Velocity Adaptation for Flow-Matching Models

Anonymous Authors[1]

Abstract

Flow-Matching models generate samples by learning a continuous-time velocity field that transports an initial distribution to a target data distribution. Although these models achieve strong empirical performance, the learned dynamics often exhibit suboptimal structure that can degrade sample quality as measured by Fréchet Inception Distance (FID). We propose a relatively lightweight post-training velocity adaptation framework for pre-trained Flow-Matching models. The adaptation leverages the Helmholtz–Hodge decomposition to decompose the learned velocity field into its conservative and solenoidal components. We introduce a fully convolutional scalar potential network (ϕ_{NET}, $\sim$3M parameters), whose gradient defines the conservative component. The adapted velocity is obtained by reweighting the conservative and solenoidal components during inference. Across CIFAR-10, ImageNet-64, ImageNet-32, CelebA-128, and AFHQ datasets, the proposed adaptation consistently improves generative quality and substantially reduces Fréchet Inception Distance (FID) compared to the corresponding pretrained Flow-Matching models. The adaptation also has strong cross-domain transferability: a ϕ_{NET} trained on one dataset effectively adapts velocity fields trained across different datasets and image resolutions. Our results suggest that structured decompositions of learned velocity fields provide an effective mechanism for improving sampling in continuous-time generative models.

1. Introduction

Generative modeling aims to learn a mapping from an initial distribution to a target data distribution. Prominent approaches include Generative Adversarial Networks (GANs) (Goodfellow et al., 2014; Nowozin et al., 2016), Variational Autoencoders (VAEs) (Kingma & Welling, 2022), Normalizing Flows (Rezende & Mohamed, 2016), Diffusion Models (Song & Ermon, 2020; Ho et al., 2020; Sohl-Dickstein et al., 2015), and Flow-Matching (Lipman et al., 2023; Albergo et al., 2025; Tong et al., 2024b). Among these, Flow-Matching has emerged as a scalable continuous-time framework that learns a time-dependent velocity field $\boldsymbol{v}_t(\boldsymbol{x}_t)$ governing the ODE $\dot{\boldsymbol{x}}_t = \boldsymbol{v}_t(\boldsymbol{x}_t)$. By directly regressing the network against a target velocity field, Flow-Matching avoids the expensive likelihood estimation required in standard continuous normalizing flows. Despite strong empirical performance, learnt velocity fields often exhibit rotational structure that does not contribute to the target probability flow and can degrade sampling quality.

In this work, we introduce a relatively lightweight post-training velocity adaptor (referred to as **Velociraptor**) for Flow-Matching models. Our approach leverages the Helmholtz–Hodge decomposition (von Helmholtz, 1858; Bhatia et al., 2012) to decompose the learnt velocity field into its conservative and solenoidal components. We estimate the conservative component using a fully-convolutional scalar potential network (ϕ_{NET}) and construct an adapted velocity field by reweighting the two components during sampling. Velociraptor consistently improves sample quality quantified by Fréchet Inception Distance (FID) (Heusel et al., 2018). Furthermore, the learnt adaptor exhibits strong transferability across datasets and image resolutions, suggesting that the conservative component captures structural regularities shared across generative dynamics.

2. Related Work

Structure-Aware Flow-Matching. Recent work has identified limitations of Flow-Matching objectives in accurately recovering probability paths. Huang et al. (Huang et al., 2026) show that matching the velocity field alone is insufficient, and propose augmenting the objective with divergence regularization. In contrast, our approach does not modify the training objective and instead operates post hoc on pre-trained velocity fields.

Helmholtz Decomposition in Neural Flows. Qi et al. (Qi et al., 2024) incorporate Helmholtz decomposition within

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the FoGen Workshop at ICML 2026. Do not distribute.

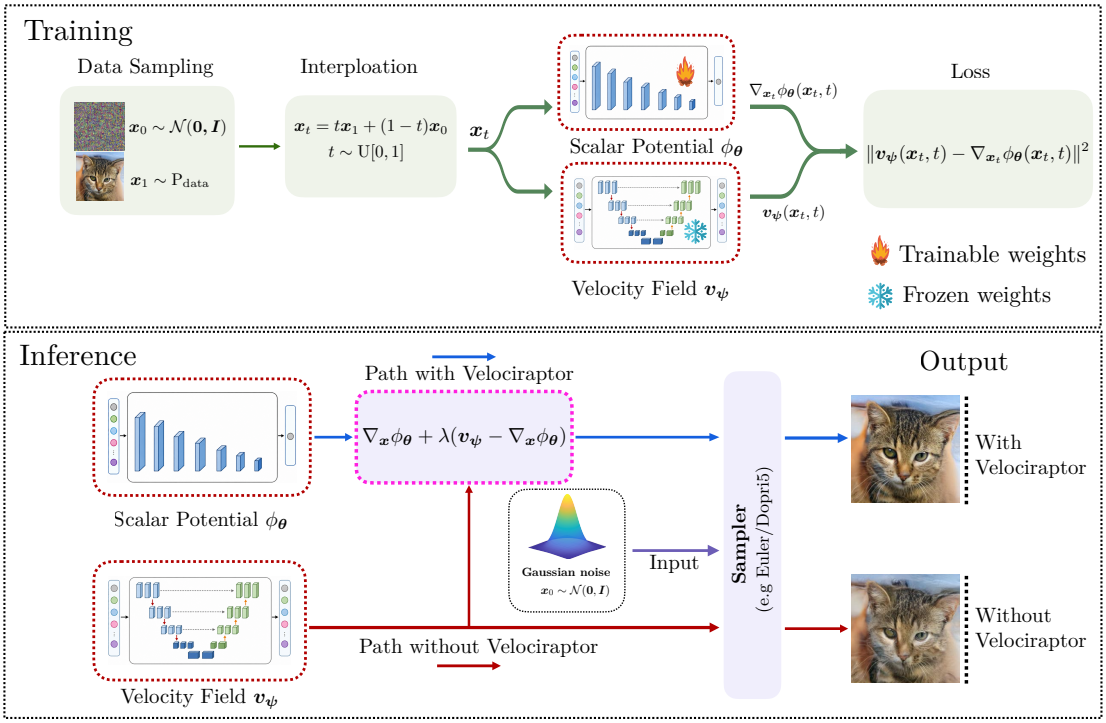

Figure 1. Overview of **Velociraptor**, the proposed velocity adaptor. **Top (Training):** A scalar potential network ϕ_{θ} is trained to approximate the conservative component of a pretrained Flow-Matching velocity field v_{ψ} by minimizing $\|v_{\psi} - \nabla_{x}\phi_{\theta}\|^2$. **Bottom (Inference):** The adapted velocity is constructed as $\nabla_{x}\phi_{\theta} + \lambda(v_{\psi} - \nabla_{x}\phi_{\theta})$ and used within a sampler (e.g., Euler or DOPRI5) to generate samples starting from Gaussian noise. The red arrows denote sampling with the original velocity field, while the blue arrows indicate the adapted trajectory. The adapted dynamics produce sharper and more coherent samples compared to the baseline.

