{U}[0, 1]$ is the uniform distribution.

In practice, however, this field is intractable to compute. The authors of Lipman et al. (2023) instead propose to estimate the conditional vector field $\mathbf{u}_t(\psi_t(\mathbf{x}_0) \mid \mathbf{x}_1)$, and minimize what we call the Conditional Flow Matching (CFM) loss,

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t,\mathbf{x}_1,\mathbf{x}_0}\left[\|\mathbf{v}_\theta(\psi_t(\mathbf{x}_0)) - \mathbf{u}_t(\psi_t(\mathbf{x}_0) \mid \mathbf{x}_1)\|^2\right]. \tag{4}$$

It has been shown that the gradients of the two losses with respect to θ are equal.

There remains, however, an infinite number of possible conditional probability paths between two distributions. A convenient choice for flow matching, again exposed in Lipman et al. (2023), is a linear interpolation in time:

$$\psi_t(\mathbf{x}_0) = (1 - t)\,\mathbf{x}_0 + t\,\mathbf{x}_1, \qquad t \in [0, 1], \tag{5}$$

for which the conditional velocity field admits the closed form $\mathbf{u}_t(\psi_t(\mathbf{x}_0) \mid \mathbf{x}_1) = \mathbf{x}_1 - \mathbf{x}_0$. This eventually leads to the loss we use in this work:

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}_1}\left[\|\mathbf{v}_\theta(\psi_t(\mathbf{x}_0)) - (\mathbf{x}_1 - \mathbf{x}_0)\|^2\right]. \tag{6}$$

It is worth mentioning that to further straighten the learned trajectories and reduce the number of integration steps at inference, it is common to apply the iterative *reflow* procedure of Liu et al. (2023a) (see Appendix D.1 for details).

3 A Calibrated Test by Transport to a Gaussian

In detection, it is common to consider, for data $\mathbf{x} \in \mathbb{R}^d$, the following test:

$$\mathcal{H}_0 : \mathbf{x}^{\mathcal{H}_0} \sim P_{\mathcal{H}_0}, \tag{7}$$

$$\mathcal{H}_1 : \mathbf{x}^{\mathcal{H}_1} \sim P_{\mathcal{H}_1}, \tag{8}$$

with $P_{\mathcal{H}_0}$ the reference distribution, known through data samples, and $P_{\mathcal{H}_1}$ distinct from $P_{\mathcal{H}_0}$, so that $\mathcal{H}_0$ describes reference, or normal, samples.

3.1 Ideal transport map to a Gaussian

Given a diffeomorphic map $f : \mathbb{R}^d \to \mathbb{R}^d$ and a distribution with density p, the *push-forward* of p by f, denoted $f_{\#}\,p$, is the density of $f(\mathbf{x})$ when $\mathbf{x} \sim p$..

Generative modeling can be framed as the problem of finding a transport map $T : \mathbb{R}^d \to \mathbb{R}^d$ from $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ to a target distribution. For solving the detection problem described by equation 7, we reason the other way around: we map data to $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ through T^{-1}, and use this mapping to check whether the sample is likely drawn from $P_{\mathcal{H}_0}$ or not, as developed below.

We assume throughout that velocity fields are Lipschitz in space, uniformly in t. By the Cauchy–Lipschitz theorem, ψ_t is a diffeomorphism for every $t \in [0, 1]$: distinct trajectories cannot cross, since doing so would force them to coincide for all t. We denote $T := \psi_1$, so that T is invertible — its inverse obtained by integrating $-\mathbf{u}_t$ backward in time — and satisfies, by construction,

$$T_{\#}\mathcal{N}(\mathbf{0}, \mathbf{I}_d) = P_{\mathcal{H}_0}. \tag{9}$$

We emphasize that no optimality of T (in the sense of Brenier (1991)) is required: it suffices that

$$T_{\#}^{-1} P_{\mathcal{H}_0} = \mathcal{N}(\mathbf{0}, \mathbf{I}_d). \tag{10}$$

This property holds only under $\mathcal{H}_0$: there is no reason for $T_{\#}^{-1} P_{\mathcal{H}_1}$ to be Gaussian as well. We refer to T as the *ideal* transport map; in practice, it will be estimated by a map T_θ, introduced in Section 3.3. The detection test proposed in this work, defined in the next section, operates on data after transportation through T^{-1}.

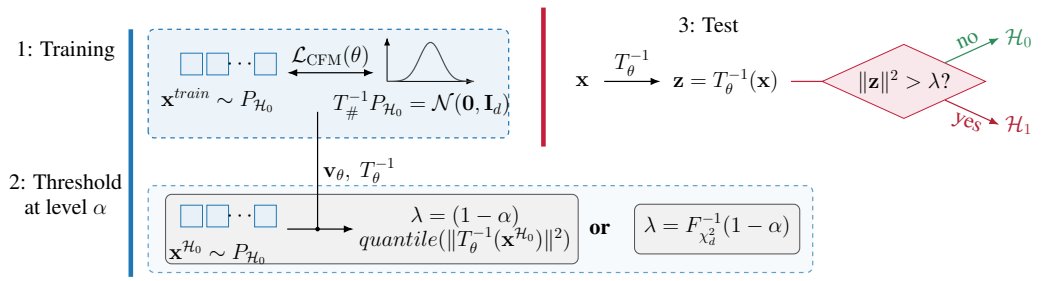

Figure 1: The algorithm proceeds in three steps: (1) training of the velocity field $\mathbf{v}_\theta$ by minimizing $\mathcal{L}_{\text{CFM}}(\theta)$; (2) estimation of reference distribution ${T_\theta^{-1}}_\# P_{\mathcal{H}_0}$ from held-out $\mathcal{H}_0$ samples, and calibration of threshold λ from resulting scores $\|T_\theta^{-1}(\mathbf{x}^{\mathcal{H}_0})\|^2$ or by using the exact expression $F_d^{-1}(1-\alpha)$; (3) test-time decision by comparing squared norm $\|\mathbf{z}\|^2$ of a transported sample $\mathbf{z} := T_\theta^{-1}(\mathbf{x})$ to λ.

3.2 The Detection Test and Its Level

For any sample $\mathbf{x} \in \mathbb{R}^d$, we define the detection statistic

$$S(\mathbf{x}) := \left\|T^{-1}(\mathbf{x})\right\|^2. \tag{11}$$

Under $\mathcal{H}_0$, since $T_\#^{-1} P_{\mathcal{H}_0} = \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ (see Section 3.1), the statistic follows a chi-squared distribution with d degrees of freedom: $S(\mathbf{x}^{\mathcal{H}_0}) \sim \chi^2_d$.

Given a threshold $\lambda \geq 0$, we define the detection test $\varphi(\mathbf{x}) = \mathbf{1}_{\{S(\mathbf{x})>\lambda\}}$, where $\varphi(\mathbf{x}) = 1$ indicates a decision in favor of $\mathcal{H}_1$.

Type I error. The type I error, also called probability of false alarm, is the probability that the test decides in favor of $\mathcal{H}_1$ while $\mathcal{H}_0$ holds:

$$\alpha(\lambda) := \mathbb{P}(\mathcal{H}_1 \mid \mathcal{H}_0) = \mathbb{P}(\varphi(\mathbf{x}) = 1 \mid \mathcal{H}_0) = \mathbb{P}(S(\mathbf{x}) > \lambda \mid \mathcal{H}_0). \tag{12}$$

Since $S(\mathbf{x}^{\mathcal{H}_0}) \sim \chi^2_d$ under $\mathcal{H}_0$, this admits the closed form $\alpha(\lambda) = 1 - F_{\chi^2_d}(\lambda)$, where $F_{\chi^2_d}$ is the CDF of χ^2_d. Conversely, a threshold achieving a target level α is obtained by setting λ to the $(1-\alpha)$-quantile of χ^2_d. Note that this threshold depends only on d, not on $P_{\mathcal{H}_0}$ itself: the level α can be guaranteed exactly regardless of the reference distribution.

Type II error and power. The type II error, or missed detection, is the probability that the test decides in favor of $\mathcal{H}_0$ while $\mathcal{H}_1$ holds:

$$\beta(\lambda) := \mathbb{P}(\mathcal{H}_0 \mid \mathcal{H}_1) = \mathbb{P}(\varphi(\mathbf{x}) = 0 \,|\, \mathcal{H}_1) = \mathbb{P}(S(\mathbf{x}) \leq \lambda \,|\, \mathcal{H}_1)\,, \tag{13}$$

and the power of the test, or probability of detection, is $P_d(\lambda) := 1 - \beta(\lambda)$.

Contrary to the level α given in equation 12, the power cannot be expressed in closed form in general: it depends on $P_{\mathcal{H}_1}$, which is generally not modeled explicitly. This asymmetry is intrinsic to the construction, since the transport is fitted to $P_{\mathcal{H}_0}$ alone. It also delimits what can be established: namely, through the paper, we characterize the power exactly in two tractable settings, when $\mathcal{H}_1$ is a deterministic signal added to a spherical $\mathcal{H}_0$ (Section 5, Example 1), and its first-order behavior for spherically symmetric $\mathcal{H}_0$ (Example 2); and we otherwise evaluate it empirically as P_d at a fixed probability of false alarm (Section 6.2) and through the AUROC (Section 6.1).

3.3 Detection algorithm presentation

We denote T_θ^{-1} the estimated flow corresponding to the estimated velocity $\mathbf{v}_\theta$, as described in section 2. This flow is employed in the detection scheme of this work, described by figure 1 : T_θ is an estimation of the ideal map T defined in section 3.1. For the detection guarantee formalized in Section 4 to hold, training data should be drawn from $\mathcal{H}_0$: under this condition, if the flow is well estimated, ${T_\theta^{-1}}_\# P_{\mathcal{H}_0}$ is expected to be close to a Gaussian distribution as stated by equation 10, and corresponding squared norms should follow a χ^2_d distribution. Figure 2 shows how closely the transported data follows a normal distribution on K-distribution, CIFAR and SVHN datasets.

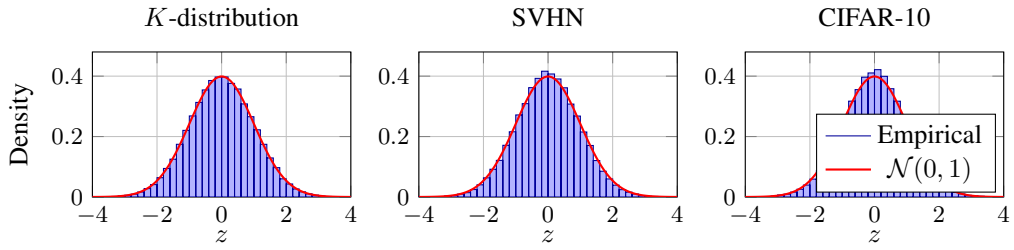

Figure 2: Marginal distribution of the transported data $T_\theta^{-1}(\mathbf{x})$ (one coordinate, pooled across dimensions where applicable), compared to $\mathcal{N}(0,1)$. Kolmogorov–Smirnov statistics on an independent held-out sample ($n = 20{,}000$) confirm the close agreement: $D = 0.006$ (K-distribution), $D = 0.011$ (SVHN), $D = 0.012$ (CIFAR-10). Quantile–quantile plots are reported in Figure 8.

Effect of the transport on distribution separation To illustrate the relevance of transport for out-of-distribution detection, we consider two K-distributions with close parameters for $\mathcal{H}_0$ and $\mathcal{H}_1$. A similar setting is considered for the mixture-of-K distributions. Since the transport is trained to map $\mathcal{H}_0$ samples to a standard Gaussian, it has no reason to do the same for $\mathcal{H}_1$ samples: Figure 3 shows that, while the two distributions are nearly indistinguishable before transport, they become markedly more separated afterward.

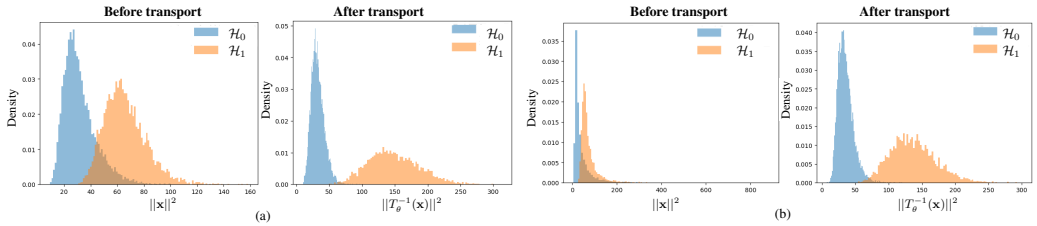

Figure 3: Effect of the learned inverse transport on the separability of $\mathcal{H}_0$ and $\mathcal{H}_1$ distributions. Both panels show the density of a squared-norm statistic for a held-out sample under $\mathcal{H}_0$, following the same distribution as the training data of the transport, and for an out-of-distribution sample under $\mathcal{H}_1$. (a) K-distributions. (b) Mixture of K-distributions. Simulation details are in Appendix D.3.

