
DFM: Interpolant-free Dual Flow Matching

Denis Gudovskiy¹

Tomoyuki Okuno²

Yohei Nakata²

¹Panasonic AI Lab, USA

²Panasonic Holdings Corporation, Japan

denis.gudovskiy@us.panasonic.com

{okuno.tomoyuki, nakata.yohei}@jp.panasonic.com

Abstract

Continuous normalizing flows (CNFs) can model data distributions with expressive infinite-length architectures. But this modeling involves computationally expensive process of solving an ordinary differential equation (ODE) during maximum likelihood training. Recently proposed flow matching (FM) framework allows to substantially simplify the training phase using a regression objective with the interpolated forward vector field. In this paper, we propose an interpolant-free dual flow matching (DFM) approach without explicit assumptions about the modeled vector field. DFM optimizes the forward and, additionally, a reverse vector field model using a novel objective that facilitates bijectivity of the forward and reverse transformations. Our experiments with the SMAP unsupervised anomaly detection show advantages of DFM when compared to the CNF trained with either maximum likelihood or FM objectives with the state-of-the-art performance metrics.

1 Introduction

Discrete-time (DNF) and continuous-time (CNF) normalizing flows have been previously extensively studied and compared in details [Ruthotto and Haber, 2021]. With the pros and cons in each approach, we are motivated to apply CNFs in real-world generative and density estimation applications. Theoretically, an infinite-length architecture with arbitrary parameterization can lead to a significant advantages of CNFs when compared to DNF shortcomings. Meanwhile, CNFs with the ordinary differential equation (ODE) integration step endure higher computational complexity, numerical instabilities and approximation errors [Chen et al., 2018]. In particular, the latter is crucial in density estimation which requires accurate estimates of the Jacobian matrix trace.

Recent flow matching (FM) framework [Lipman et al., 2023] simplifies the training phase in CNFs by introducing a new regression objective. This objective minimizes mean square difference of a parameterized vector field model and an interpolated vector field between two data distributions. While the former is a conventional neural network with time-dependent conditioning, the latter relies on certain assumptions about modeled data distributions. As a result, there is an extensive line of research that proposes various forms of the interpolated vector fields and the corresponding probability paths. We summarize recent works in Table 1 with the formal introduction in Section 2. Diffusion models [Song and Ermon, 2019, Ho et al., 2020] that solve stochastic differential equations (SDEs) can also be generalized using the FM framework [Tong et al., 2024].

In this paper, we analyze current interpolation-based FM approach and its inherent limitations i.e. the Gaussian probability path assumption between two data distributions [Lipman et al., 2023]. To address this limitation, we propose a novel *interpolant-free dual flow matching* (DFM) method. Specifically, we accomplish interpolant-free FM using an additional parameterized reverse vector field model. For simplicity, we model the reverse vector field using exactly the same architecture as for the forward one. Then, we optimize our DFM using an objective that enforces transformation bijectivity of the modeled forward and the reverse vector fields.

Table 1: Generalization of probability paths for diffusion and flow matching methods by the [Tong et al., 2024] framework. Unlike ours, these methods rely on an interpolation of the probability paths.

| Probability Path | $q(\mathbf{z})$ | $\mu_t(\mathbf{z})$ | σ_t |
|---|----------------------------------|---|-------------------------------|
| Var. Exploding SDE [Song and Ermon, 2019] | $q(\mathbf{x}_1)$ | \mathbf{x}_1 | $\frac{\sigma_{1-t}}{\sigma}$ |
| Var. Preserving SDE [Ho et al., 2020] | $q(\mathbf{x}_1)$ | $\alpha_{1-t}\mathbf{x}_1$ | $\sqrt{1 - \alpha_{1-t}^2}$ |
| FM [Lipman et al., 2023] | $q(\mathbf{x}_1)$ | $t\mathbf{x}_1$ | $t\sigma - t + 1$ |
| Rectified FM [Liu et al., 2023] | $q(\mathbf{x}_0)q(\mathbf{x}_1)$ | $t\mathbf{x}_1 + (1-t)\mathbf{x}_0$ | 0 |
| Var. Preserving FM [Albergo and Vanden-Eijnden, 2023] | $q(\mathbf{x}_0)q(\mathbf{x}_1)$ | $\cos(\pi t/2)\mathbf{x}_0 + \sin(\pi t/2)\mathbf{x}_1$ | 0 |
| I-CFM [Tong et al., 2024] | $q(\mathbf{x}_0)q(\mathbf{x}_1)$ | $t\mathbf{x}_1 + (1-t)\mathbf{x}_0$ | σ |