a supervised framework using neural network for 2D flow reconstruction, learning conservative and solenoidal components. Unlike HDNet, our method operates in high-dimensional generative settings and learns the conservative component directly from pretrained Flow-Matching dynamics without access to the ground-truth decomposition.

Decomposition in Continuous-Time Generative Models. Horvat and Pfister (Horvat & Pfister, 2024) analyze diffusion vector fields through conservative and orthogonal decompositions to study density estimation and intrinsic dimensionality. In contrast, our approach using the Helmholtz–Hodge decomposition is a practical post-training adaptation mechanism for modifying pretrained generative dynamics and improving sample quality.

3. Background

Helmholtz–Hodge Decomposition. A smooth vector field $v : \mathbb{R}^n \to \mathbb{R}^n$ can be decomposed into conservative and solenoidal components (von Helmholtz, 1858; Weyl, 1940; Bhatia et al., 2012): $v(x) = \nabla_x \phi(x) + u(x)$, where ϕ is a scalar potential and u satisfies $\nabla_x \cdot u = 0$. The gradient component governs global transport, while the solenoidal component corresponds to localized rotational dynamics.

Flow-Matching. Flow-Matching (Lipman et al., 2023) learns a time-dependent velocity field $v_{\theta}(x, t)$ transporting an initial distribution p_0 to a target distribution p_1. Flow-Matching minimizes the following objective:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{\substack{t\sim\mathcal{U}[0,1] \\ x_0\sim p_0 \\ x_1\sim p_1}} \left[\|v_{\theta}(x_t, t) - v_t(x_t \mid x_0, x_1)\|_2^2 \right]. \tag{1}$$

For Optimal Transport Flow-Matching (Tong et al., 2024b), the interpolation is linear: $x_t = (1-t)x_0 + tx_1$, yielding the conditional velocity $v_t(x_t \mid x_0, x_1) = x_1 - x_0$

Velocity Non-Uniqueness. The density evolution induced by velocity field v follows the continuity equation (Benamou & Brenier, 2000): $\frac{\partial p_t}{\partial t} + \nabla_x \cdot (p_t v) = 0$. This equation does not uniquely determine the velocity field. If v satisfies the continuity equation, then any field $v' = v + u$ produces the same marginal dynamics whenever $\nabla_x \cdot (p_t u) = 0$. Thus, divergence-free solenoidal components can alter trajectories. In practice, finite model capacity and optimization error can lead Flow-Matching models to learn velocity fields containing solenoidal rotational structure. Velociraptor exploits this non-uniqueness by decomposing and reweighting the conservative and solenoidal components of pretrained velocity fields.

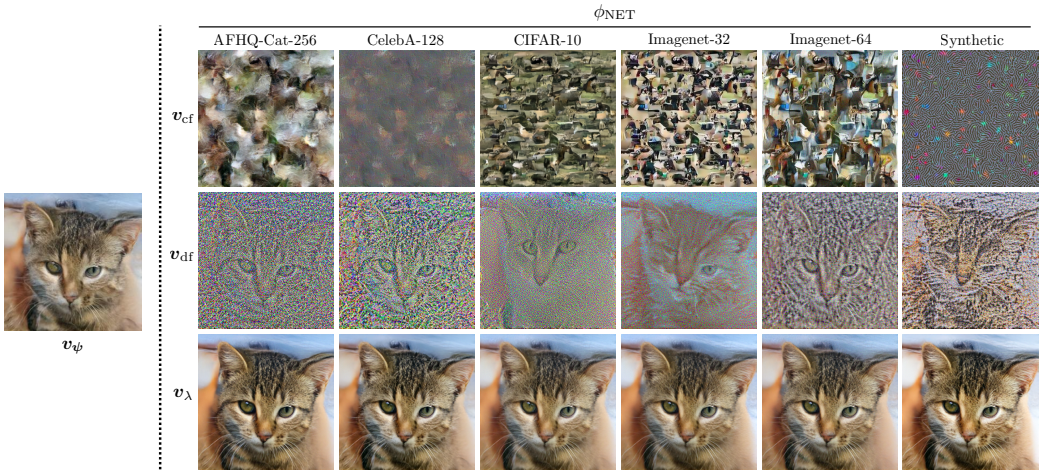

Figure 2. Qualitative comparison of generated images using different velocity components. (a) $\boldsymbol{v}_{\boldsymbol{\psi}}$: baseline velocity field trained on AFHQ-Cat-256; (b) $\boldsymbol{v}_{\text{cf}}$: conservative component; (c) $\boldsymbol{v}_{\text{df}}$: solenoidal component; (d) $\boldsymbol{v}_\lambda$: adapted velocity field **(Velociraptor)**. The right panel shows samples generated using a velocity model trained on AFHQ-Cat with ϕ_{NET} trained independently on different datasets. The adapted velocity produces sharper and more coherent samples, and generalizes across datasets.