4 From Training Loss to Level Guarantees

While the flow matching model is trained by minimizing a velocity loss, the detection procedure relies on the transported representation provided by T_θ^{-1}.

Lemma 1. *Assume the estimated velocity field $\mathbf{v}_\theta$ is L_t-Lipschitz in space. Then, for any $\eta \in (0,1)$, there exists a set $\mathcal{X}$ with $\mathbb{P}(\mathbf{x} \in \mathcal{X}) \geq 1-\eta$ such that*

$$\forall \mathbf{x} \in \mathcal{X}, \quad \left\| T_\theta^{-1}(\mathbf{x}) - T^{-1}(\mathbf{x}) \right\| \leq \varepsilon := K\sqrt{\mathcal{L}_{\mathrm{CFM}}(\theta)/\eta}, \qquad K = \exp\left(\int_0^1 L_t \, dt\right). \tag{14}$$

Proof. See Appendix A.1. □

In particular, $\varepsilon \to 0$ as the training loss $\mathcal{L}_{\mathrm{CFM}}(\theta) \to 0$ (see Appendix A.1, remark 1, for a discussion of this limit). This bound holds regardless of the training dataset. As discussed in Section 3.3, however, the results below are most directly relevant when training data follow $P_{\mathcal{H}_0}$; we then simply write $\mathbf{x}$ for $\mathbf{x}^{\mathcal{H}_0} \sim P_{\mathcal{H}_0}$.

4.1 Learned Detection Statistic Bounds

We first ask how far the learned detection statistic, introduced in Section 3.2, can drift from drift from the ideal one, as a direct consequence of equation 14.

Proposition 1. *Let $S(\mathbf{x}) = \|T^{-1}(\mathbf{x})\|^2$ and $S_\theta(\mathbf{x}) = \|T_\theta^{-1}(\mathbf{x})\|^2$. Suppose the bound equation 14 holds on a set $\mathcal{X}$ with $\mathbb{P}(\mathbf{x} \in \mathcal{X}) \geq 1-\eta$, and let $C_S := \sup_{\mathbf{x}} |S_\theta(\mathbf{x}) - S(\mathbf{x})|$. Then,*

$$\mathbb{E}\left[|S_\theta(\mathbf{x}) - S(\mathbf{x})|\right] \leq \varepsilon\,(2\delta + \varepsilon) + \eta\, C_S \qquad \textit{and} \qquad \mathbb{P}\left(|S_\theta(\mathbf{x}) - S(\mathbf{x})| > \sqrt{\varepsilon}\right) \leq \eta + (\varepsilon + 2\delta)\sqrt{\varepsilon}, \tag{15}$$

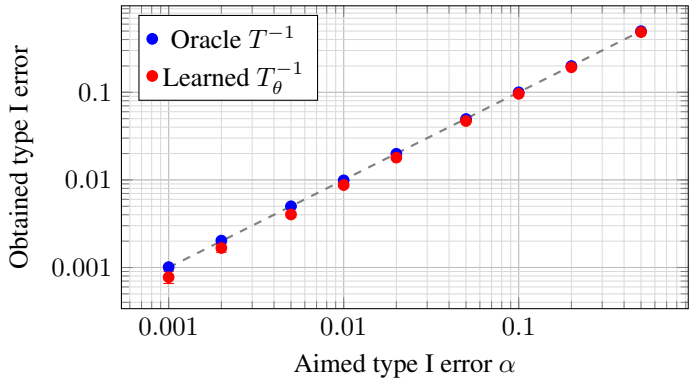

Figure 4: Realized vs. aimed type I error for a K-distribution, comparing the oracle transport T^{-1} to the learned flow T_θ^{-1}, both using the analytical χ^2_d threshold. Dashed line: $y = x$.

where $\delta := \mathbb{E}\big[\|T^{-1}(\mathbf{x})\|\big] = \dfrac{\sqrt{2}\,\Gamma\big(\frac{d+1}{2}\big)}{\Gamma\big(\frac{d}{2}\big)}$, *the mean of the* χ_d *distribution.*

The proof is derived in Appendix A.2. The first bound shows that S_θ stays within $O(\varepsilon)$ of S on average, up to a residual term ηC_S accounting for the (probability-η) event where the transport bound equation 14 does not hold. The second bound is obtained from the first via Markov's inequality; it similarly caps the probability that two statistics differ by more than $\sqrt{\varepsilon}$, up to the same residual η. Both bounds vanish as $\varepsilon \to 0$ at any fixed η; for fixed ε, they reduce to ηC_S and η respectively.

We then derive the type I error of the test $\{S_\theta(\mathbf{x}) > \lambda\}$ actually implemented, compared to the nominal level obtained with the ideal statistic S.

Proposition 2. *Let* $S(\mathbf{x}) = \|T^{-1}(\mathbf{x})\|^2$ *and* $S_\theta(\mathbf{x}) = \|T_\theta^{-1}(\mathbf{x})\|^2$. *Suppose the bound equation 14 holds on a set* $\mathcal{X}$ *with* $\mathbb{P}(\mathbf{x} \in \mathcal{X}) \geq 1 - \eta$. *Then, for any* $\lambda > \varepsilon^2$ *and* $d > 2$,

$$\big|\,\mathbb{P}\big(S_\theta(\mathbf{x}) > \lambda\big) - \mathbb{P}\big(S(\mathbf{x}) > \lambda\big)\,\big| \;\leq\; \eta + L_d\left(2\varepsilon\sqrt{\lambda} + \varepsilon^2\right) \;=:\; \rho(\varepsilon, \lambda), \tag{16}$$

where L_d *is the Lipschitz constant of the CDF* $F_{\chi^2_d}$ *of* χ^2_d, *and* $\rho(\varepsilon, \lambda) \to \eta$ *as* $\varepsilon \to 0$ *for fixed* λ.

The proof is derived in Appendix A.3. Figure 4 validates this result numerically on data following a K-distribution: using the threshold λ obtained analytically from the χ^2_d quantile, with no calibration data required, the realized false alarm rate closely matches the nominal level α across three orders of magnitude, even when T^{-1} is replaced by the learned flow T_θ^{-1}.

Evaluating the bounds of Propositions 1 and 2 in practice requires an explicit form of T^{-1}, so that ε can be estimated directly rather than only through the training loss. We now present two cases where T^{-1} admits a closed form.

Example 1 : Gaussian distribution. If $P_{\mathcal{H}_0} = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the optimal transport map (in the sense of Brenier (1991)) is affine:

$$T^{-1}(\mathbf{x}) = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu}). \tag{17}$$

Example 2: Spherical distribution. Spherically symmetric distributions are widely used to model non-Gaussian noise across signal processing and statistics. If $P_{\mathcal{H}_0}$ is spherically symmetric, i.e. $\mathbf{x} = r\boldsymbol{\omega}$ with $r = \|\mathbf{x}\|$ and $\boldsymbol{\omega}$ uniform on the sphere independent of r, the optimal transport map is given in closed form by

$$T^{-1}(\mathbf{x}) = \phi(r)\boldsymbol{\omega}, \tag{18}$$

with $\phi = F^{-1} \circ G$ the composition of the CDFs of $\|\mathbf{z}\|$ (for $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$) and r; see Appendix B.1 for the derivation. This is the case of the K-distributions, widely used in this work both to validate our theoretical results numerically (Figures 2 and 4) and to model radar clutter (Section 6.2), which are a particular case of spherically symmetric distributions.

Both expressions above are, in fact, exactly the optimal transport map. Appendix B.2 shows that flow matching exactly recovers this map in both cases.

5 DETECTING A SIGNAL OF INTEREST

Detecting a signal in noise is a classical and practically important hypothesis testing problem, arising in applications such as signal processing, radar, and communications. A particularly relevant setting is the discrimination between a noise-only hypothesis and a signal-plus-noise hypothesis. While, as discussed in Section 3.2, the type II error generally depends on $P_{\mathcal{H}_1}$ and admits no closed-form expression, this setting yields a tractable special case when $\mathcal{H}_1$ is obtained by adding a deterministic signal of interest to $\mathcal{H}_0$. This structure enables an explicit characterization of the test, including a closed-form expression for the type II error in the Gaussian case. Concretely, we consider:

$$\mathcal{H}_0 : \mathbf{x}^{\mathcal{H}_0} \sim P_{\mathcal{H}_0}, \tag{19}$$

$$\mathcal{H}_1 : \mathbf{x}^{\mathcal{H}_1} = \mathbf{p} + \mathbf{x}^{\mathcal{H}_0}, \;\; \mathbf{x}^{\mathcal{H}_0} \sim P_{\mathcal{H}_0}, \tag{20}$$

with $\mathbf{p}$, the signal of interest, a deterministic element of $\mathbb{R}^d$. In Appendix C, we show that we can describe $S(\mathbf{x}^{\mathcal{H}_1})$ as function if the reference distribution by approximating T^{-1} via a Taylor expansion (assuming T^{-1} is differentiable with an invertible Jacobian), giving the detection statistic

$$S(\mathbf{x}^{\mathcal{H}_1}) = \|T^{-1}(\mathbf{x}^{\mathcal{H}_1})\|^2 \approx \left\|\mathbf{z}^{\mathcal{H}_0} + J_{T^{-1}}(\mathbf{x}^{\mathcal{H}_0})\,\mathbf{p}\right\|^2, \qquad \mathbf{z}^{\mathcal{H}_0} := T^{-1}(\mathbf{x}^{\mathcal{H}_0}).$$

Even here, $S(\mathbf{x}^{\mathcal{H}_1})$ has no closed form: the Jacobian is itself a function of $\mathbf{z}^{\mathcal{H}_0}$, so the statistic is a nonlinear transformation of $\mathbf{z}^{\mathcal{H}_0}$ whose law depends on the geometry of T (see Appendix C). We therefore study its expectation over $\mathbf{x}^{\mathcal{H}_0}$, which, as shown in Appendix C, decomposes as

$$\mathbb{E}\left[\|T^{-1}(\mathbf{x}^{\mathcal{H}_1})\|^2\right] \approx d + C + M, \tag{21}$$

where $\mathbb{E}\left[\|T^{-1}(\mathbf{x}^{\mathcal{H}_0})\|^2\right] = d$ since $S(\mathbf{x}^{\mathcal{H}_0}) \sim \chi^2_d$, and where the cross term $C = 2\,\mathbb{E}\left[T^{-1}(\mathbf{x}^{\mathcal{H}_0})^\top J_{T^{-1}}(\mathbf{x}^{\mathcal{H}_0})\right]\mathbf{p}$ and the quadratic term $M = \mathbf{p}^\top \mathbb{E}\left[J_{T^{-1}}(\mathbf{x}^{\mathcal{H}_0})^\top J_{T^{-1}}(\mathbf{x}^{\mathcal{H}_0})\right]\mathbf{p}$, are respectively linear and quadratic in $\mathbf{p}$. Since the statistic is compared to a threshold, we want this expectation to exceed d, so that the signal of interest is, on average, pushed towards the decision region (this is a first-order indication rather than a guarantee on P_d, which also depends on the dispersion of S under $\mathcal{H}_1$). As $M \geq 0$, the only term of concern is the cross term C, whose sign is unknown in general.

Example 1: Gaussian distribution. If $\mathbf{x}^{\mathcal{H}_0} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the transport map is affine, $T^{-1}(\mathbf{x}) = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$, so its Jacobian is constant and the Taylor expansion above is exact. The signal is mapped to a deterministic shift of the standard Gaussian, $T^{-1}(\mathbf{x}^{\mathcal{H}_1}) = \mathbf{z}^{\mathcal{H}_0} + \boldsymbol{\Sigma}^{-1/2}\mathbf{p}$ with $\mathbf{z}^{\mathcal{H}_0} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, and the detection statistic follows a noncentral chi-squared distribution,

$$S(\mathbf{x}^{\mathcal{H}_1}) \sim \chi'^2_d(M), \qquad M = \mathbf{p}^\top \boldsymbol{\Sigma}^{-1}\mathbf{p}.$$

Here $C = 0$ and equation 21 is exact, giving $\mathbb{E}\left[\|T^{-1}(\mathbf{x}^{\mathcal{H}_1})\|^2\right] = d + M$. This is the one case where the power is fully characterized: the type II error is $F_{\chi'^2_d(M)}(\lambda)$, so it decreases monotonically with M, i.e. with signal-to-noise ratio (as defined later in section 6.2). Figure 6 confirms this empirically.