2 CNF Preliminaries and Prior FM Methods

Continuous normalizing flows. We follow Lipman et al. [2023] and Tong et al. [2024] notation. We consider a pair of data distributions $q(\mathbf{x}_0)$ and $q(\mathbf{x}_1)$ over \mathbb{R}^D with densities $p(\mathbf{x}_0)$ and $p(\mathbf{x}_1)$, respectively. Often, the $p_0 = p(\mathbf{x}_0)$ density represents a known prior distribution while the data density $p_1 = p(\mathbf{x}_1)$ is not given with only access to an empirical $\hat{q}(\mathbf{x}_1)$ and p_1 needs to be estimated.

Then, there are a *probability density path* $p : [0, 1] \times \mathbb{R}^D \rightarrow \mathbb{R}_{>0}$, which is a time-dependent probability density function $p_t(\mathbf{x})$ with $t \in [0, 1]$ such that $\int p_t(\mathbf{x})d\mathbf{x} = 1$, and a Lipschitz-smooth time-dependent *vector field* $u : [0, 1] \times \mathbb{R}^D \rightarrow \mathbb{R}^D$. The vector field u_t is used to construct a time-dependent diffeomorphism i.e., the CNF $\phi : [0, 1] \times \mathbb{R}^D \rightarrow \mathbb{R}^D$ that is defined via the ODE as

$$d\phi_t(\mathbf{x})/dt = u_t(\phi_t(\mathbf{x})) \text{ and } \phi_0(\mathbf{x}) = \mathbf{x}_0, \quad (1)$$

where $\phi_t(\mathbf{x})$ is the ODE solution with $\phi_0(\mathbf{x})$ initial condition that transports \mathbf{x} from time 0 to time t .

On the other hand, ϕ_t induces a push-forward $p_t = [\phi_t]_{\#}(p_0)$ that transports the density p_0 from time 0 to time t . The time-dependent density p_t is characterized by the *continuity equation* written by

$$\partial p_t(\mathbf{x})/\partial t = -\text{div}(p_t(\mathbf{x})u_t(\phi_t(\mathbf{x}))) = -\text{div}(f_t(\mathbf{x})), \quad (2)$$

where the divergence operator, div , is defined as the sum of derivatives of $f_t(\mathbf{x}) \in \mathbb{R}^D$ w.r.t. all elements x_d or, simply, the Jacobian matrix trace: $\text{div}(f(\mathbf{x})) = \sum_{d=1}^D \partial f_d(\mathbf{x})/\partial x_d = \text{Tr}(\mathbf{J})$.

The vector field $u_t(\phi_t(\mathbf{x}))$ is often modeled without $\phi_t(\mathbf{x})$ invertibility requirement by an arbitrary neural network $v_{\theta}(t, \mathbf{x}_t)$ with the learnable weight vector θ . Then, the continuity equation in (2) for (1) neural ODE can be written using the instantaneous change of variables [Chen et al., 2018] as

$$d \log p_t(\mathbf{x}_t)/dt + \text{Tr}(\partial v_{\theta}(t, \mathbf{x}_t)/\partial \mathbf{x}_t^T) = 0. \quad (3)$$

The (3) neural ODE can be solved both for a point \mathbf{x}_0 and the log-likelihood change as integration

$$\begin{bmatrix} \mathbf{x}_0 \\ \log(p_1/p_0) \end{bmatrix} = \int_{t=1}^{t=0} \begin{bmatrix} v_{\theta}(t, \mathbf{x}_t) \\ -\text{Tr}(\partial v_{\theta}(t, \mathbf{x}_t)/\partial \mathbf{x}_t^T) \end{bmatrix} dt, \text{ and initially } \begin{bmatrix} \mathbf{x}_t \\ \log(p_1/p_t) \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ 0 \end{bmatrix}. \quad (4)$$