Algorithm 1 Training the ϕ_{NET} Scalar Potential

Require: Pretrained velocity field $\boldsymbol{v}_{\boldsymbol{\psi}}$, target data distribution p_1, initial distribution p_0
Ensure: $\boldsymbol{\theta}^\star$
Initialize $\boldsymbol{\theta} \in \mathbb{R}^d$
while not converged **do**
  $\boldsymbol{x}_0 \sim p_0, \boldsymbol{x}_1 \sim p_1, \quad t \sim \mathcal{U}(0,1)$
  $\boldsymbol{x}_t \coloneqq (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$
  $\boldsymbol{v}_t \coloneqq \boldsymbol{v}_{\boldsymbol{\psi}}(\boldsymbol{x}_t, t)$
  $\mathcal{L}(\boldsymbol{\theta}) \coloneqq \|\boldsymbol{v}_t - \nabla_{\boldsymbol{x}_t}\phi_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\|_2^2$
  $\boldsymbol{\theta} \leftarrow \text{OptimizerStep}(\boldsymbol{\theta}, \nabla_{\boldsymbol{\theta}}\mathcal{L})$
end while
return $\boldsymbol{\theta}^\star$ (optimized $\boldsymbol{\theta}$)

Algorithm 2 Sampling with Velociraptor

Require: Pretrained velocity field $\boldsymbol{v}_{\boldsymbol{\psi}}$, optimized ϕ_{NET}, scale factor λ, ODE solver
Ensure: $\boldsymbol{x}_1$
$\boldsymbol{x}_0 \sim p_0$
$\boldsymbol{v}_{\text{cf}}(\boldsymbol{x}, t) \coloneqq \nabla_{\boldsymbol{x}}\phi_{\text{NET}}(\boldsymbol{x}, t)$
$\boldsymbol{v}_{\text{df}}(\boldsymbol{x}, t) \coloneqq \boldsymbol{v}_{\boldsymbol{\psi}}(\boldsymbol{x}, t) - \boldsymbol{v}_{\text{cf}}(\boldsymbol{x}, t)$
$\boldsymbol{v}_\lambda(\boldsymbol{x}, t) \coloneqq \boldsymbol{v}_{\text{cf}}(\boldsymbol{x}, t) + \lambda\, \boldsymbol{v}_{\text{df}}(\boldsymbol{x}, t)$
$\boldsymbol{x}_1 \coloneqq \text{ODESolver}(\boldsymbol{v}_\lambda, \boldsymbol{x}_0, t \in [0,1])$ ▷ DOPRI5, Euler
return $\boldsymbol{x}_1$

4. Methodology

4.1. Extraction of the Conservative Field Component

We begin by isolating the conservative field component of a pretrained Flow-Matching velocity field $\boldsymbol{v}_{\boldsymbol{\psi}}$ by introducing a time-dependent scalar potential $\phi_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$ parameterized by a lightweight neural network θ. This potential is optimized to approximate the conservative field component of $\boldsymbol{v}_{\boldsymbol{\psi}}$ via the L^2 minimization objective:

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},t}\left[\|\boldsymbol{v}_{\boldsymbol{\psi}}(\boldsymbol{x}, t) - \nabla_{\boldsymbol{x}}\phi_{\boldsymbol{\theta}}(\boldsymbol{x}, t)\|_2^2\right] \tag{2}$$

This objective encourages the gradient field $\nabla_{\boldsymbol{x}}\phi_{\boldsymbol{\theta}}$ to approximate the conservative component of the pretrained velocity field $\boldsymbol{v}_{\boldsymbol{\psi}}$. The optimization steps are provided in Algorithm 1. We refer to the estimated field as $\boldsymbol{v}_{\text{cf}} = \nabla_{\boldsymbol{x}}\phi_{\boldsymbol{\theta}^\star}$ and $\phi_{\boldsymbol{\theta}^\star}(\boldsymbol{x}, t)$ as $\phi_{\text{NET}}(\boldsymbol{x}, t)$.

4.2. Velocity Adaptation by Solenoidal Scaling

The solenoidal residue is given by $\boldsymbol{v}_{\text{df}} = \boldsymbol{v}_{\boldsymbol{\psi}} - \boldsymbol{v}_{\text{cf}}$. We construct the adapted velocity field $\boldsymbol{v}_\lambda$ as follows:

$$\boldsymbol{v}_\lambda(\boldsymbol{x}, t) = \boldsymbol{v}_{\text{cf}}(\boldsymbol{x}, t) + \lambda \boldsymbol{v}_{\text{df}}(\boldsymbol{x}, t) \tag{3}$$

where λ is a scalar hyperparameter. $\lambda = 1$ recovers the base pretrained velocity field $\boldsymbol{v}_{\boldsymbol{\psi}}$, while $\lambda > 1$ recalibrates the rotational dynamics. The Velociraptor pipeline is shown in Figure 1.

5. Experimental Results

We evaluate **Velociraptor** across several datasets and resolutions, including CIFAR-10 (Krizhevsky et al., 2009), CelebA (128×128) (Liu et al., 2015), AFHQ-Cat (256×256) (Choi et al., 2020), and downsampled ImageNet at 32×32 and 64×64 (Russakovsky et al., 2015). The sampling algorithm is as given in Algorithm 2.

Dataset	Model	FID ↓	KID $\times 10^{-3}$ ↓	Avg-NFE ↓	Precision ↑	Recall ↑	F1-Score ↑	Coverage ↑
ImageNet-32	v_ψ (Baseline)	7.63	3.50 ± 0.62	89.60	**0.874**	0.706	0.781	0.902
	v_λ (**Velociraptor**)	**4.94**	$\mathbf{1.21 \pm 0.32}$	**80.10**	0.843	**0.783**	**0.812**	**0.916**
ImageNet-64	v_ψ (Baseline)	17.30	10.68 ± 1.32	108.08	**0.875**	0.703	0.780	0.739
	v_λ (**Velociraptor**)	**10.77**	$\mathbf{5.04 \pm 0.79}$	**84.26**	0.837	**0.780**	**0.808**	**0.770**
AFHQ-Cat-256	v_ψ (Baseline)	15.70	6.87 ± 0.49	88.82	0.737	0.621	0.674	0.787
	v_λ (**Velociraptor**)	**11.74**	$\mathbf{4.50 \pm 0.57}$	**86.09**	**0.816**	**0.652**	**0.725**	**0.881**
CelebA-128	v_ψ (Baseline)	17.46	7.93 ± 0.67	93.26	0.827	0.613	0.704	0.776
	v_λ (**Velociraptor**)	**13.55**	$\mathbf{3.67 \pm 0.34}$	**91.58**	**0.836**	**0.639**	**0.724**	**0.830**
CIFAR-10	v_ψ (TorchCFM)	3.57	1.64 ± 0.46	**130.40**	**0.782**	0.654	0.712	0.828
	v_λ (**Velociraptor**)	**3.04**	$\mathbf{0.84 \pm 0.26}$	144.20	0.778	**0.674**	**0.722**	**0.832**
CIFAR-10	v_ψ (Rectified Flow++)	3.08	1.06 ± 0.30	1	**0.775**	0.672	0.720	**0.831**
	v_λ (**Velociraptor**)	**2.69**	$\mathbf{0.71 \pm 0.24}$	1	0.766	**0.692**	**0.727**	0.830

Table 1. Quantitative comparison between the baseline velocity v_ψ and the adapted velocity v_λ (**Velociraptor**). For CelebA-128 and AFHQ-Cat-256, the baseline models are taken from the PnP flow (Martin et al., 2025) paper. Additionally, for CIFAR-10, we report results on publicly available implementations, including TorchCFM (OT-CFM) (Tong et al., 2024a;c) and Rectified-Flow++ (Lee et al., 2024). Metrics are evaluated on five benchmark datasets using DoPri5 adaptive ODE solver.