Example 2: Spherical distribution. With the closed form of equation 18, $T^{-1}(\mathbf{x}) = \phi(r)\boldsymbol{\omega}$, the Jacobian is no longer constant and S has no closed form. We show in Appendix C.1 that the cross term nonetheless vanishes, $C = 0$, so that the mean of the statistic increases by $M \geq 0$ under $\mathcal{H}_1$, as desired to first order. Numerical validation is given in Figure 7.

A concrete application of this model is radar target detection, whose experiments are presented in Section 6.2.

6 EXPERIMENTS

6.1 VISION-BASED OOD DETECTION

We evaluate our approach on image data: distinguishing images of common objects such as dogs, trucks, and horses (CIFAR-10) from images of house numbers (SVHN). We consider both CIFAR-10 and SVHN as ID datasets, and evaluate out-of-distribution (OOD) detection against CIFAR-10, SVHN, CelebA, and CIFAR-100. For each ID dataset, the transport is trained exclusively on ID samples and has no access to OOD data during training. At test time, an input is mapped to the reference space using the learned transport, and detection is performed as discussed in Section 3.3.

We compare our approach with a broad set of OOD detection methods trained on the ID data (or $\mathcal{H}_0$) used as the reference for detection: energy-based generative models (IGEBM Du & Mordatch (2019) and VAEBM Xiao et al. (2021)), likelihood-based methods (WAIC Choi et al. (2018), likelihood-ratio scoring (LR) Ren et al. (2019), and negative log-likelihood (NLL)), classifier-based methods (ReAct Sun et al. (2021), ASH Djurisic et al. (2022)), and diffusion-based (DDPM-OOD Graham et al. (2023b)). For each baseline, we use the authors' released pretrained models when available; otherwise, we train the method using the settings reported in the original work on the same ID/OOD dataset pairs. Further implementation details are provided in Appendix D.3.

Table 1 reports AUROC for each ID/OOD pair, with higher values indicating better OOD discrimination. Among methods trained solely on $\mathcal{H}_0$, our algorithm achieves the best performance. More recent diffusion-based methods that rely on a cross-backbone pretrained on a broad image corpus, such as SCOPED (Barkley et al., 2026) or GEPC (Rouzoumka et al., 2026), still achieve slightly higher AUROC (see Table 9 for a separate comparison). However, such methods presuppose the existence of a suitable pretrained backbone, an assumption that does not hold for many application domains, particularly when data differ substantially from natural images. The following two sections illustrate this on two such domains: biological classification and radar target detection.

Table 1: OOD detection performance comparison in terms of AUROC (higher is better). All methods are trained or pretrained on ID data and evaluated on OOD data. "-" indicates that the corresponding result is unavailable. Best value per column in **bold**.

	CIFAR-10 (ID)			**SVHN (ID)**			**Avg**
Method	SVHN	CelebA	C100	C10	CelebA	C100	
IGEBM Du & Mordatch (2019)	0.634	0.700	0.500	-	-	-	-
VAEBM Xiao et al. (2021)	**0.830**	0.770	**0.620**	-	-	-	-
WAIC Choi et al. (2018)	0.143	**0.928**	0.532	0.802	0.991	0.831	0.704
LR Ren et al. (2019)	0.064	0.914	0.520	0.819	0.912	0.779	0.668
NLL	0.091	0.574	0.521	**0.990**	**0.999**	**0.992**	0.694
DDPM-OOD Graham et al. (2023b)	0.390	0.659	0.536	0.951	0.986	0.945	0.744
ReAct Sun et al. (2021)	0.334	0.343	0.475	0.932	0.948	0.930	0.660
ASH Djurisic et al. (2022)	0.528	0.345	0.493	0.320	0.804	0.338	0.471
Ours	0.503	0.695	0.511	0.897	0.993	0.963	**0.760**

6.2 Radar Target Detection

Radar target detection is an instance of anomaly detection (Maio & Greco, 2015). Clutter — for instance, an ocean surface — is modeled by the distribution under $\mathcal{H}_0$, and the goal is to detect a target (e.g., a boat) characterized by a steering vector $\mathbf{p}$, so that $\mathcal{H}_1$ corresponds to the presence of this signal on top of the clutter, exactly as in the additive signal model of Section 5. Two classical clutter models are considered: the Gaussian and the K-distribution, which better captures texture variations across the scene. The difficulty of the detection problem is governed by the target amplitude A, which determines the signal-to-noise ratio $\mathrm{SNR} = |A|^2\, \mathbf{p}^H \hat{\mathbf{\Sigma}}^{-1} \mathbf{p}$. Detection is traditionally performed via the generalized likelihood ratio test (GLRT) between $\mathcal{H}_0$ and $\mathcal{H}_1$ (Kelly, 1986), and methods from the machine learning literature, such as SVDD, have also been tested for radar detection (Pinsolle et al., 2026). All setup parameters and benchmark algorithms are detailed in Appendix D.5.

As mentioned in Section 3.3, a detection is counted when the detection statistic exceeds a threshold λ, fixed to achieve a desired probability of false alarm P_{fa} (CFAR property). In classical detection, a closed-form relation between λ and P_{fa} is only tractable when the clutter is Gaussian (or elliptical, under mild conditions). Our method instead admits a closed-form expression for P_{fa} that holds regardless of the clutter distribution, as derived in Section 3.2. In practice, thresholds are calibrated empirically by taking the corresponding quantile of the score distribution under $\mathcal{H}_0$ — as is standard even for classical CFAR detectors — and we verified in Figure 4 that both estimates closely agree under K-distributed clutter.

Figure 5 compares detection probability across SNR values for the detectors introduced above. Under Gaussian clutter, our detector performs slightly better than the adaptive matched filter (AMF) and SVDD, approaching the oracle bounds Matched Filter (MF). Under K-distributed clutter, our detector remains competitive with ANMF-Tyler – although the latter exploits knowledge of $\mathbf{p}$ – and clearly outperforms both AMF and Support Vector Data Description (SVDD).

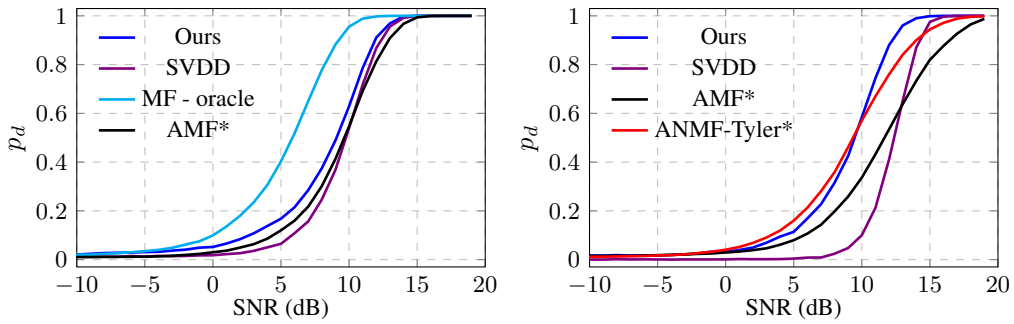

Figure 5: Performance comparison in terms of p_d, for a type I error $\alpha(\lambda) = P_{fa}(\lambda) = 10^{-2}$. Left: Gaussian clutter. Right: K-distributed clutter. See all details in appendix D.5.*Assumes $\mathbf{p}$ known.

6.3 Bioinformatics

Imbalanced Data Classification: We evaluate the proposed transport-based approach against existing imbalance-learning methods, including RUSBoost Seiffert et al. (2010), SMOTEBoost Chawla et al. (2003), UnderBagging Barandela et al. (2003), HUE Ng et al. (2022), BESS Ning et al. (2022), SPE Liu et al. (2020), EasyEnsemble Liu et al. (2008), SMOTEBagging Wang & Yao (2009), MDSampler Ning et al. (2025), and ImbXGBoost Wang et al. (2020) on Protein Homology dataset Caruana et al. (2004) from the 2004 KDD Cup (BIO) using metrics, F1 score, AUROC, and AUPRC. We present this comparison in Table 2. We find that the proposed approach achieves an F1 score of 0.88 and an AUROC of 0.99. In terms of F1 score, it outperforms all evaluated methods except ImbXGBoost (0.89), which achieves a marginally higher score by 0.01. Its AUROC is comparable to SPE and ImbXGBoost, while it outperforms all in terms of AUPRC. Overall, these results demonstrate that the proposed transport-based approach achieves strong, competitive performance on imbalanced classification tasks.

Table 2: Imbalanced classification results. Higher values indicate better performance.

Method	RUSBoost	SMOTEBoost	UnderBagging	HUE	BESS	SPE	EasyEnsemble	SMOTEBagging	MDSampler	ImbXGBoost	Ours
F1	0.21	0.22	0.33	0.34	0.34	0.52	0.55	0.86	0.87	**0.89**	0.88
AUROC	0.89	0.97	0.97	0.95	0.96	**0.99**	0.98	0.96	0.98	**0.99**	**0.99**
AUPRC	0.18	0.75	0.70	0.92	0.24	0.81	0.84	0.86	0.92	0.91	**0.93**

For completeness, we report in Appendix D.10 an additional experiment on attribute matching, where the transport-based representation is used as a two-sample test to identify which pairs of empirically observed attributes correspond to the same underlying quantity.

7 Conclusion, Limitations, and Future Directions

This work proposes a detector for anomaly detection that transports in-distribution data toward a standard Gaussian via flow matching. The resulting squared norm follows a chi-squared distribution under the reference hypothesis, yielding an explicit expression for the type I error, $1 - F_{\chi^2_d}(\lambda)$. We further derive bounds on the transport estimation error and on the resulting type I error, both depending on the training loss of the velocity field, and validate these results numerically by comparing the estimated flow to the exact transport on distributions for which the latter is known in closed form (Gaussian and spherically symmetric references). We finally evaluate the detector across three domains: out-of-distribution detection in images, radar target detection, and biological imbalanced data classification. The detector is competitive in all three settings, matching or outperforming most evaluated baselines.

A promising direction is to further exploit the explicit type I error expression in settings where it cannot be validated empirically against ground-truth labels, such as safety-critical deployments where out-of-distribution data are never observed during calibration.

AI USE STATEMENT

In this work, we used generative AI tools to assist with writing, structuring the paper, correcting typos, and generating plotting code. We have not used generative AI tools for research ideation or the derivation of theoretical results. We have reviewed all AI-assisted work : the writing of the paper, the code used for experiments. We take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide the source code and supplementary materials to facilitate the reproducibility of our work. The proposed method, proofs, and its implementation details are provided in the main paper and Appendix. The supplementary material includes details of the datasets, model configuration, hyperparameters, training settings, and evaluation protocols. Additional experimental results, along with ablation studies, are also provided in the supplementary material to support our work.

REFERENCES

Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Computer Vision – ACCV 2018*, pp. 622–637. Springer, 2019. doi: 10.1007/978-3-030-20893-6_39.

Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2(1):1–18, 2015.

Ricardo Barandela, Rosa Maria Valdovinos, and José Salvador Sánchez. New applications of ensembles of classifiers. *Pattern Analysis & Applications*, 6(3):245–256, 2003.

Brett Barkley, Preston Culbertson, and David Fridovich-Keil. Scoped: Score-curvature out-of-distribution proximity evaluator for diffusion. 2026.

Christopher M Bishop. Novelty detection and neural network validation. *IEE Proceedings-Vision, Image and Signal Processing*, 141(4):217–222, 1994.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

Rich Caruana, Thorsten Joachims, and Lars Backstrom. Kdd-cup 2004: results and analysis. *ACM SIGKDD Explorations Newsletter*, 6(2):95–108, 2004. URL `https://www.kdd.org/kdd-cup/view/kdd-cup-2004/Data`.

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):1–58, 2009.

Nitesh V Chawla, Aleksandar Lazarevic, Lawrence O Hall, and Kevin W Bowyer. Smoteboost: Improving prediction of the minority class in boosting. In *European conference on principles of data mining and knowledge discovery*, pp. 107–119. Springer, 2003.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge–Kantorovich depth, quantiles, ranks, and signs. *Annals of Statistics*, 45(1):223–256, 2017. doi: 10.1214/16-AOS1450.

Hyunsun Choi, Eric Jang, and Alexander A Alemi. Waic, but why? generative ensembles for robust anomaly detection. *arXiv preprint arXiv:1810.01392*, 2018.