Then, the CNF maximizes likelihood (MLE) during training and uses its estimate at the evaluation as

$$\arg \max_{\theta} \mathcal{L}_{\text{MLE}} := \log \hat{p}_1 = \log p_0 - \int_{t=1}^{t=0} \text{Tr}(\partial v_{\theta}(t, \mathbf{x}_t)/\partial \mathbf{x}_t^T) dt, \quad (5)$$

where $\log p_0$ is the likelihood of a point \mathbf{x}_0 from (4) when evaluated using a known prior $q(\mathbf{x}_0)$.

Flow matching training. Typically, solving the ODE (4) for the MLE objective (5) is computationally expensive [Grathwohl et al., 2019, Zhuang et al., 2020]. The FM framework [Lipman et al., 2023] proposes an alternative objective that regresses the $v_{\theta}(t, \mathbf{x}_t)$ to u_t by conditioning the latter by a vector $\mathbf{z} = \mathbf{x}_1$. This has been extended by the conditional FM (CFM) framework [Tong et al., 2024] where the $u_t(\mathbf{x}|\mathbf{z})$ and $p_t(\mathbf{x}|\mathbf{z})$ are conditioned on a more general $\mathbf{z} \sim q(\mathbf{z})$ such that the marginal probability density path and the corresponding marginal vector field are defined as

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x}|\mathbf{z})q(\mathbf{z}) d\mathbf{z}, \text{ and } u_t(\mathbf{x}) = \mathbb{E}_{\mathbf{z} \sim q(\mathbf{z})} [u_t(\mathbf{x}|\mathbf{z})p_t(\mathbf{x}|\mathbf{z})/p_t(\mathbf{x})]. \quad (6)$$

The Gaussian conditional probability path in (6) has the unique conditional vector field such that

$$p_t(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x} | \mu_t(\mathbf{z}), \sigma_t(\mathbf{z})^2 \mathbf{I}) \implies u_t(\mathbf{x}|\mathbf{z}) = (\mathbf{x} - \mu_t(\mathbf{z})) \sigma_t'(\mathbf{z}) / \sigma_t(\mathbf{z}) + \mu_t'(\mathbf{z}), \quad (7)$$

where the mean $\mu_t(\mathbf{z})$ and the standard deviation $\sigma_t(\mathbf{z})$ functions parameterize the path $p_t(\mathbf{x}|\mathbf{z})$.

Finally, the CFM objective for the simplified CNF training using (7) result can be written as

$$\arg \min_{\theta} \mathcal{L}_{\text{CFM}} := \mathbb{E}_{t \sim \mathcal{U}(0,1), \mathbf{x}_t \sim p_t(\mathbf{x}|\mathbf{z}), \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{z} \sim q(\mathbf{z})} \|v_{\theta}(t, \mathbf{x}_t) - u_t(\mathbf{x}|\mathbf{z})\|^2. \quad (8)$$

Recently proposed $\mu_t(\mathbf{z})$ and $\sigma_t(\mathbf{z})$ for interpolation of Gaussian conditional probability path $p_t(\mathbf{x}|\mathbf{z})$ in (7) are given in Table 1. The above CFM framework is inherently limited by the (6) interpolation assumption and (7) path choice. To overcome this, Chen and Lipman [2024] extend the Euclidean FM to Riemannian manifolds. Kapusniak et al. [2024] introduce the metric FM that learns parameterized interpolants. There are attempts to support FM for categorical data Stark et al. [2024], Cheng et al. [2024], Campbell et al. [2024] and mixed categorical-continuous data [Dunn and Koes, 2024]. In this paper, we are motivated by a different perspective: *is it feasible to avoid the interpolation in FM framework with a reasonable computational cost?*

3 The Proposed Interpolant-free Dual Flow Matching

Dual CNF via a reverse vector field. Let's extend the CNF framework that is defined in (1-4). We introduce a *dual CNF* where its first part, implemented as the above non-invertible $v_{\theta}(t, \mathbf{x}_t)$, approximates the vector field $u_t(\phi_t(\mathbf{x}))$. In addition, we employ an extension $v_{\lambda}(t, \mathbf{y}_t)$ with learnable parameters λ that models a *reverse vector field* model $u_t(\phi_t^{-1}(\mathbf{y}))$. In other words, there is the forward transformation $\mathbf{x}_t = \phi_t(\mathbf{x})$ and the inverse transformation $\mathbf{y}_t = \phi_t^{-1}(\mathbf{y}) = \phi_t^{-1}(\mathbf{x})$ of the bijective map ϕ_t .