Baselines: We use pretrained Flow-Matching models from PnP-Flow (Martin et al., 2025) for CelebA-128 and AFHQ-Cat-256, and train ImageNet-32 and ImageNet-64 models from scratch. For CIFAR-10, we leverage publicly available implementations including TorchCFM (OT-CFM) (Tong et al., 2024a;b) and Rectified-Flow++ (Lee et al., 2024).

Synthetic Dataset: We synthesize a dataset to study the effect of data diversity and multimodality on the transferability of ϕ_{NET}. The design is motivated by the observation that ϕ_{NET} trained on structured datasets with limited variability (e.g., CelebA-128) generalizes poorly to diverse datasets such as CIFAR-10 and ImageNet. We seek a controlled dataset that exhibits strong intraclass variation, multimodal structure, and smooth manifold geometry. The synthetic dataset consists of 32×32 RGB images organized into 100 balanced classes, each parameterized by a configuration controlling color statistics, spatial layouts, texture frequencies, and blob composition. For diversity, the classes are assigned distinct combinations of geometric patterns, frequency ranges, and chromatic distributions.

Metrics: We report Fréchet Inception Distance (FID) (Heusel et al., 2018), Kernel Inception Distance (KID) (Bińkowski et al., 2021), Precision, Recall, F1-score (Kynkäänniemi et al., 2019), and Coverage (Naeem et al., 2020).

Training and Sampling: ϕ_{NET} is trained using AdamW (Loshchilov & Hutter, 2019) with learning rate 2×10^{-4} for 20 epochs. The scale parameter λ is selected on a validation set. For sampling, we use Euler (Griffiths & Higham, 2010) and DOPRI5 (Dormand & Prince, 1980) ODE solvers, with atol and rtol of 10^{-3} for DOPRI5.

5.1. Quantitative Results

We compare the baseline velocity v_ψ and the adapted velocity v_λ across multiple datasets in Table 1. Velociraptor consistently improves FID across datasets, including reducing CIFAR-10 FID from 3.57 to 3.04 on the publicly available TorchCFM model, and ImageNet-64 FID from 17.30 to 10.77. We additionally observe improved Recall, F1-score, and Coverage, indicating better distributional alignment and diversity. Although Precision decreases slightly compared to the baseline, the overall gain in F1-score suggests a more favorable balance between sample fidelity and diversity. In most cases, these improvements are achieved with similar or lower NFE, though with additional computational overhead due to gradient evaluation of ϕ_{NET}. The quantitative results are consistent with the qualitative comparisons in Figure 2, where the adapted velocity produces sharper and more coherent samples.

5.2. Cross-Dataset/Model Generalization or Transferability

We evaluate the transferability of the learnt conservative component v_{cf}, parameterized by ϕ_{NET}, across pretrained Flow-Matching velocity fields trained on different datasets. In this setup, v_ψ denotes the pretrained baseline velocity model (rows in Table 2), while ϕ_{NET} is trained independently on a given dataset (columns) using Algorithm 1. At inference, the adapted velocity field is constructed using Equation (3).

As shown in Table 2, the learnt adaptation generalizes effectively across datasets. For example, applying a CIFAR-10 trained ϕ_{NET} to an ImageNet-32 baseline model improves

Base Model	Baseline (FID)	Adaptor Domain ϕ_{NET} (FID ↓, $\lambda^\star$)					
Domain (v_ψ)	(No Adaptation)	CIFAR-10	ImageNet-32	ImageNet-64	CelebA-128	AFHQ-Cat-256	Synthetic
CIFAR-10 (TorchCFM)	3.57	**3.04** (1.125)	3.11 (1.05)	3.30(1.05)	3.71(1.025)	3.60(1.025)	3.45(1.025)
ImageNet-32	7.63	6.77 (1.10)	**4.94** (1.40)	7.30 (1.05)	7.33 (1.025)	7.31 (1.025)	7.13 (1.05)
ImageNet-64	17.30	13.55 (1.10)	13.00 (1.10)	**10.77** (1.25)	16.09 (1.05)	14.85 (1.10)	13.63 (1.15)
CelebA-128	17.46	13.30 (1.10)	**12.53** (1.075)	12.60 (1.10)	13.55 (1.075)	13.58 (1.05)	14.30 (1.05)
AFHQ-Cat-256	15.7	13.63 (1.05)	14.28 (1.075)	14.12 (1.05)	14.17 (1.05)	**11.74** (1.20)	14.55 (1.10)

Table 2. Cross-dataset generalization of **Velociraptor**. Rows denote the dataset used to train the baseline velocity field v_ψ, while columns indicate the training domain of ϕ_{NET}. Each entry reports the FID score (lower is better), with the optimal scaling parameter λ^* shown in parentheses. Diagonal entries correspond to in-domain evaluation, while off-diagonal entries reflect cross-domain transfer. For each row, the best results are highlighted in **bold** and the second-best results are underlined.

the FID from 7.63 to 6.77. Similarly, for a CIFAR-10 baseline model, an ImageNet-32 trained ϕ_{NET} reduces the FID from 3.57 to 3.11. These results suggest that the learnt conservative component captures structural properties of the velocity field that transfer across datasets.

However, we observe weaker transferability for ϕ_{NET} trained on CelebA-128 when applied to highly multimodal datasets such as CIFAR-10 and ImageNet variants. Compared to adaptors trained on CIFAR-10 or ImageNet variants, CelebA-128 trained ϕ_{NET} consistently yields smaller improvements in cross-domain settings.