Andrija Djurisic, Nebojsa Bozanic, Arjun Ashok, and Rosanne Liu. Extremely simple activation shaping for out-of-distribution detection. 35, 2022.

Yilun Du and Igor Mordatch. Implicit generation and modeling with energy based models. *Advances in neural information processing systems*, 32, 2019.

Mark S. Graham, Walter H. L. Pinaya, Petru-Daniel Tudosiu, Parashkev Nachev, Sebastien Ourselin, and Jorge Cardoso. Denoising diffusion models for out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 2948–2957, 2023a.

Mark S Graham, Walter HL Pinaya, Petru-Daniel Tudosiu, Parashkev Nachev, Sebastien Ourselin, and Jorge Cardoso. Denoising diffusion models for out-of-distribution detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2948–2957, 2023b.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The journal of machine learning research*, 13(1):723–773, 2012.

T. H. Gronwall. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Annals of Mathematics*, 20(4):292–296, 1919.

Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations*, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

E. J. Kelly. An adaptive detection algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 22(1):115–127, 1986.

Zhong Li, Qi Huang, Yuxuan Zhu, Lincen Yang, Mohammad Mohammadi Amiri, Niki van Stein, and Matthijs van Leeuwen. Scalable, explainable and provably robust anomaly detection with one-step flow matching. In *Advances in Neural Information Processing Systems*, 2025.

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.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023a.

Xu-Ying Liu, Jianxin Wu, and Zhi-Hua Zhou. Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(2):539–550, 2008.

Zhenzhen Liu, Jin Peng Zhou, Yufan Wang, and Kilian Q. Weinberger. Unsupervised out-of-distribution detection with diffusion inpainting. In *International Conference on Machine Learning*, pp. 22528–22538, 2023b.

Zhining Liu, Wei Cao, Zhifeng Gao, Jiang Bian, Hechang Chen, Yi Chang, and Tie-Yan Liu. Self-paced ensemble for highly imbalanced massive data classification. In *2020 IEEE 36th international conference on data engineering (ICDE)*, pp. 841–852. IEEE, 2020.

Antonio De Maio and Maria Sabrina Greco (eds.). *Modern Radar Detection Theory*. The Institution of Engineering and Technology, 2015. doi: 10.1049/SBRA509E. URL `https://digital-library.theiet.org/doi/abs/10.1049/SBRA509E`.

Eric Nalisnick, Akihiro Matsukawa, Yee Whye Teh, Dilan Gorur, and Balaji Lakshminarayanan. Do deep generative models know what they don't know? In *International Conference on Learning Representations*, 2019.

Wing WY Ng, Shichao Xu, Jianjun Zhang, Xing Tian, Tongwen Rong, and Sam Kwong. Hashing-based undersampling ensemble for imbalanced pattern classification problems. *IEEE Transactions on Cybernetics*, 52(2):1269–1279, 2022.

Zhihan Ning, Ziqing Ye, Zhixing Jiang, and David Zhang. Bess: Balanced evolutionary semi-stacking for disease detection using partially labeled imbalanced data. *Information Sciences*, 594: 233–248, 2022.

Zhihan Ning, Chaoxun Guo, and David Zhang. Meta-distribution-based ensemble sampler for imbalanced semi-supervised learning. *Pattern Recognition*, 164:111552, 2025.

Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Computing Surveys*, 54(2):1–38, 2021.

Long HC Pham. Decomposing out-of-distribution error in conditional flow matching via wasserstein geometry. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=orzcDIBxa0.

Jean Pinsolle, Yadang Alexis Rouzoumka, Chengfang Ren, Christèle Morisseau, and Jean-Philippe Ovarlez. Support vector data description for radar target detection. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2026. doi: 10.1109/ICASSP55912.2026.11464086.

Jie Ren, Peter J Liu, Emily Fertig, Jasper Snoek, Ryan Poplin, Mark Depristo, Joshua Dillon, and Balaji Lakshminarayanan. Likelihood ratios for out-of-distribution detection. *Advances in neural information processing systems*, 32, 2019.

R. Tyrrell Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, Princeton, NJ, 1997. Originally published in 1970.

Yadang Alexis Rouzoumka, Jean Pinsolle, Eugénie Terreaux, Christèle Morisseau, Jean-Philippe Ovarlez, and Chengfang Ren. Gepc: Group-equivariant posterior consistency for out-of-distribution detection in diffusion models. In *International Conference on Machine Learning*, 2026.

Lukas Ruff, Robert Vandermeulen, Nico Görnitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International Conference on Machine Learning*, pp. 4393–4402, 2018.

Chris Seiffert, Taghi M Khoshgoftaar, Jason Van Hulse, and Amri Napolitano. Rusboost: A hybrid approach to alleviating class imbalance. *IEEE transactions on systems, man, and cybernetics-part A: systems and humans*, 40(1):185–197, 2010.

Shirin Shoushtari, Yi Wang, Xiao Shi, M. Salman Asif, and Ulugbek S. Kamilov. Eigenscore: Ood detection using posterior covariance in diffusion models. In *International Conference on Learning Representations*, 2026.

Luzhe Sun, Jingtian Ji, Haoran Chen, Jiawei Zhou, and Matthew R. Walter. Flow-based policy adaptation without policy updates, 2026. URL https://arxiv.org/abs/2606.06461.

Yiyou Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. *Advances in neural information processing systems*, 34:144–157, 2021.

David MJ Tax and Robert PW Duin. Support vector data description. *Machine Learning*, 54:45–66, 2004.

David E. Tyler. A distribution-free m-estimator of multivariate scatter. *The Annals of Statistics*, 15 (1):234–251, 1987.

Chen Wang, Chengyuan Deng, and Suzhen Wang. Imbalance-xgboost: leveraging weighted and focal losses for binary label-imbalanced classification with xgboost. *Pattern recognition letters*, 136:190–197, 2020.

Shuo Wang and Xin Yao. Diversity analysis on imbalanced data sets by using ensemble models. In *2009 IEEE symposium on computational intelligence and data mining*, pp. 324–331. IEEE, 2009.

Zhisheng Xiao, Karsten Kreis, Jan Kautz, and Arash Vahdat. Vaebm: A symbiosis between variational autoencoders and energy-based models. In *International Conference on Learning Representations*, 2021.

Chen Xu, Tony Khuong Nguyen, Emma Dixon, Christopher Rodriguez, Patrick Miller, Robert Lee, Paarth Shah, Rares Ambrus, Haruki Nishimura, and Masha Itkina. Can we detect failures without failure data? uncertainty-aware runtime failure detection for imitation learning policies. In *Robotics: Science and Systems (RSS)*, 2025.

A BOUNDS COMPUTATION DETAILS

A.1 BOUND OF THE ESTIMATION TRANSPORT MAP WITH THE LOSS

We show here how the bound equation 14 can be justified from the CFM training loss.

We assume that $\mathbf{v}_\theta(\cdot, t)$ and $\mathbf{u}_t(\cdot)$ are both Lipschitz in space for each t, and denote by L_t the Lipschitz constant of $\mathbf{v}_\theta(\cdot, t)$.

Fix $\mathbf{x} \in \mathbb{R}^d$, we consider the backwards trajectories from $\mathbf{x}$ at $t = 1$: the exact one, $\boldsymbol{\psi}_t(\mathbf{x})$, driven by $-\mathbf{u}_t$, and the learned one, $\boldsymbol{\psi}_t^\theta(\mathbf{x})$, driven by $-\mathbf{v}_\theta$. At $t = 0$ they give $T^{-1}(\mathbf{x})$ and $T_\theta^{-1}(\mathbf{x})$ respectively. We want to bound how far apart they end up, in terms of how well $\mathbf{v}_\theta$ approximates $\mathbf{u}_t$.

Write the difference between the two trajectories as $\delta_t := \boldsymbol{\psi}_t^\theta(\mathbf{x}) - \boldsymbol{\psi}_t(\mathbf{x})$, so $\delta_1 = \mathbf{0}$. Differentiating and adding and subtracting $\mathbf{v}_\theta$ evaluated on the exact trajectory,

$$-\dot{\delta}_t = \mathbf{v}_\theta(\boldsymbol{\psi}_t^\theta(\mathbf{x}), t) - \mathbf{u}_t(\boldsymbol{\psi}_t(\mathbf{x})) = \underbrace{\mathbf{v}_\theta(\boldsymbol{\psi}_t^\theta(\mathbf{x}), t) - \mathbf{v}_\theta(\boldsymbol{\psi}_t(\mathbf{x}), t)}_{\text{controlled by the Lipschitz constant}} + \underbrace{\mathbf{v}_\theta(\boldsymbol{\psi}_t(\mathbf{x}), t) - \mathbf{u}_t(\boldsymbol{\psi}_t(\mathbf{x}))}_{\text{the model's error}}. \tag{22}$$

By hypothesis, the first term is bounded by $L_t\|\delta_t\|$. The second term is the velocity error, evaluated along the exact trajectory. Taking norms,

$$\|\dot{\delta}_t\| \leq L_t\|\delta_t\| + \big\|\mathbf{v}_\theta(\boldsymbol{\psi}_t(\mathbf{x}), t) - \mathbf{u}_t(\boldsymbol{\psi}_t(\mathbf{x}))\big\|. \tag{23}$$

Since $\delta_1 = \mathbf{0}$, Grönwall's lemma Gronwall (1919) applied to this differential inequality gives

$$\big\|T_\theta^{-1}(\mathbf{x}) - T^{-1}(\mathbf{x})\big\| \leq K \int_0^1 \big\|\mathbf{v}_\theta(\boldsymbol{\psi}_t(\mathbf{x}), t) - \mathbf{u}_t(\boldsymbol{\psi}_t(\mathbf{x}))\big\|\, dt. \tag{24}$$

where $K := \exp(\int_0^1 L_t\, dt)$. This inequality holds for every fixed $\mathbf{x}$; To connect this to the training loss, we use Cauchy-Schwarz: the integral in equation 24 is at most $\sqrt{e(\mathbf{x})}$, where

$$e(\mathbf{x}) := \int_0^1 \big\|\mathbf{v}_\theta(\boldsymbol{\psi}_t(\mathbf{x}), t) - \mathbf{u}_t(\boldsymbol{\psi}_t(\mathbf{x}))\big\|^2\, dt \tag{25}$$

is the accumulated squared error along the trajectory through $\mathbf{x}$. Now, if we let $\mathbf{x}$ be random, drawn from $P_{\mathcal{H}_0}$, the trajectory $\boldsymbol{\psi}_t(\mathbf{x})$ has the same law as the flow matching which interpolates $\mathbf{x}_t$ at each time t. Averaging $e(\mathbf{x})$ over $\mathbf{x} \sim P_{\mathcal{H}_0}$ therefore gives exactly the marginal flow matching loss,

$$\mathbb{E}_{\mathbf{x}\sim P_{\mathcal{H}_0}}[e(\mathbf{x})] = \int_0^1 \mathbb{E}\big[\|\mathbf{v}_\theta(\mathbf{x}_t, t) - \mathbf{u}_t(\mathbf{x}_t)\|^2\big]\, dt = \mathcal{L}_{\mathrm{FM}}(\theta). \tag{26}$$

Since $\mathcal{L}_{\mathrm{CFM}}(\theta) - \mathcal{L}_{\mathrm{FM}}(\theta)$ is a conditional variance, it is always nonnegative. This means $\mathbb{E}_{\mathbf{x}\sim P_{\mathcal{H}_0}}[e(\mathbf{x})] \leq \mathcal{L}_{\mathrm{CFM}}(\theta)$: the loss we actually minimize is a valid upper bound on the average pathwise error.