Then, we can reformulate the equations (4) for $v_{\lambda}(t, \mathbf{y}_t)$ model with minor modifications. The neural ODE in (3) can be solved in reverse simultaneously for a point \mathbf{x}_1 and the log-likelihood change with the initial condition $\mathbf{y}_t \sim q(\mathbf{x}_0)$ and $\log(p_0/p_t) = 0$ as integration

$$\begin{bmatrix} \mathbf{x}_1 \\ \log(p_0/p_1) \end{bmatrix} = \int_{t=0}^{t=1} \begin{bmatrix} v_{\lambda}(t, \mathbf{y}_t) \\ -\text{Tr} \left(\frac{\partial v_{\lambda}(t, \mathbf{y}_t)}{\partial \mathbf{y}_t^T} \right) \end{bmatrix} dt. \quad (9)$$

Interestingly, the (9) approach with the modified maximum likelihood $\mathcal{L}_{\text{MLE}}(\lambda)$ is known in the DNF literature as a reverse divergence objective [Papamakarios et al., 2021]. When the target data p_1 cannot be analytically evaluated, the (9) is impractical for CNF training with the MLE objective.

Interpolant-free DFM. On the other hand, the proposed dual CNF with the reverse model in (9) can be used for the interpolant-free flow matching. Instead of the less expressive *affine transformation* $(\phi_t(\mathbf{x}|\mathbf{z}) = \mu_t(\mathbf{z}) + \sigma_t \mathbf{x}, \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}))$ induced by the Gaussian interpolation (7), the proposed DFM only requires *bijectivity of the free-form transformations* ϕ_t and ϕ_t^{-1} produced by, correspondingly, the forward $v_{\theta}(t, \mathbf{x}_t)$ and the reverse $v_{\lambda}(t, \mathbf{y}_t)$ vector field models.

Then, the proposed dual CNF with the bijective ϕ_t can be expressed as ODEs expressed by

$$\begin{cases} d\phi_t(\mathbf{x})/dt = u_t(\phi_t(\mathbf{x})) = v_{\theta}(t, \mathbf{x}_t) \\ d\phi_t(\mathbf{y})/dt = u_t(\phi_t(\mathbf{y})) = v_{\lambda}(t, \mathbf{y}_t). \end{cases} \quad (10)$$

Assuming the $\phi_t(\mathbf{y}) = \phi_t^{-1}(\mathbf{x})$ bijectivity in a neighborhood of t for \mathbf{x} and \mathbf{y} , (10) can be rewritten using the univariate inverse function theorem by substituting the top to bottom as

$$d\phi_t^{-1}(\mathbf{x})/dt = 1 / (d\phi_t(\mathbf{x})/dt) \implies \text{diag}(v_{\theta}(t, \mathbf{x}_t) \odot v_{\lambda}(t, \mathbf{y}_t)) = \mathbf{I}. \quad (11)$$

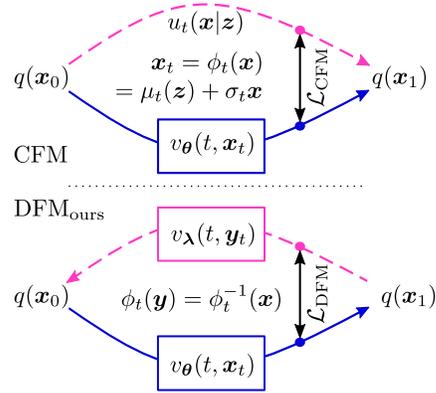


Figure 1: The CFM (top) regresses the Gaussian-interpolated forward vector field by a neural network with the *affine transformation* $\phi_t(\mathbf{x})$. Our DFM (bottom) has two neural networks with the *free-form transformations* with only the bijectivity objective $\mathbf{x}_t = \phi_{\lambda}^{-1}(\phi_{\theta}(\mathbf{x}_t))$ for an arbitrary vector field and a probability path.