We hypothesize that this behavior arises from differences in data diversity. CIFAR-10 and ImageNet exhibit substantial intra-class variation and multimodal structure, whereas CelebA-128 primarily contains aligned face images with comparatively limited variability. As a result, ϕ_{NET} trained on CelebA-128 learns adaptations specialized to a narrower distribution and generalizes less effectively to broader multimodal datasets.

As shown in Figure 2, samples generated using only v_{cf} or v_{df} exhibit oversmoothed or noisy artifacts, whereas the baseline v_ψ and adapted velocity v_λ produce coherent samples.

To further evaluate whether the proposed adaptation memorizes training data, we perform a nearest-neighbour analysis in feature space. For each generated image, we retrieve the $k = 5$ nearest training samples using Inception-V3 (Szegedy et al., 2016) features and Euclidean distance. As shown in Figure 3 retrieved neighbours are visually similar but not identical to generated images, suggesting that Velociraptor captures the underlying data distribution rather than memorizing training samples.

5.3. Effect of Solenoidal Scaling

We investigate the role of the scaling parameter λ, which balances the conservative and solenoidal components in the adapted velocity field (Equation (3)). Figure 4 shows the effect of varying λ on FID across multiple datasets. Across all datasets, we observe a consistent U-shaped trend: moderate amplification of the solenodial component (typically around $\lambda \approx 1.2$) significantly improves sample quality, whereas both under-scaling and over-scaling degrade performance.

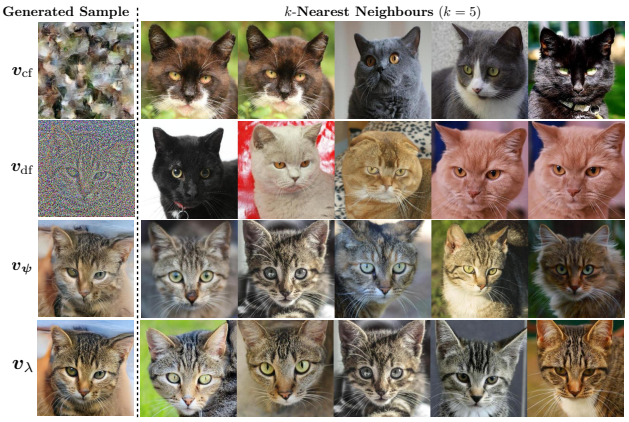

Figure 3. Nearest-neighbour analysis of generated samples. For each generated image (left), we retrieve its $k = 5$ nearest neighbours from the training set using Inception-V3 (Szegedy et al., 2016) features and Euclidean distance in feature space. Rows correspond to samples generated using v_{cf}, v_{df}, v_ψ, and **Velociraptor** v_λ.

This behavior can be understood from the distinctive roles played by the two components. Using only the conservative field (v_{cf}), produces globally coherent/smooth structure, but without any details pertaining to the identity of the sample. The gradient field enforces globally consistent transport but does not capture localized variations. The solenoidal component (v_{df}) primarily encodes high-frequency, localized spatial variations and contains the identity of the sample. When used in isolation, v_{df} leads to noisy samples, indicating that it does not provide a suitable global transport mechanism. Combining the two components optimally through v_λ yields the best qualitative and quantitative performance.

Velocity Components	Dataset (FID)				
	CIFAR-10	ImageNet-32	ImageNet-64	CelebA-128	AFHQ-Cat-256
$\boldsymbol{v}_\psi$ (Velocity)	3.57	7.63	17.30	17.46	15.70
$\boldsymbol{v}_{\mathrm{cf}}$ (Conservative Field)	369.51	23.68	80.88	305.86	265.90
$\boldsymbol{v}_{\mathrm{df}}$ (Solenoidal Field)	369.24	423.76	395.37	371.70	392.63
$\boldsymbol{v}_\lambda$ (Velociraptor)	**3.04**	**4.94**	**10.77**	**13.55**	**11.74**

Table 3. Quantitative evaluation of generative performance across datasets using (FID). We compare the baseline velocity field $\boldsymbol{v}_\psi$ with its decomposed components: the conservative (irrotational) field $\boldsymbol{v}_{\mathrm{cf}}$ and the solenoidal field $\boldsymbol{v}_{\mathrm{df}}$.

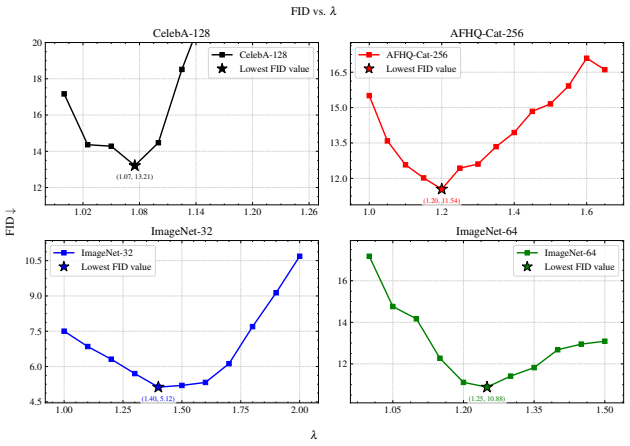

Figure 4. FID as a function of the scaling parameter λ across datasets. Each plot shows the effect of varying λ on sample quality, with the optimal value marked by a star. Across all datasets, performance exhibits a U-shaped trend: moderate scaling of the solenodial component improves FID, while under-scaling or over-scaling degrades performance.

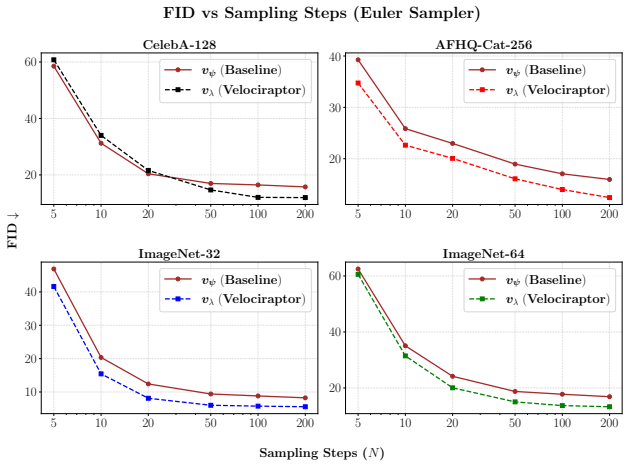

Figure 5. FID as a function of the number of sampling steps across datasets using the Euler sampler. Across all datasets, $\boldsymbol{v}_\lambda$ consistently achieves lower FID than the baseline $\boldsymbol{v}_\psi$, particularly in the low-step regime.