We now just need to turn this average control into control on most individual trajectories. Since $e(\mathbf{x}) \geq 0$, Markov's inequality gives, for any $\eta \in (0, 1)$,

$$\mathbb{P}_{\mathbf{x}\sim P_{\mathcal{H}_0}}\Big(e(\mathbf{x}) > \mathcal{L}_{\mathrm{CFM}}(\theta)/\eta\Big) \leq \eta. \tag{27}$$

Putting everything together: on the set $\mathcal{X} := \{\mathbf{x} : e(\mathbf{x}) \leq \mathcal{L}_{\mathrm{CFM}}(\theta)/\eta\}$, which has probability at least $1 - \eta$ under $P_{\mathcal{H}_0}$, the bound equation 14 holds with

$$\boxed{\varepsilon = K\sqrt{\mathcal{L}_{\mathrm{CFM}}(\theta)/\eta}\,.} \tag{28}$$

Remark 1. As explained in the proof , $\mathcal{L}_{\mathrm{CFM}}(\theta) - \mathcal{L}_{\mathrm{FM}}(\theta)$ is a conditional variance,

$$D := \mathcal{L}_{\mathrm{CFM}}(\theta) - \mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{t,\mathbf{x}_t}\big[\mathrm{Var}(\mathbf{x}_1 - \mathbf{x}_0 \mid \mathbf{x}_t)\big], \tag{29}$$

which does not depend on θ: it reflects how many distinct pairs $(\mathbf{x}_0, \mathbf{x}_1)$, drawn independently from $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ and $P_{\mathcal{H}_0}$, can produce the same interpolated point $\mathbf{x}_t$ at a given t. With an independent

coupling, this ambiguity is generally unavoidable, so $D > 0$: even the best possible model, $\theta^\star = \arg\min_\theta \mathcal{L}_{\mathrm{CFM}}(\theta)$, satisfies $\mathcal{L}_{\mathrm{CFM}}(\theta^\star) = D > 0$, not 0. Consequently, the bound in equation 14 does not vanish as training proceeds: ε decreases toward an irreducible floor $\varepsilon_{\min} = K\sqrt{D/\eta} > 0$, rather than to 0. The reflow procedure of Appendix D.1 directly addresses this gap. At round k, the coupling $(\mathbf{x}_0, \hat{\mathbf{x}}_1)$ becomes *deterministic*: $\hat{\mathbf{x}}_1$ is a function of $\mathbf{x}_0$ through the previous flow $\psi_1^{(k)}$, rather than an independent draw. Under a deterministic coupling, the ambiguity in equation 29 vanishes almost everywhere, so the corresponding gap $D^{(k)}$ is expected to decreases with k.

Remark 2. The bound above concerns the exact backward flow T_θ^{-1}, obtained by solving the ODE $\dot{\psi}_t^\theta = -\mathbf{v}_\theta(\psi_t^\theta, t)$ exactly. In practice, this ODE is integrated numerically, e.g. with an explicit Euler scheme with N steps ($h = 1/N$), yielding an approximation $\widehat{T}_{\theta,h}^{-1}$ rather than T_θ^{-1} itself. Standard results on the convergence of Euler's method for Lipschitz vector fields give

$$\big\|\widehat{T}_{\theta,h}^{-1}(\mathbf{x}) - T_\theta^{-1}(\mathbf{x})\big\| = O(h) = O(1/N), \tag{30}$$

with a constant depending on L_t and on a bound on the time-derivative of $\mathbf{v}_\theta$ along the trajectory. Combined with the triangle inequality, this yields

$$\big\|\widehat{T}_{\theta,h}^{-1}(\mathbf{x}) - T^{-1}(\mathbf{x})\big\| \le \varepsilon + O(1/N), \tag{31}$$

so that all bounds in Propositions 1 and 2 carry over to the statistic actually computed in practice, $\widehat{S}_{\theta,h}(\mathbf{x}) := \|\widehat{T}_{\theta,h}^{-1}(\mathbf{x})\|^2$, up to this additional, vanishing term.

A.2 Detection score bound

Here we derive the proof of Proposition 1.

Proof. On $\mathcal{X}$, by equation 14, for all $\mathbf{x} \in \mathcal{X}$,

$$\begin{aligned} |S_\theta(\mathbf{x}) - S(\mathbf{x})| &= \big|\|T_\theta^{-1}(\mathbf{x})\|^2 - \|T^{-1}(\mathbf{x})\|^2\big| \\ &= \big(\|T_\theta^{-1}(\mathbf{x})\| + \|T^{-1}(\mathbf{x})\|\big)\,\big|\|T_\theta^{-1}(\mathbf{x})\| - \|T^{-1}(\mathbf{x})\|\big| \\ &\le \big(\|T_\theta^{-1}(\mathbf{x})\| + \|T^{-1}(\mathbf{x})\|\big)\,\big\|T_\theta^{-1}(\mathbf{x}) - T^{-1}(\mathbf{x})\big\| \\ &\le \big(2\|T^{-1}(\mathbf{x})\| + \varepsilon\big)\,\varepsilon, \end{aligned} \tag{32}$$

where the second line uses $\big|\|a\| - \|b\|\big| \le \|a - b\|$ (reverse triangle inequality), and the last line uses equation 14 together with $\|T_\theta^{-1}(\mathbf{x})\| \le \|T^{-1}(\mathbf{x})\| + \varepsilon$.

Taking expectations and splitting over $\mathcal{X}$ and its complement,

$$\mathbb{E}\big[|S_\theta(\mathbf{x}) - S(\mathbf{x})|\big] = \mathbb{E}\big[|S_\theta(\mathbf{x}) - S(\mathbf{x})|\,\mathbf{1}_{\mathcal{X}}(\mathbf{x})\big] + \mathbb{E}\big[|S_\theta(\mathbf{x}) - S(\mathbf{x})|\,\mathbf{1}_{\mathcal{X}^c}(\mathbf{x})\big]. \tag{33}$$

For the first term, using equation 32 and $\mathbb{E}[\|T^{-1}(\mathbf{x})\|] = \delta$,

$$\mathbb{E}\big[|S_\theta(\mathbf{x}) - S(\mathbf{x})|\,\mathbf{1}_{\mathcal{X}}(\mathbf{x})\big] \le \varepsilon(2\delta + \varepsilon). \tag{34}$$

For the second term, we bound it by $\eta \cdot C_S$, where $C_S := \sup_{\mathbf{x}} |S_\theta(\mathbf{x}) - S(\mathbf{x})|$ and $\mathbb{P}(\mathbf{x} \notin \mathcal{X}) \le \eta$; this term vanishes as $\eta \to 0$ for fixed C_S, or can be neglected whenever $\eta \ll \varepsilon$. Altogether,

$$\mathbb{E}\big[|S_\theta(\mathbf{x}) - S(\mathbf{x})|\big] \le \varepsilon(2\delta + \varepsilon) + \eta\, C_S. \tag{35}$$

For the probability bound, Markov's inequality applied to the first term gives, for any $\gamma > 0$,

$$\mathbb{P}\big(|S_\theta(\mathbf{x}) - S(\mathbf{x})| \ge \gamma\big) \le \mathbb{P}(\mathbf{x} \notin \mathcal{X}) + \frac{\mathbb{E}\big[|S_\theta(\mathbf{x}) - S(\mathbf{x})|\,\mathbf{1}_{\mathcal{X}}(\mathbf{x})\big]}{\gamma} \le \eta + \frac{\varepsilon(2\delta + \varepsilon)}{\gamma}. \tag{36}$$

Setting $\gamma = \sqrt{\varepsilon}$ yields

$$\mathbb{P}\big(|S_\theta(\mathbf{x}) - S(\mathbf{x})| > \sqrt{\varepsilon}\big) \le \eta + (\varepsilon + 2\delta)\sqrt{\varepsilon}. \tag{37}$$

□

A.3 Bound on type I error

Proof of Proposition 2.

Proof. For $x \in \mathbb{R}$, write $(x)_+ := \max(x, 0)$. For $d > 2$, the density $f_{\chi^2_d}$ of χ^2_d is bounded, so its CDF $F_{\chi^2_d}$ is L_d-Lipschitz with $L_d := \sup_t f_{\chi^2_d}(t) < \infty$.

By equation 14, for all $\mathbf{x} \in \mathcal{X}$, the reverse triangle inequality gives $\left|\|T_\theta^{-1}(\mathbf{x})\| - \|T^{-1}(\mathbf{x})\|\right| \leq \varepsilon$, i.e. deterministically,

$$\left(\sqrt{S(\mathbf{x})} - \varepsilon\right)_+^2 \;\leq\; S_\theta(\mathbf{x}) \;\leq\; \left(\sqrt{S(\mathbf{x})} + \varepsilon\right)^2. \tag{38}$$

Upper direction. If $S_\theta(\mathbf{x}) > \lambda$ and $\mathbf{x} \in \mathcal{X}$, the right-hand side of equation 38 gives $\sqrt{S(\mathbf{x})} > \sqrt{\lambda} - \varepsilon$, hence $S(\mathbf{x}) > (\sqrt{\lambda} - \varepsilon)_+^2$. Thus $\{S_\theta > \lambda\} \cap \mathcal{X} \subseteq \{S > (\sqrt{\lambda} - \varepsilon)_+^2\}$, and

$$\mathbb{P}(S_\theta > \lambda) - \mathbb{P}(S > \lambda) \leq \eta + F_{\chi^2_d}(\lambda) - F_{\chi^2_d}\big((\sqrt{\lambda} - \varepsilon)_+^2\big) \leq \eta + L_d \cdot 2\varepsilon\sqrt{\lambda}. \tag{39}$$

Lower direction. $\mathbb{P}(S > \lambda) - \mathbb{P}(S_\theta > \lambda) \leq \mathbb{P}(S > \lambda,\, S_\theta \leq \lambda,\, \mathbf{x} \in \mathcal{X}) + \eta$. On this event, the left-hand side of equation 38 gives $(\sqrt{S(\mathbf{x})} - \varepsilon)_+^2 \leq S_\theta(\mathbf{x}) \leq \lambda$, hence $\sqrt{S(\mathbf{x})} \leq \sqrt{\lambda} + \varepsilon$, i.e. $\lambda < S(\mathbf{x}) \leq (\sqrt{\lambda} + \varepsilon)^2$. Thus

$$\mathbb{P}(S > \lambda) - \mathbb{P}(S_\theta > \lambda) \leq \eta + F_{\chi^2_d}\big((\sqrt{\lambda} + \varepsilon)^2\big) - F_{\chi^2_d}(\lambda) \leq \eta + L_d\left(2\varepsilon\sqrt{\lambda} + \varepsilon^2\right). \tag{40}$$

Combining both directions yields equation 16. □

B Transport between a spherical distribution and a Gaussian

B.1 Expression of the transport map

Let $\mathbf{x} \sim P$, P a spherically symmetric distribution assumed absolutely continuous. By definition, we can write $\mathbf{x} = r\,\boldsymbol{\omega}$, with $r = \|\mathbf{x}\|$, $\boldsymbol{\omega} = \mathbf{x}/r$, r and $\boldsymbol{\omega}$ independent, and $\boldsymbol{\omega}$ uniform on the unit sphere. Let $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, let G denote the CDF of r, and F the CDF of $\|\mathbf{z}\|$.

The optimal transport map from P to $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, denoted T^{-1}, is

$$T^{-1}(\mathbf{x}) = \phi(r)\,\boldsymbol{\omega}, \qquad \phi = F^{-1} \circ G. \tag{41}$$

A good reference for this result is Section 2 (particularly 2.4) of Chernozhukov et al. (2017). It states that, since P and $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ are absolutely continuous, there exists a unique optimal transport map, which is the gradient of a convex function.

We define $\Phi(t) = \int_{-\infty}^{t} \phi(s)\, ds$, and consider the map

$$T^{-1}(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}\,\phi(\|\mathbf{x}\|) = \nabla\phi^\star(\mathbf{x}), \qquad \phi^\star(\mathbf{v}) = \Phi(\|\mathbf{v}\|). \tag{42}$$

The function $\phi^\star$ is convex: it is the composition of Φ, which is convex and non-decreasing (as $\phi = F^{-1} \circ G$ is non-decreasing), with the convex function $\|\cdot\|$. The composition rule ensuring convexity in this case is non-trivial and is stated in (Rockafellar, 1997, Th. 5.1).

It remains to show that $T^{-1}_{\#}P = \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. By construction, we can leave $\boldsymbol{\omega}$ unchanged, as it is independent of r (by definition of a spherically symmetric distribution); it is therefore sufficient for $\phi = F^{-1} \circ G$ to push the law of r onto that of $\|\mathbf{z}\|$, i.e. $\phi(r) \sim F$. This follows from the probability integral transform theorem, since G is continuous, as the CDF of $r = \|\mathbf{x}\|$ with $\mathbf{x}$ admitting a density by hypothesis.

In short, the absolute continuity hypothesis on the distributions guarantees the uniqueness of the transport map. We then construct a transport by separating $\mathbf{x} = r\boldsymbol{\omega}$ (definition of a spherically symmetric distribution) and transporting only r onto $\|\mathbf{z}\|$ (the Gaussian being itself spherically symmetric, we can write $\mathbf{z} = \|\mathbf{z}\|\,\boldsymbol{\omega}$ for the same angular part $\boldsymbol{\omega}$) via the function ϕ, given simply by the composition of the two CDFs.

Remark. The result extends directly to a mixture $P = \sum_k w_k P_k$ of spherically symmetric distributions, which is itself spherically symmetric. In this case, the radial CDF G appearing in equation 41 is simply the corresponding mixture of the componentwise radial CDFs, $G = \sum_k w_k G_k$, and the transport map T^{-1} is obtained exactly as above, replacing G by this combined CDF.