The (11) objective can be achieved by minimizing the cosine distance between two unit vectors as

$$\arg \min_{\theta, \lambda} \mathcal{L}_{\text{DFM}} := \mathbb{E}_{t \sim \mathcal{U}(0,1), \mathbf{x}_t \sim \hat{q}(\mathbf{x}_1), \mathbf{y}_t \sim q(\mathbf{x}_0)} \text{dist}_{\text{cos}}(v_{\theta}(t, \mathbf{x}_t), v_{\lambda}(t, \mathbf{y}_t)), \quad (12)$$

where this loss with vector normalization is more numerically stable in practice.

Once the (12) loss is minimized, the density estimation i.e. $\log \hat{p}_1$ can be performed using the conventional MLE approach (5) without the extension $v_{\lambda}(t, \mathbf{y}_t)$. On the other hand, the extension can be used to improve $\log \hat{p}_1$ by integrating (9) from $t = 1$ to $t = 0$. We use the latter strategy. While we yet to accomplish sampling experiments, we expect DFM to outperform previous FM methods due to the enforced bijectivity which is a common issue in ODEs [Gholami et al., 2019].

4 Experiments

Benchmark. We employ real-world SMAP [Hundman et al., 2018] time series benchmark for unsupervised anomaly detection. The soil moisture active passive satellite (SMAP) dataset contains soil samples and telemetry information from the Mars rover with 135K and 428K data points in the training (without anomalies) and test sets, respectively. SMAP data has 25 data dimensions collected from 55 entities. We follow Su et al. [2019] and transform the regression task into a classification task using sliding windows (window size = 8) and replication padding [Tuli et al., 2022].

Flow models. We report experimental results for the Glow-type DNF [Kingma and Dhariwal, 2018] from [Gudovskiy et al., 2024] with the state-of-the-art baselines. Second, we experiment with the vanilla CNF from Section 2 and the CNF trained using the CFM framework i.e., the FM from [Lipman et al., 2023] and I-CFM from [Tong et al., 2024]. All CNF models have exactly the same U-Net architecture [Ronneberger et al., 2015], learnable $\mathcal{N}(\mu, \sigma^2 I)$ prior and identical evaluation using (5).

Evaluation. We follow Su et al. [2019] and report precision (P), recall (R), AuC and F₁ score. We provide results when we solve ODE using the fixed-step (F) Euler method with 4 steps and the variable-step (V) Dopri5 method (atol=1e-1, rtol=1e-2) from the [Zhuang et al., 2021] library. We use the Hutchinson stochastic estimator of the Jacobian matrix trace [Hutchinson, 1990].

Quantitative results. We compare flow models to other popular baselines: OmniAnomaly [Su et al., 2019], CAE-M [Zhang et al., 2021], TranAD [Tuli et al., 2022]. It is common in these baselines to train and evaluate a separate model for each SMAP entity. In contrast, all our flow models use a single model for all entities in Table 2 i.e. they are entity-unconditional.

We can derive several important conclusions from Table 2 results. First, continuous-time normalizing flows models, if properly trained and evaluated, are able to outperform discrete-time normalizing flows as well as other non-flow models in this density estimation task. Second, recent integration-free FM training methods using (8) perform similarly or better than the CNF trained by computationally-expensive

MLE from (5). Third, the proposed DFM significantly outperforms prior FM methods with only the 2× complexity increase. In particular, DFM increases the non-saturated metrics such as precision and F₁ score by, correspondingly, 6.5 (88.2% → 94.7%) and 3.1 (93.3% → 96.4%) percentage points.

5 Conclusions

In this paper, we analyzed limitations of the interpolation-based flow matching framework that allows to efficiently train a CNF model. To address the limitations, we proposed the interpolant-free dual flow matching method. Our experiments with the SMAP benchmark showed that our DFM achieves state-of-the-art results for the entity-unconditional unsupervised anomaly detection. In future, the DFM objective (11) and the practical loss (12) can further be extended to multivariate case.