For datasets such as AFHQ-Cat, the solenodial component recovers fine textures, fur details and high-contrast facial details, without disrupting the overall structure, as illustrated in Figure 6. This qualitative behavior aligns with the quantitative ablations as shown in Table 3, where the combined velocity field consistently achieves the best FID. Overall, the results suggest that the solenoidal acts as a refinement mechanism that improves local structure while maintaining globally coherent transport.

5.4. FID versus Sampling Steps

We analyze the effect of velocity adaptation under different sampling budgets using the Euler sampler. Figure 5 reports FID as a function of the number of sampling steps for CelebA-128, AFHQ-Cat-256, ImageNet-32, and ImageNet-64 datasets. Across all datasets, **Velociraptor** consistently achieves lower FID than the baseline velocity field. For instance, on ImageNet-32 and ImageNet-64, the adapted velocity significantly improves sample quality as measured by FID even with 10–20 Euler steps. A similar trend is observed on AFHQ-Cat-256, where the gap between the baseline and adapted dynamics widens as the sampling budget increases.

These results suggest that the adapted velocity field induces smoother and more efficient trajectories, enabling improved generation quality even with fewer integration steps. Importantly, this behavior is consistent with the DOPRI5 results reported in Table 1, where **Velociraptor** achieves lower FID with comparable or reduced average NFE across datasets.

6. Training ϕ_{NET} Without Requiring a Pretrained Velocity Model

Recall from Equation (2) that we rely on a pretrained velocity field $\boldsymbol{v}_\psi$ to approximate the conservative field component. However, the dependence on a pretrained velocity field can be removed by directly using the data with Conditional Flow-Matching (Lipman et al., 2023). We sample $\boldsymbol{x}_0 \sim p_0$ from the initial distribution, and $\boldsymbol{x}_1 \sim p_1$ from the

Base Model	**Baseline (FID)**	**Adaptor Domain ϕ_{NET} (FID ↓, $\lambda^\star$)**					
Domain (v_ψ)	**(No Adaptation)**	CIFAR-10	ImageNet-32	ImageNet-64	CelebA-128	AFHQ-Cat	Synthetic
CIFAR-10 (TorchCFM)	3.57	3.30 (1.05)	**3.15** (1.075)	3.24 (1.05)	3.88 (1.025)	3.53 (1.025)	3.44 (1.05)
ImageNet-32	7.63	6.85 (1.05)	**5.90** (1.375)	7.04 (1.1)	7.48 (1.025)	7.27 (1.025)	7.52 (1.05)
ImageNet-64	17.30	14.45 (1.075)	13.38 (1.1)	**11.19** (1.275)	16.55 (1.025)	14.78 (1.1)	15.40 (1.1)
CelebA-128	17.46	**12.37** (1.075)	12.55 (1.075)	13.18 (1.125)	15.80 (1.025)	12.50 (1.075)	13.63 (1.05)
AFHQ-Cat-256	15.70	13.95 (1.05)	14.58 (1.05)	13.96 (1.1)	14.72 (1.05)	**13.74** (1.15)	15.17 (1.05)

Table 4. Cross-dataset evaluation of **Velociraptor** without a pretrained velocity model. Rows denote the dataset used to train the baseline velocity field, while columns indicate the training domain of ϕ_{NET}. Each entry reports the FID score (lower is better), with the optimal scaling parameter λ^* shown in parentheses. Diagonal entries correspond to in-domain evaluation, while off-diagonal entries reflect cross-domain transfer. **Bold** and underline indicate the best and second-best performance per row, respectively. The results demonstrate that ϕ_{NET} can be trained independently and still generalize across datasets, enabling effective velocity adaptation without relying on a pretrained velocity field.

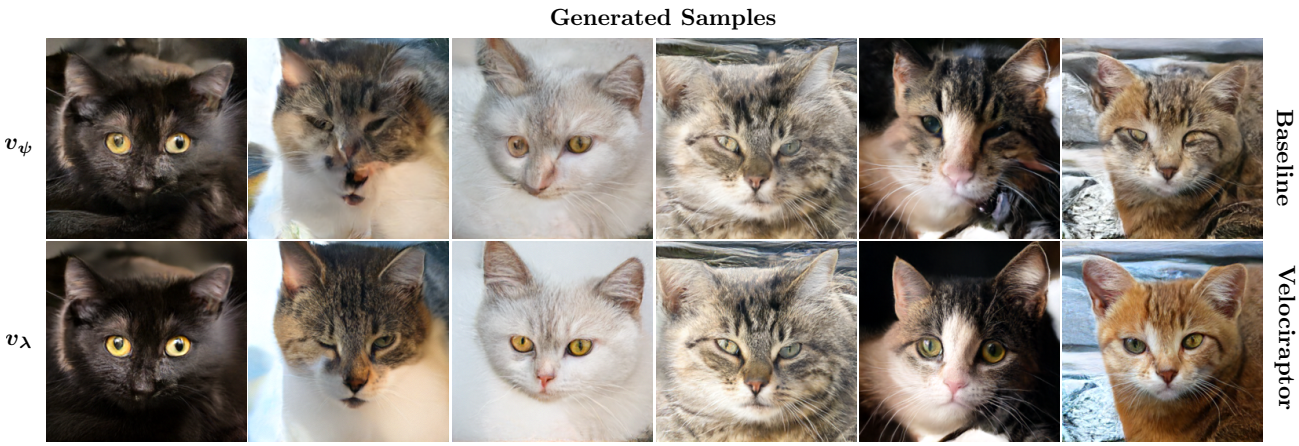

Figure 6. Qualitative comparison between samples generated using the baseline velocity v_ψ (top) and the adapted velocity v_λ (**Velociraptor**, bottom). The adapted dynamics consistently produce sharper facial structures, improved high-frequency details, and more coherent textures, while reducing the blur and structural artifacts present in the baseline samples.