B.2 When Does Flow Matching Recover the Optimal Transport Map?

Flow matching, as used in this work, is trained with an independent coupling $(\mathbf{x}_0, \mathbf{x}_1)$ and a linear conditional path. In general, the resulting map does not coincide with the optimal transport map of Brenier (1991): it is well known that the corresponding rectified flow only decreases transport costs but does not, in general, minimize them under this construction (Liu et al., 2023a). We detail here the two special cases used in Section 4 where the two maps do coincide, and explain why this equivalence breaks down for a general elliptical $P_{\mathcal{H}_0}$.

Gaussian case. Let $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ and $\mathbf{x}_1 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be independent, and $\mathbf{x}_t = (1-t)\mathbf{x}_0 + t\mathbf{x}_1$. Since $\mathbf{x}_0$ is isotropic, its law is unchanged by any orthonormal change of basis; we work in the eigenbasis of $\boldsymbol{\Sigma}$, in which $(\mathbf{x}_0, \mathbf{x}_1)$ decouples into d independent one-dimensional problems, one per coordinate.

In each coordinate, $(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_t)$ is jointly Gaussian, so the marginal velocity field $\mathbf{u}_t(\mathbf{x}) = \mathbb{E}[\mathbf{x}_1 - \mathbf{x}_0 \mid \mathbf{x}_t = \mathbf{x}]$ is affine in $\mathbf{x}$. Solving the resulting flow ODE then yields, in closed form, the one-dimensional optimal transport map (the monotone rearrangement) for that coordinate.

Recombining all coordinates gives the affine map $T(\mathbf{z}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}$ — exactly the optimal transport map between $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This holds for any $\boldsymbol{\Sigma}$.

Spherical case. Let $P_{\mathcal{H}_0}$ be spherically symmetric and consider again the independent coupling with linear path. Since $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ is itself spherically symmetric, the joint law of $(\mathbf{x}_0, \mathbf{x}_1)$ is invariant under simultaneous rotation: for any $R \in O(d)$, $(R\mathbf{x}_0, R\mathbf{x}_1) \overset{d}{=} (\mathbf{x}_0, \mathbf{x}_1)$. This forces the marginal velocity field to be equivariant, $\mathbf{u}_t(R\mathbf{x}) = R\mathbf{u}_t(\mathbf{x})$; taking R in the stabilizer of $\mathbf{x}$ shows $\mathbf{u}_t(\mathbf{x})$ must be radial, $\mathbf{u}_t(\mathbf{x}) = a_t(\|\mathbf{x}\|)\,\mathbf{x}/\|\mathbf{x}\|$. The resulting flow map preserves rays and reduces to a one-dimensional monotone problem on the radius, whose unique solution is $\phi = F^{-1} \circ G$, with G the CDF of $\|\mathbf{x}_1\|$ under $P_{\mathcal{H}_0}$ and F the CDF of $\|\mathbf{x}_0\|$, matching the optimal transport map of Appendix B.1.

Remark For a general elliptical distribution $P_{\mathcal{H}_0}$, neither argument applies, and composing the whitening map $\boldsymbol{\Sigma}^{-1/2}$ with the spherical transport map does not yield a valid substitute: a composition of two optimal transport maps is not in general itself a optimal transport map.

C Detection where $\mathcal{H}_1$ is a signal of interest: computations

We detail here the expansion leading to the detection statistic and to the decomposition equation 21 of its expectation. Throughout, we write

$$\mathbf{z}^{\mathcal{H}_0} := T^{-1}(\mathbf{x}^{\mathcal{H}_0}), \qquad J := J_{T^{-1}}(\mathbf{x}^{\mathcal{H}_0}),$$

and we assume that T^{-1} is twice continuously differentiable with a Hessian bounded on a neighborhood of the support of $P_{\mathcal{H}_0}$, so that the second-order Taylor remainder below is controlled.

Under $\mathcal{H}_1$, the observation is $\mathbf{x}^{\mathcal{H}_1} = \mathbf{x}^{\mathcal{H}_0} + \mathbf{p}$. A first-order Taylor expansion of T^{-1} around $\mathbf{x}^{\mathcal{H}_0}$ gives

$$T^{-1}(\mathbf{x}^{\mathcal{H}_1}) \;=\; \mathbf{z}^{\mathcal{H}_0} \;+\; J\,\mathbf{p} \;+\; \mathbf{R}(\mathbf{x}^{\mathcal{H}_0}, \mathbf{p}), \qquad \|\mathbf{R}\| = O\big(\|\mathbf{p}\|^2\big), \tag{43}$$

the remainder being controlled by the Hessian bound above.

We expand T^{-1} to first order in $\mathbf{p}$ and take the squared norm of the resulting map, which gives the approximation used in Section 5,

$$S(\mathbf{x}^{\mathcal{H}_1}) \;\approx\; \big\|\mathbf{z}^{\mathcal{H}_0} + J\,\mathbf{p}\big\|^2 \;=\; \underbrace{\|\mathbf{z}^{\mathcal{H}_0}\|^2}_{=\,S(\mathbf{x}^{\mathcal{H}_0})} \;+\, 2\,(\mathbf{z}^{\mathcal{H}_0})^\top J\,\mathbf{p} \;+\; \|J\mathbf{p}\|^2. \tag{44}$$

The quadratic term $\|J\mathbf{p}\|^2$ arises from this squaring rather than from a second-order expansion of the map itself. The approximation is exact when T^{-1} is affine, in which case $\mathbf{R} \equiv \mathbf{0}$; when T^{-1} is nonlinear, it degrades as $\|\mathbf{p}\|$ grows, i.e., at high signal-to-noise ratio (defined in Section 6.2) — precisely the regime where detection is in practice easiest, and thus least critical to characterize precisely.

Even under equation 44, the law of $S(\mathbf{x}^{\mathcal{H}_1})$ has no closed form. The reason is that J is evaluated at $\mathbf{x}^{\mathcal{H}_0}$, which is random, so J is random too. Under $\mathcal{H}_0$ the statistic depends on a single Gaussian vector, $\mathbf{z}^{\mathcal{H}_0} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, which is why it follows a χ^2_d law. Under $\mathcal{H}_1$, it depends on two coupled quantities, $\mathbf{z}^{\mathcal{H}_0}$ and J, and computing its law would require their joint distribution. This joint law is determined by the geometry of T and is not available in general. This is the same asymmetry as evoked in Section 3.2, now manifesting through the randomness of J.

We therefore turn to the first moment, which is tractable. Taking the expectation of equation 44 over $\mathbf{x}^{\mathcal{H}_0} \sim P_{\mathcal{H}_0}$ and using the linearity of $\mathbb{E}$,

$$\mathbb{E}\big[S(\mathbf{x}^{\mathcal{H}_1})\big] \;\approx\; \underbrace{\mathbb{E}\big[\|\mathbf{z}^{\mathcal{H}_0}\|^2\big]}_{=d} \;+\; \underbrace{2\,\mathbb{E}\big[(\mathbf{z}^{\mathcal{H}_0})^\top J\big]\,\mathbf{p}}_{=:C} \;+\; \underbrace{\mathbf{p}^\top\,\mathbb{E}\big[J^\top J\big]\,\mathbf{p}}_{=:M}, \tag{45}$$

where $\mathbb{E}[\|\mathbf{z}^{\mathcal{H}_0}\|^2] = d$ because $S(\mathbf{x}^{\mathcal{H}_0}) \sim \chi^2_d$ under the exact transport. Explicitly,

$$C := 2\,\mathbb{E}\big[T^{-1}(\mathbf{x}^{\mathcal{H}_0})^\top J_{T^{-1}}(\mathbf{x}^{\mathcal{H}_0})\big]\,\mathbf{p}, \qquad M := \mathbf{p}^\top\,\mathbb{E}\big[J_{T^{-1}}(\mathbf{x}^{\mathcal{H}_0})^\top J_{T^{-1}}(\mathbf{x}^{\mathcal{H}_0})\big]\,\mathbf{p}. \tag{46}$$

Both are deterministic, $\mathbf{p}$ being deterministic: C is linear in $\mathbf{p}$, M is quadratic. Since $\mathbb{E}[J^\top J] \succeq 0$ as an expectation of positive semi-definite matrices, we have $M \geq 0$, with $M > 0$ as soon as J is almost surely invertible and $\mathbf{p} \neq \mathbf{0}$. The cross term C, on the other hand, has no definite sign in general: it is an average of the alignment between the transported observation $\mathbf{z}^{\mathcal{H}_0}$ and the transported signal direction $J\mathbf{p}$, and nothing forces this alignment to be positive. Sections 5 and Appendix C.1 examine the two cases where it can be evaluated.

Consistency with Example 1. When $P_{\mathcal{H}_0} = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the map $T^{-1}(\mathbf{x}) = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ is affine, so $J = \boldsymbol{\Sigma}^{-1/2}$ is deterministic and the expansion above is exact. Then $C = 2\,\mathbb{E}[\mathbf{z}^{\mathcal{H}_0}]^\top \boldsymbol{\Sigma}^{-1/2}\mathbf{p} = 0$ since $\mathbf{z}^{\mathcal{H}_0} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, and $M = \mathbf{p}^\top \boldsymbol{\Sigma}^{-1}\mathbf{p}$, recovering the noncentrality parameter of the $\chi'^2_d(M)$ law and the exact identity $\mathbb{E}[S(\mathbf{x}^{\mathcal{H}_1})] = d + M$.

Figure 6 uses this exact law as a reference to assess the learned transport. We set $\mathbf{p} = A\mathbf{s}$, with $\mathbf{s}$ a fixed unit vector and A swept from 0 to 5. For each A, 5000 $\mathcal{H}_1$ samples are drawn and thresholded against λ, set at $P_{fa} = 0.05$ via the χ^2_d quantile. The maximum gap between the empirical power obtained with T_θ^{-1} and the exact curve is 0.019 over the grid, consistent with the small inversion error of the model used here ($\varepsilon \approx 0.10$).

C.1 Computation of C and M for a spherical distribution

We compute here the terms C and M derived in section 5 and appendix C. The expression of the transport is given by equation 41 :

$T^{-1}(\mathbf{x}) = \phi(r)\boldsymbol{\omega} = g(r)\boldsymbol{\omega}$ with $g = \phi$. For a general radial field $h(\mathbf{x}) = g(r)\boldsymbol{\omega} = f(r)\,\mathbf{x}$, with $f(r) := g(r)/r$, we compute its Jacobian componentwise, using $\partial r/\partial x_j = \theta_j$:

$$\frac{\partial h_i}{\partial x_j} = f'(r)\,\theta_j\,x_i + f(r)\,\delta_{ij} = f'(r)\,r\,\theta_i\theta_j + f(r)\,\delta_{ij}, \tag{47}$$

since $x_i = r\theta_i$. Substituting $f(r) = g(r)/r$ and $f'(r)\,r = g'(r) - g(r)/r$, this gives

$$J[g(r)\boldsymbol{\omega}] = g'(r)\,\boldsymbol{\omega}\boldsymbol{\omega}^T + \frac{g(r)}{r}\big(\mathbf{I} - \boldsymbol{\omega}\boldsymbol{\omega}^T\big). \tag{48}$$

Applying this with $g = \phi$, the Jacobian of T^{-1} reads

$$J_{T^{-1}}(\mathbf{x}) = \phi'(r)\,\boldsymbol{\omega}\boldsymbol{\omega}^T + \frac{\phi(r)}{r}\big(\mathbf{I} - \boldsymbol{\omega}\boldsymbol{\omega}^T\big). \tag{49}$$

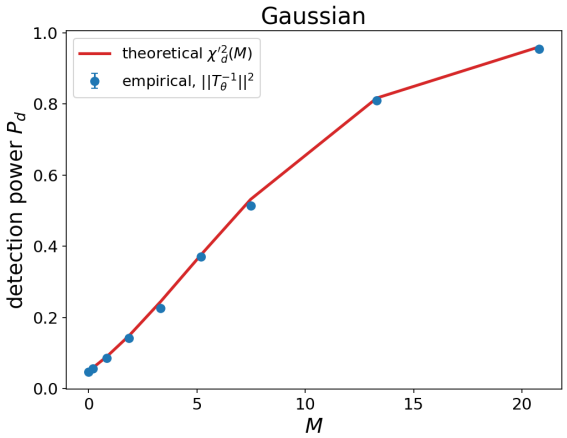

Figure 6: Detection power P_d as a function of the non-centrality parameter M, for the Gaussian scenario. Solid red: exact theoretical power $1 - F_{\chi'^2_d(M)}(\lambda)$. Blue markers with binomial standard-error bars: empirical power using the learned transport T_θ^{-1}.