Table 2: SMAP unsupervised anomaly detection. The **best** and the second best metrics, %.

| Model | \int | P | R | AUC | F ₁ |
|------------|--------------|-------------|-------------|-------------|----------------|
| OmniAnom. | X | 81.3 | 94.2 | 98.9 | 87.3 |
| CAE-M | X | 81.9 | 95.7 | 99.0 | 88.3 |
| TranAD | X | 80.4 | 99.9 | 99.2 | 89.2 |
| Glow DNF | X | 87.4 | 84.9 | 91.6 | 86.1 |
| Base CNF | F | 87.5 | 98.8 | 98.4 | 92.8 |
| FM | F | <u>88.2</u> | <u>98.9</u> | 98.5 | <u>93.3</u> |
| I-CFM | F | 88.0 | 99.2 | <u>98.6</u> | <u>93.3</u> |
| DFM (ours) | F | 94.7 | 98.1 | 98.7 | 96.4 |
| Base CNF | V | 86.5 | 91.9 | 94.9 | 89.1 |
| FM | V | 87.4 | 99.6 | 98.7 | 93.1 |
| I-CFM | V | <u>89.3</u> | 98.2 | 98.3 | <u>93.6</u> |
| DFM (ours) | V | 89.7 | <u>98.9</u> | <u>98.6</u> | 94.1 |

References

- Lars Ruthotto and Eldad Haber. An introduction to deep generative modeling. *GAMM-Mitteilungen*, 2021.
- Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *NeurIPS*, 2018.
- Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *ICLR*, 2023.
- Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *NeurIPS*, 2019.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.
- Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguette, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.
- Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *ICLR*, 2023.
- Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *ICLR*, 2023.
- Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, and David Duvenaud. Scalable reversible generative models with free-form continuous dynamics. In *ICLR*, 2019.
- Juntang Zhuang, Nicha Dvornek, Xiaoxiao Li, Sekhar Tatikonda, Xenophon Papademetris, and James Duncan. Adaptive checkpoint adjoint method for gradient estimation in neural ODE. In *ICML*, 2020.
- Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *ICLR*, 2024.
- Kacper Kapusniak, Peter Potapchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Avishek Joey Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. *NeurIPS*, 2024.
- Hannes Stark, Bowen Jing, Chenyu Wang, Gabriele Corso, Bonnie Berger, Regina Barzilay, and Tommi Jaakkola. Dirichlet flow matching with applications to DNA sequence design. In *ICML*, 2024.
- Chaoran Cheng, Jiahua Li, Jian Peng, and Ge Liu. Categorical flow matching on statistical manifolds. *NeurIPS*, 2024.
- Andrew Campbell, Jason Yim, Regina Barzilay, Tom Rainforth, and Tommi Jaakkola. Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. In *ICML*, 2024.
- Ian Dunn and David Ryan Koes. Mixed continuous and categorical flow matching for 3D De Novo molecule generation. *arXiv:2404.19739*, 2024.
- George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *JMLR*, 2021.
- Amir Gholami, Kurt Keutzer, and George Biros. ANODE: Unconditionally accurate memory-efficient gradients for neural odes. In *IJCAI*, 2019.
- Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In *SIGKDD*, 2018.
- Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *SIGKDD*, 2019.

- Shreshth Tuli, Giuliano Casale, and Nicholas R. Jennings. TranAD: deep transformer networks for anomaly detection in multivariate time series data. *VLDB*, 2022.
- Diederik P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1×1 convolutions. In *NeurIPS*, 2018.
- Denis Gudovskiy, Tomoyuki Okuno, and Yohei Nakata. ContextFlow++: Generalist-specialist flow-based generative models with mixed-variable context encoding. In *UAI*, 2024.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.
- Juntang Zhuang, Nicha C Dvornek, sekhar tatikonda, and James s Duncan. MALI: A memory efficient and reverse accurate integrator for neural ODEs. In *ICLR*, 2021.
- M.F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics - Simulation and Computation*, 1990.
- Yuxin Zhang, Yiqiang Chen, Jindong Wang, and Zhiwen Pan. Unsupervised deep anomaly detection for multi-sensor time-series signals. *IEEE TKDE*, 2021.