Dataset	Model	FID ↓	KID $\times 10^{-3}$ ↓	Avg-NFE ↓	Precision ↑	Recall ↑	F1-Score ↑	Coverage ↑
ImageNet-32	v_ψ (Baseline)	7.63	3.50 ± 0.62	89.60	**0.874**	0.706	0.781	**0.902**
	v_λ (**Velociraptor**)	**5.90**	$\mathbf{2.31 \pm 0.51}$	**84.60**	0.823	**0.793**	**0.808**	0.898
ImageNet-64	v_ψ (Baseline)	17.30	10.68 ± 1.32	108.08	**0.875**	0.703	0.780	0.739
	v_λ (**Velociraptor**)	**11.19**	$\mathbf{5.00 \pm 0.70}$	**70.98**	0.840	**0.775**	**0.807**	**0.765**
AFHQ-Cat-256	v_ψ (Baseline)	15.70	6.87 ± 0.49	88.82	0.737	0.621	0.674	**0.787**
	v_λ (**Velociraptor**)	**13.74**	$\mathbf{6.00 \pm 0.59}$	**86.43**	**0.791**	**0.671**	**0.726**	0.729
CelebA-128	v_ψ (Baseline)	17.46	7.93 ± 0.67	93.26	**0.827**	0.613	0.704	0.776
	v_λ (**Velociraptor**)	**15.80**	$\mathbf{5.45 \pm 0.56}$	**91.84**	0.825	**0.633**	**0.716**	**0.798**

Table 5. Quantitative comparison between the baseline velocity v_ψ and the adapted velocity v_λ when ϕ_{NET} is trained directly using the conditional flow matching velocity field, without relying on a pretrained velocity model. Metrics are evaluated using the DOPRI5 adaptive ODE solver across multiple datasets.

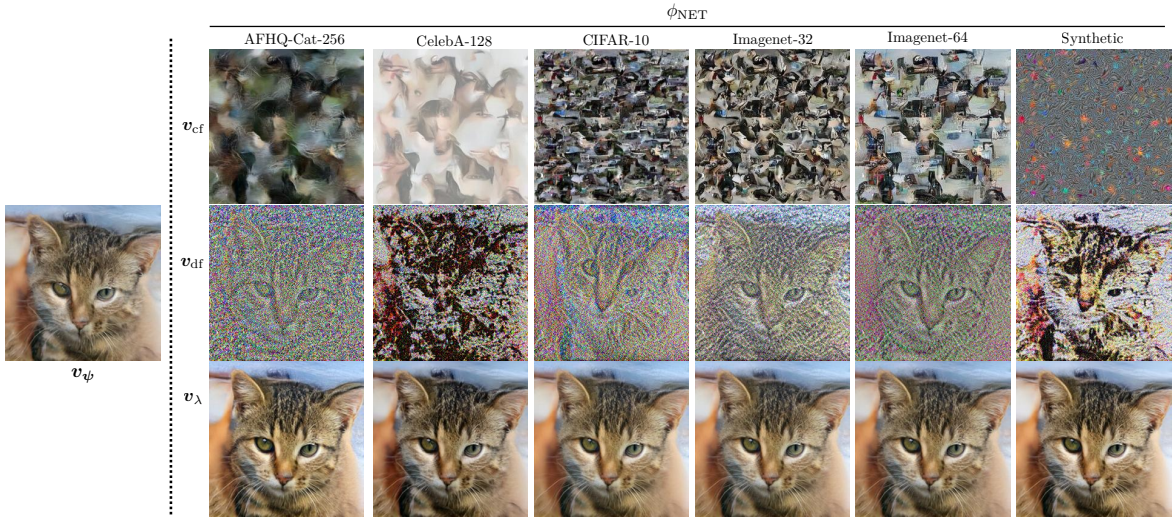

Figure 7. Qualitative comparison of velocity components. (a) $\boldsymbol{v}_{\psi}$: baseline velocity, (b) $\boldsymbol{v}_{\text{cf}}$: conservative component, (c) $\boldsymbol{v}_{\text{df}}$: solenoidal component, and (d) $\boldsymbol{v}_{\lambda}$: adapted velocity. The columns correspond to ϕ_{NET} trained independently on different datasets, without a pretrained velocity model. The conservative component captures global structure, while the solenoidal component encodes high-frequency details. Their combination yields sharper and coherent samples, demonstrating consistent cross-dataset generalization.

target data distribution, and construct the interpolation $\boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$. Instead of regressing against a pretrained velocity field, we directly match gradient of ϕ_{NET} to the conditional velocity field $\boldsymbol{v}_t(\boldsymbol{x}_t \mid \boldsymbol{x}_0, \boldsymbol{x}_1) = \boldsymbol{x}_1 - \boldsymbol{x}_0$ with the objective: $\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\substack{t\sim\mathcal{U}[0,1] \\ \boldsymbol{x}_0\sim p_0 \\ \boldsymbol{x}_1\sim p_1}} \left[\|\nabla_{\boldsymbol{x}} \phi_\theta(\boldsymbol{x}_t, t) - (\boldsymbol{x}_1 - \boldsymbol{x}_0)\|_2^2 \right]$. Minimizing this objective, following the steps in Algorithm 1, isolates the curl-free component of the target dynamics directly within the potential network. Table 5 reports quantitative comparisons between the baseline $\boldsymbol{v}_{\psi}$ and the adapted velocity $\boldsymbol{v}_{\lambda}$, across four datasets and Table 4 reports cross-dataset performance in this setting. **Velociraptor**, in the absence of a pretrained velocity field achieves competitive FID scores across all datasets. Notably, ϕ_{NET} trained on one dataset generalizes effectively to other datasets, maintaining strong performance in cross-domain settings. This suggests that the learnt scalar potential captures transferable structural information that is not tied to a specific data distribution. Qualitative results in Figure 7 support these findings. This experiment demonstrates that ϕ_{NET} can be trained directly on the data while retaining strong generalization capability and sample quality.

7. Computational Cost Analysis

We analyze the computational cost of using the Velociraptor $\boldsymbol{v}_{\lambda}$ compared to the baseline $\boldsymbol{v}_{\psi}$. As shown in Table 6, Velociraptor incurs additional runtime and memory overhead due to the gradient computation of the scalar potential. The additional cost is justified in view of the substantial improvements in sample quality as evidenced by FID scores.