Cross term C. Using $\mathbf{z}^{\mathcal{H}_0} = \phi(r)\,\boldsymbol{\omega}$ and left-multiplying equation 49 by $(\mathbf{z}^{\mathcal{H}_0})^T = \phi(r)\boldsymbol{\omega}^T$,

$$(\mathbf{z}^{\mathcal{H}_0})^T J_{T^{-1}}(\mathbf{x}) = \phi(r)\,\boldsymbol{\omega}^T\Big[\phi'(r)\,\boldsymbol{\omega}\boldsymbol{\omega}^T + \frac{\phi(r)}{r}\big(\mathbf{I} - \boldsymbol{\omega}\boldsymbol{\omega}^T\big)\Big]. \tag{50}$$

Since $\boldsymbol{\omega}^T\boldsymbol{\omega} = 1$ and $\boldsymbol{\omega}^T(\mathbf{I} - \boldsymbol{\omega}\boldsymbol{\omega}^T) = \mathbf{0}$, the second term vanishes and

$$(\mathbf{z}^{\mathcal{H}_0})^T J_{T^{-1}}(\mathbf{x}) = \phi(r)\phi'(r)\,\boldsymbol{\omega}^T. \tag{51}$$

Recalling $C = 2\,\mathbb{E}[(\mathbf{z}^{\mathcal{H}_0})^T J_{T^{-1}}(\mathbf{x})]\,\mathbf{p}$ and using independence of r and $\boldsymbol{\omega}$,

$$C = 2\,\mathbb{E}[\phi(r)\phi'(r)]\,\mathbb{E}[\boldsymbol{\omega}^T]\,\mathbf{p} = 0, \tag{52}$$

since $\mathbb{E}[\boldsymbol{\omega}] = \mathbf{0}$ by spherical symmetry.

Quadratic term M. Recall $M = \mathbf{p}^T\,\mathbb{E}\big[J_{T^{-1}}(\mathbf{x})^T J_{T^{-1}}(\mathbf{x})\big]\,\mathbf{p}$. Since $\boldsymbol{\omega}\boldsymbol{\omega}^T$ and $(\mathbf{I} - \boldsymbol{\omega}\boldsymbol{\omega}^T)$ are symmetric, idempotent, and mutually orthogonal projectors, the bracket in equation 49 squares to

$$J_{T^{-1}}(\mathbf{x})^T J_{T^{-1}}(\mathbf{x}) = \phi'(r)^2\,\boldsymbol{\omega}\boldsymbol{\omega}^T + \frac{\phi(r)^2}{r^2}\big(\mathbf{I} - \boldsymbol{\omega}\boldsymbol{\omega}^T\big). \tag{53}$$

Taking expectations and using independence of r and $\boldsymbol{\omega}$,

$$\mathbb{E}\big[J_{T^{-1}}(\mathbf{x})^T J_{T^{-1}}(\mathbf{x})\big] = \mathbb{E}[\phi'(r)^2]\,\mathbb{E}[\boldsymbol{\omega}\boldsymbol{\omega}^T] + \mathbb{E}\Big[\frac{\phi(r)^2}{r^2}\Big]\big(\mathbf{I} - \mathbb{E}[\boldsymbol{\omega}\boldsymbol{\omega}^T]\big). \tag{54}$$

By spherical symmetry, $\boldsymbol{\omega}$ is uniform on the unit sphere, so $\mathbb{E}[\boldsymbol{\omega}\boldsymbol{\omega}^T] = \mathbf{I}/d$, and $\mathbf{I} - \mathbb{E}[\boldsymbol{\omega}\boldsymbol{\omega}^T] = \frac{d-1}{d}\mathbf{I}$. Hence, the expectation reduces to a scalar multiple of the identity,

$$\mathbb{E}\big[J_{T^{-1}}(\mathbf{x})^T J_{T^{-1}}(\mathbf{x})\big] = \kappa\,\mathbf{I}, \qquad \kappa := \frac{1}{d}\left(\mathbb{E}[\phi'(r)^2] + (d-1)\,\mathbb{E}\Big[\frac{\phi(r)^2}{r^2}\Big]\right), \tag{55}$$

and therefore

$$M = \kappa\,\|\mathbf{p}\|^2. \tag{56}$$

C.2 Numerical validation for a spherical distribution

Figure 7 experimentally addresses that C vanishes in expectation. The experiments use only the exact transport, not the learned model: the Jacobian $J_{T^{-1}}$ is estimated via batched central finite differences, since automatic differentiation is not available through the scipy-based radial CDF inversion.

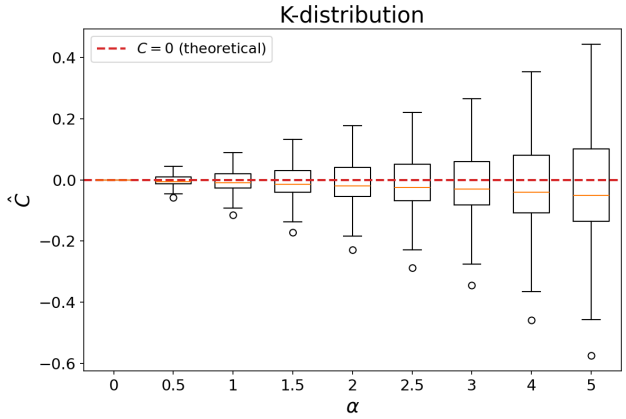

Figure 7: Distribution of the estimated cross term $\hat{C}$ as a function of the signal amplitude A, for the K-distribution scenario. One box per A, aggregated over 30 independent batches. Dashed red line: theoretical value $C = 0$.

D Supplementary materials for experiments

D.1 Reflow

When specified in the experiments, to further straighten the trajectories induced by the flow matching model, we apply the *reflow* procedure of Liu et al. (2023a), summarized in Algorithm 1. Starting from the model $v_\theta^{(0)}$ obtained by standard CFM training with an independent coupling $(\mathbf{x}_0, \mathbf{x}_1)$, reflow proceeds in rounds. At round k, new source samples $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ are drawn and integrated forward through the current model $v_\theta^{(k)}$ to obtain $\hat{\mathbf{x}}_1$, yielding a new, *deterministic* coupling $(\mathbf{x}_0, \hat{\mathbf{x}}_1)$ in place of the original independent one. The forward integration uses N_r numerical ODE steps to construct the reflow coupling. A new model $v_\theta^{(k+1)}$ is then trained by CFM on this coupling, exactly as at round 0.

This construction has two useful properties. First, it is *marginal-preserving*: regardless of how accurately $v_\theta^{(k)}$ approximates the true velocity field, the resulting flow still pushes $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ forward to (an approximation of) $P_{\mathcal{H}_0}$ at every round — reflow only changes how source and target samples are paired, never the target distribution itself. Second, each round tends to make the paired trajectories straighter, in the sense that it does not increase, and generally decreases, the average transport cost between coupled samples (Liu et al., 2023a). This is what allows fewer integration steps to be used at inference after reflow, without degrading the quality of the transport.

Algorithm 1 Reflow

1: **Input:** velocity model v_θ; number of rounds K.
2: Train $v_\theta^{(0)}$ by minimizing $\mathcal{L}_{\mathrm{CFM}}$ with independent coupling $(\mathbf{x}_0, \mathbf{x}_1)$, $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, $\mathbf{x}_1 \sim P_{\mathcal{H}_0}$.
3: **for** $k = 0, \dots, K-1$ **do**
4: Draw $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$; integrate forward through $v_\theta^{(k)}$ using N_r steps to obtain $\hat{\mathbf{x}}_1$.
5: Train $v_\theta^{(k+1)}$ by minimizing $\mathcal{L}_{\mathrm{CFM}}$ on the deterministic coupling $(\mathbf{x}_0, \hat{\mathbf{x}}_1)$.
6: **end for**
7: **Return:** $v_\theta^{(K)}$.

D.2 Details for synthetic distribution generation

A K-distributed vector of dimension d is generated as $\mathbf{x} = \sqrt{\Gamma}\,\mathbf{z}$, with $\Gamma \sim \mathrm{Gamma}(\mathrm{shape}, \mathrm{scale})$ independent of $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$; this construction yields an isotropic distribution. Mixtures are formed

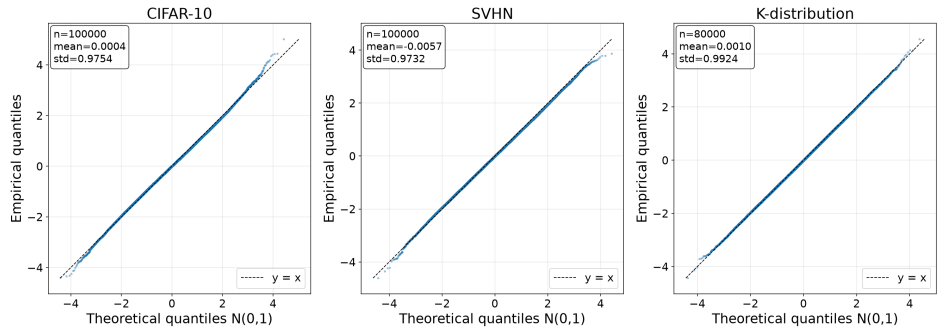

Figure 8: Quantile–quantile plots of the transported coordinates against the standard Gaussian N(0,1), for CIFAR-10 (left), SVHN (middle) and K-distributed clutter (right). Each point compares an empirical quantile with the matching N(0,1) quantile, and the dashed line y = x is a perfect Gaussian fit.

by drawing the component index according to the mixture weights before sampling Γ and z. Parameter values used for the different K-distributions are given in Table 3.

Table 3: Source distribution parameters.

Distribution	**Components**	**(shape, scale)**
Simple K-distribution	-	$(2.0,\ 1.0)$
Mixture	component 1 (weight 0.5)	$(2.0,\ 1.0)$
	component 2 (weight 0.5)	$(6.0,\ 0.5)$

D.3 Experimental setup

This section details the experimental setup used in the simulations.

Table 4 summarizes the parameters used for the velocity field estimation and the ODE solver as presented in section 2.

Table 4: Velocity field architecture and training hyperparameters for radar detection.

Component	**Value**
Input	$(\mathbf{x}, t) \in \mathbb{R}^d \times [0, 1]$
Time embedding	sinusoidal, dimension 64
Hidden layers	3 fully-connected layers, width 256
Activation	SiLU
Optimizer	Adam
Learning rate	10^{-3}, cosine annealing over training
Batch size	256
Training steps	5,000
Solver for ODE	explicit Euler
Nb of steps ODE	10

Neural parameterization. The same flow-matching framework, training objective, numerical inversion procedure, and scoring rule are used across all experiments. For vector-valued data (radar and bio-informatics), the velocity field is parameterized by an MLP, while for image-based OOD detection, it is parameterized by a time-conditioned U-Net to preserve spatial structure. The corresponding architectural and optimization details are reported in Tables 4 and 5.

Table 5: Architecture of the time-conditioned U-Net used for OOD detection.

Component	Value
Input	$3 \times 32 \times 32$ RGB image
Backbone	Time-conditioned U-Net
Time embedding	Sinusoidal, dimension 64
Encoder/decoder levels	3
Activation	SiLU
Upsampling	2×2 transposed convolution
Downsampling	2×2 max pooling

Hardware and software. All experiments were conducted on a 64-bit Windows 11 system equipped with a 13th Gen Intel Core i9-13950HX CPU, 32 GB of RAM, and an NVIDIA RTX 3500 Laptop GPU with 12 GB of VRAM. All experiments were implemented in PyTorch.

D.4 Performances metrics

We evaluate our method using task-specific performance metrics. For radar detection, we report the probability of detection and its dependence on SNR. For imbalanced classification, we report the F1 score, AUROC, and AUPRC. For attribute matching, we report the correct attribute correspondence rate. For OOD detection, we report AUROC. AUROC refers to the area under the receiver operating characteristic curve. We measure it by calculating the area under the curve when we plot true positive rate ($TPR = \frac{TP}{(TP+FN)}$) against false positive rate ($FPR = \frac{FP}{(FP+TN)}$). AUPRC refers to the area under the precision-recall curve, where precision = $\frac{TP}{(TP+FP)}$ and recall = $\frac{TP}{(TP+FN)}$. F1 score is $2\frac{(Precision \times Recall)}{(Precision+Recall)}$.