Dataset	Model	Time (s)↓	Memory (MB)↓
CIFAR-10	$\boldsymbol{v}_{\psi}$	0.5717	781.97
	$\boldsymbol{v}_{\lambda}$	0.6896	807.87
ImageNet-64	$\boldsymbol{v}_{\psi}$	0.4312	4721.76
	$\boldsymbol{v}_{\lambda}$	0.5833	4738.51
CelebA-128	$\boldsymbol{v}_{\psi}$	1.4188	364.00
	$\boldsymbol{v}_{\lambda}$	1.8112	709.91
AFHQ-Cat-256	$\boldsymbol{v}_{\psi}$	1.2028	478.91
	$\boldsymbol{v}_{\lambda}$	3.0894	1806.88

Table 6. Inference cost comparison between the baseline flow-matching model $\boldsymbol{v}_{\psi}$ and the Velociraptor adaptor $\boldsymbol{v}_{\lambda}$. These metrics have been averaged over 100 samples using the DOPRI5 solver.

8. Conclusion

We introduced a velocity adaptation mechanism for adjusting the conservative and solenoidal components of the velocity field. We demonstrated that there is a trade-off between the two components and mildly rescaling the solenoidal component versus the curl-free component results in substantial improvements in image generation quality. The optimal combination could only be determined empirically. A theoretical approach to determine the optimal combination remains an open problem. An interesting aspect of this formulation is that velocity adaptation is transferable from one dataset to another, which opens up numerous possibilities from a generalization perspective.

Impact Statement

This work introduces a post-training velocity adaptation framework for improving the performance of continuous-time generative models. By leveraging structured decompositions of the learnt velocity fields, the proposed method improves sample quality and provides additional insights into the role of conservative and solenoidal dynamics in Flow-Matching models. This work contributes to a better understanding of the geometric structure underlying continuous-time generative modeling.

As with all generative modeling research, there are broader societal risks associated with synthetic media generation, including misinformation, impersonation, and deepfake content. Safeguards pertaining to responsible deployment remain important considerations for future applications.

References

Albergo, M. S., Boffi, N. M., and Vanden-Eijnden, E. Stochastic interpolants: A unifying framework for flows and diffusions, 2025. URL https://arxiv.org/abs/2303.08797.

Benamou, J.-D. and Brenier, Y. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

Bhatia, H., Norgard, G., Pascucci, V., and Bremer, P.-T. The helmholtz-hodge decomposition—a survey. *IEEE Transactions on visualization and computer graphics*, 19 (8):1386–1404, 2012.

Bińkowski, M., Sutherland, D. J., Arbel, M., and Gretton, A. Demystifying mmd gans, 2021. URL https://arxiv.org/abs/1801.01401.

Choi, Y., Uh, Y., Yoo, J., and Ha, J.-W. StarGAN v2: Diverse image synthesis for multiple domains. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8188–8197, 2020.

Dormand, J. R. and Prince, P. J. A family of embedded runge-kutta formulae. *Journal of Computational and Applied Mathematics*, 6(1):19–26, 1980. doi: 10.1016/0771-050X(80)90013-3.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial networks, 2014. URL https://arxiv.org/abs/1406.2661.

Griffiths, D. F. and Higham, D. J. *Euler's Method*, pp. 19–31. Springer London, London, 2010. ISBN 978-0-85729-148-6. doi: 10.1007/978-0-85729-148-6_2. URL https://doi.org/10.1007/978-0-85729-148-6_2.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium, 2018. URL https://arxiv.org/abs/1706.08500.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models, 2020. URL https://arxiv.org/abs/2006.11239.

Horvat, C. and Pfister, J.-P. On gauge freedom, conservativity and intrinsic dimensionality estimation in diffusion models, 2024. URL https://arxiv.org/abs/2402.03845.

Huang, Y., Transue, T., Wang, S.-H., Feldman, W., Zhang, H., and Wang, B. Improving flow matching by aligning flow divergence, 2026. URL https://arxiv.org/abs/2602.00869.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes, 2022. URL https://arxiv.org/abs/1312.6114.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. Technical report, Toronto, ON, Canada, 2009.

Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., and Aila, T. Improved precision and recall metric for assessing generative models, 2019. URL https://arxiv.org/abs/1904.06991.

Lee, S., Lin, Z., and Fanti, G. Improving the training of rectified flows. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=mSHs6C7Nfa.

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling, 2023. URL https://arxiv.org/abs/2210.02747.

Liu, Z., Luo, P., Wang, X., and Tang, X. Deep learning face attributes in the wild. *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 3730–3738, 2015. doi: 10.1109/iccv.2015.425.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization, 2019. URL https://arxiv.org/abs/1711.05101.

Martin, S., Gagneux, A., Hagemann, P., and Steidl, G. Pnp-flow: Plug-and-play image restoration with flow matching, 2025. URL https://arxiv.org/abs/2410.02423.

Naeem, M. F., Oh, S. J., Uh, Y., Choi, Y., and Yoo, J. Reliable fidelity and diversity metrics for generative models, 2020. URL https://arxiv.org/abs/2002.09797.

Nowozin, S., Cseke, B., and Tomioka, R. f-gan: Training generative neural samplers using variational divergence minimization. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/cedebb6e872f539bef8c3f919874e9d7-Paper.pdf.

Qi, M., Idoughi, R., and Heidrich, W. Hdnet: Physics-inspired neural network for flow estimation based on helmholtz decomposition. *ArXiv*, abs/2406.08570, 2024. URL https://api.semanticscholar.org/CorpusID:270440985.

Rezende, D. J. and Mohamed, S. Variational inference with normalizing flows, 2016. URL https://arxiv.org/abs/1505.05770.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics, 2015. URL https://arxiv.org/abs/1503.03585.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution, 2020. URL https://arxiv.org/abs/1907.05600.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

Tong, A., FATRAS, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., and Bengio, Y. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024a. ISSN 2835-8856. URL https://openreview.net/forum?id=CD9Snc73AW. Expert Certification.

Tong, A., Fatras, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., and Bengio, Y. Improving and generalizing flow-based generative models with minibatch optimal transport, 2024b. URL https://arxiv.org/abs/2302.00482.

Tong, A., Malkin, N., Fatras, K., Atanackovic, L., Zhang, Y., Huguet, G., Wolf, G., and Bengio, Y. Simulation-free schrödinger bridges via score and flow matching, 2024c. URL https://arxiv.org/abs/2307.03672.

von Helmholtz, H. Über integrale der hydrodynamischen gleichungen, welche den wirbelbewegungen entsprechen. *Journal für die reine und angewandte Mathematik*, 55: 25–55, 1858.

Weyl, H. The method of orthogonal projection in potential theory. *Duke Mathematical Journal*, 7(1):411–444, 1940.