D.5 Details for radar target simulations

Setup. We consider complex-valued data $\mathbf{x} \in \mathbb{C}^d$, $d = 16$, corresponding to d pulses of a coherent processing interval. The clutter covariance matrix $\mathbf{\Sigma}$ has Toeplitz structure, with correlation coefficient $\rho = 0.5$. Under $\mathcal{H}_0$, $\mathbf{x}$ follows either a Gaussian clutter model, $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$, or a K-distributed clutter model with the same covariance structure. Under $\mathcal{H}_1$, a target is added, $\mathbf{x} = A\,\mathbf{p} + \mathbf{c}$, with $\mathbf{c}$ drawn from the same clutter model as under $\mathcal{H}_0$, $A \in \mathbb{C}$ the unknown target complex amplitude, and $\mathbf{p}$ the steering vector at normalized target Doppler frequency f_0,

$$\mathbf{p} = \frac{1}{\sqrt{d}}\left[1,\, e^{i2\pi f_0},\, \ldots,\, e^{i2\pi(d-1)f_0}\right]^\top, \tag{57}$$

assumed perfectly known (matched case). The resulting signal-to-noise ratio is $\mathrm{SNR} = |A|^2\,\mathbf{p}^H\mathbf{\Sigma}^{-1}\mathbf{p}$.

Benchmark algorithms. Detection is traditionally performed via the generalized likelihood ratio test (GLRT) between $\mathcal{H}_0$ and $\mathcal{H}_1$, assuming Gaussian clutter under both hypotheses. This yields the Matched Filter (MF) statistic (Kelly, 1986),

$$\Lambda_{\mathrm{MF}}(\mathbf{x}) = \frac{\left|\mathbf{p}^H\mathbf{\Sigma}^{-1}\mathbf{x}\right|^2}{\mathbf{p}^H\mathbf{\Sigma}^{-1}\mathbf{p}}, \tag{58}$$

optimal under this Gaussian assumption. We compare against the following detectors:

- **MF** Kelly (1986): assumes $\mathbf{\Sigma}$ perfectly known; since this is not achievable in practice, we report it as an oracle upper bound.
- **AMF**: replaces $\mathbf{\Sigma}$ in Λ_{MF} by the sample covariance matrix (SCM) estimator, $\widehat{\mathbf{\Sigma}}_{\mathrm{SCM}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_l\mathbf{x}_l^H$. Since the SCM is itself derived under a Gaussian assumption, it becomes mismatched under K-distributed clutter.

- **ANMF-Tyler**: better suited to non-Gaussian, compound-Gaussian clutter, by replacing the SCM with Tyler's M-estimator (Tyler, 1987), the fixed point of

$$\widehat{\mathbf{\Sigma}}_{\text{Tyler}} = \frac{d}{L}\sum_{l=1}^{L}\frac{\mathbf{x}_l\mathbf{x}_l^H}{\mathbf{x}_l^H\widehat{\mathbf{\Sigma}}_{\text{Tyler}}^{-1}\mathbf{x}_l}, \tag{59}$$

  and by normalizing Λ_{MF} to make the test invariant to the (unknown) clutter texture, yielding the ANMF statistic.
- **SVDD** Pinsolle et al. (2026): trained on $\mathcal{H}_0$ data to project observations onto a compact hypersphere; the detection score for a test sample is its distance to the center of this sphere.

For AMF and ANMF-Tyler, $\mathbf{\Sigma}$ is estimated from $L = 32$ secondary data samples, drawn independently from the same clutter distribution as the cell under test and free of target contamination.

D.6 Inference time

We further evaluate the computational efficiency of the proposed method in terms of inference time for radar target detection.

Table 6 reports the average per-sample detection time of the evaluated radar target detectors. The time is calculated as $\bar{t} = \frac{1}{\mathcal{S}}\sum_{i=1}^{\mathcal{S}}\hat{t}_i$, where $\hat{t}_i = \frac{T_i}{\mathcal{N}}$, $\mathcal{S}$ is the number of SNR values considered (0-20dB), T_i is the total detection time over all test samples at the i-th SNR for a given Doppler bin, and $\mathcal{N}$ is the number of test samples per SNR. In our experiments, $\mathcal{N} = 1000$ for a single Doppler bin. As shown, the proposed method has the lowest average per-sample inference time among the evaluated methods. This indicates that the proposed approach provides favorable computational efficiency while maintaining its detection performance under non-Gaussian clutter and for weak targets.

Table 6: Average per-sample inference time $\bar{t}$ (ms) for radar target detection across SNR values from 0-20dB. Lower is better.

Method	MF	AMF	ANMF-Tyler	SVDD	ours
Time ($\bar{t}$)	0.0516	0.0970	1.9255	0.4042	**0.0209**

D.7 Additional Results

This section reports additional experimental results that complement the experiments mentioned in the main paper.

Radar Target Detection Analysis across multiple Doppler Bins. To analyze detection performance across multiple Doppler bins, we average P_d over 16 Doppler bins and report the resulting mean detection probability at representative SNR levels in Table 7. Under K-distributed clutter, ANMF-Tyler achieves higher mean P_d than our method in the low-SNR regime ($8 - 10$dB). At medium SNR, e.g., 12dB, the performance gap narrows, while our method achieves a higher mean P_d at 14dB. In addition, our method reaches a mean $P_d \geq 0.9$ at 14dB, compared with 15dB for AMF and AMF-Tyler. These results demonstrate that the proposed detector is competitive with the considered baselines and achieves high detection probability at medium-to-high SNRs.

OOD Detection evaluation using AUPR and FPR95. We further compare our method with several recent OOD detection methods using AUPR and FPR95. As shown in Table 8, our method achieves competitive performance, with the lowest FPR95 among the compared methods on SVHN→C100. These results provide further evidence of the effectiveness of our method for OOD detection.

D.8 Comparison with Pretrained-Backbone Methods

Table 1 in the main text compares methods trained solely on $\mathcal{H}_0$, the same distribution used as the reference for detection. For completeness, we report here in table 9 a separate comparison including

Table 7: Mean detection probability P_d across 16 Doppler bins at selected SNR levels under K-distributed clutter. SNR_{90} denotes the SNR at which mean detection probability across the 16 Doppler bins reaches $P_d \geq 0.9$.

Method	AMF	ANMF-Tyler	SVDD	Ours
P_d@8dB	0.193	0.348	0.160	0.198
P_d@10 dB	0.366	0.545	0.335	0.400
P_d@12dB	0.610	0.748	0.703	0.706
P_d@14dB	0.840	0.894	0.918	0.947
SNR_{90}	15	15	14	14

Table 8: OOD detection performance in terms of AUPR (higher is better) and FPR95 (lower is better). The corresponding AUROC results are reported in Table 1.

	C10→SVHN		SVHN→C100	
Method	AUPR ↑	FPR95 ↓	AUPR ↑	FPR95 ↓
ReAct Sun et al. (2021)	0.204	0.996	**0.961**	0.255
ASH Djurisic et al. (2022)	0.750	0.977	0.200	0.977
Ours	**0.769**	**0.839**	0.922	**0.180**

methods that instead rely on an external, pretrained image backbone — typically a large generative model pretrained on a broad image corpus (e.g., CelebA-32) rather than on $\mathcal{H}_0$ itself.

Table 9: OOD detection performance (AUROC) against methods relying on a pretrained backbone on CelebA-32.

	CIFAR-10 (ID)			**SVHN (ID)**			**Avg**
Method	SVHN	CelebA	C100	C10	CelebA	C100	
Ours	0.503	0.695	0.511	0.897	0.993	0.963	0.760
SCOPED Barkley et al. (2026)	0.814	0.940	0.477	0.971	0.996	0.959	0.859
GEPC Rouzoumka et al. (2026)	0.842	0.999	0.554	0.880	1.00	0.897	0.862

SCOPED and GEPC achieve a slightly higher average AUROC than our method, but at the cost of requiring a backbone pretrained on a large, broad-coverage image corpus — an assumption that does not extend to the other application domains considered in this work (biological classification, radar target detection), where no such generic pretrained backbone exists. We also note that neither SCOPED nor GEPC uniformly dominates: on CIFAR-10 → SVHN and CIFAR-10 → CIFAR-100, our method remains competitive with or ahead of at least one of the two.

D.9 Ablation studies

Effect of Reflow: To evaluate the contribution of Reflow, we compare our method with a variant in which Reflow is removed, while keeping all other settings unchanged. As shown in Table 10, incorporating Reflow improves AUROC. This result indicates that Reflow contributes meaningfully to OOD detection performance beyond the other components of our method.

Table 10: Ablation of the Reflow component for OOD detection.

ID/OOD pair	CIFAR-10→SVHN			CIFAR-10→CIFAR-100			SVHN→CIFAR-100		
Metrics	AUROC	AUPR	FPR95	AUROC	AUPR	FPR95	AUROC	AUPR	FPR95
Without	0.446	0.702	0.986	0.495	0.520	0.932	0.951	0.954	0.408
With ReFlow	0.503	0.769	0.839	0.511	0.526	0.977	0.963	0.922	0.180

To study the sensitivity of reflow to the number of numerical integration steps, we vary the number of integration steps used to construct the reflow coupling for OOD detection in the SVHN (ID) → CIFAR-100 (OOD) setting, while keeping all other training and evaluation settings fixed. As shown in Tab. 11, the number of integration steps affects OOD detection performance. Among

the evaluated settings, $N_r = 20$ achieves the highest AUROC and AUPR and the lowest FPR95, demonstrating the strongest OOD detection performance.

Table 11: Effect of number of ReFlow integration steps N_r on OOD detection performance. The number of rounds is fixed to 3, while N_r used to construct reflow coupling is varied. Results are reported on SVHN (ID) $\rightarrow$ CIFAR-100 (OOD). Higher AUROC, AUPR, and lower FPR95 is better.

Steps N_r	10	20	30
AUROC	0.898	0.963	0.954
AUPR	0.794	0.922	0.903
FPR95	0.429	0.180	0.219

Seed Robustness: To assess the sensitivity of our method to random initialization, we repeat the OOD detection experiment using five different random seeds while keeping all other hyperparameters fixed. We report the results in Table 12. In case of SVHN (ID), the method achieves an average AUROC of 0.945 ± 0.015, with consistently strong performance across all seeds, demonstrating limited sensitivity to random initialization.

Table 12: Per-seed AUROC for OOD detection using SVHN as ID and CIFAR-100 as OOD, providing an ablation of the result reported in Table 1.

ID/OOD pair	AUROC	AUPR	FPR95
CIFAR-10$\rightarrow$SVHN	0.495 ± 0.010	0.764 ± 0.007	0.982 ± 0.019
SVHN$\rightarrow$CIFAR-100	0.945 ± 0.015	0.872 ± 0.028	0.273 ± 0.076

D.10 Attribute matching

We evaluate whether the proposed transport-based representation can identify attributes that correspond to the same underlying quantity based on their empirical distributions. We used the Protein Homology dataset Caruana et al. (2004) from the 2004 KDD Cup (BIO). Following Gretton et al. Gretton et al. (2012), we randomly sample 754 observations from the dataset and split them into two subsets, A and B, of 377 observations each. We consider the first $n = 6$ attributes and compare each attribute A_i with every other attribute B_j, yielding 36 candidate pairs. The original attribute ordering provides the ground truth: A_i and B_i are corresponding pairs, while A_i and B_j for $i \neq j$ are non-corresponding pairs. We repeat the experiment 100 times with independently sampled subsets and random splits. For each A_i, we train a separate transport model from a standard Gaussian source to the empirical distribution of A_i. We then map A_i and each candidate B_j to the resulting transport space and measure their compatibility by the squared distance between their empirical mean representations. We compare against the t-test, unbiased quadratic-time maximum mean discrepancy (MMD) with bootstrap (MMD_u^2 B)Gretton et al. (2012), and MMD using linear time statistic (MMD_l^2)Gretton et al. (2012), using a Gaussian kernel for both MMD variants. The null hypothesis is accepted at $\alpha = 0.05$ for the statistical tests.

Table 13: Attribute matching results on BIO. Values denote the percentage of times the null hypothesis is accepted. Results are pooled over 100 repetitions, with 377 samples per subset and 6 attributes. (Higher is better for corresponding pairs and lower for non-corresponding pairs).

Methods	t-test	MMD_u^2 B	MMD_l^2	Ours
Corresponding	95.2	92.5	94.8	**95.7**
Non-corresponding	7.8	**3.5**	22.0	7.5

Table 13 shows that the proposed representation accepts 95.7% of corresponding pairs while accepting only 7.5% of non-corresponding pairs. These results show that our method preserves the similarity between independently sampled observations of the same attribute while effectively separating different attributes. $\text{MMD}_u^2 B$ achieves the lowest non-corresponding acceptance rate (3.5%), while our method provides a competitive balance with a higher corresponding acceptance rate (95.7%).