

生成模型的流匹配

Yaron Lipman⁽¹⁾⁽²⁾ Ricky T. Q. Chen⁽¹⁾ Heli Ben-Hamu² Maximilian Nickel¹ Matt Le⁽¹⁾⁽¹⁾ Meta AI (FAIR)
²Weizmann Institute of Science

内容摘要

我们介绍了一种基于连续归一化流（CNFs）的生成建模新范式，使我们能够以前所未有的规模训练 CNFs。具体来说，我们提出了“流匹配”（FM）的概念，这是一种基于固定条件概率路径向量场回归训练 CNF 的免模拟方法。流量匹配与用于在噪声和数据样本之间进行转换的高斯概率路径的一般族兼容--它将现有的扩散路径归纳为具体实例。有趣的是，我们发现使用带有扩散路径的 FM 可以为扩散模型的训练提供更稳健、更稳定的替代方案。此外，流量匹配还为使用其他非扩散概率路径训练 CNF 打开了大门。其中一个特别值得关注的例子是使用最优传输（OT）位移插值来定义条件概率路径。这些路径比扩散路径更有效，训练和采样速度更快，泛化效果更好。在 ImageNet 上使用流匹配法训练 CNF，在可能性和样本质量方面的性能始终优于其他基于扩散的方法，并且可以使用现成的数值 ODE 求解器快速、可靠地生成样本。

1 简介

深度生成模型是一类深度学习算法，旨在对未知数据分布进行估计和采样。最近，生成模型取得了令人惊叹的进展，例如用于图像生成的 Ramesh 等人（2022 年）；Rombach 等人（2022 年），这主要得益于基于扩散的模型的可扩展和相对稳定的训练，Ho 等人（2020 年）；Song 等人（2020 年b）。然而，对简单扩散过程的限制导致采样概率路径的空间相当有限，导致训练时间很长，需要采用专门的方法（如 Song 等人（2020a）；Zhang & Chen（2022））进行高效采样。

在这项工作中，我们考虑了连续归一化流（CNFs；Chen 等人（2018））这一通用的确定性框架。CNFs 能够对任意概率路径建模，尤其是已知的由扩散模型建模的概率路径。

概率路径（Song 等人，2021 年）。然而，除了可以通过去噪分数匹配等方法高效训练扩散外（Vincent，2011 年），目前还没有可扩展的 CNF 训练算法。事实上，最大似然训练（如 Grathwohl 等人（2018 年））需要昂贵的数值 ODE 仿真，而现有的免仿真方法要么涉及不可控积分（Rozen 等 Ben-Hamu 等人，2021 年），要么涉及有偏差的粒度（人，2022 年）。

这项工作的目标是提出一种高效的无模拟 CNF 模型训练方法--流匹配（FM），允许采用一般概率路径来监督 CNF 训练。重要的是，FM 打破了可扩展 CNF 训练的障碍，使 CNF 训练超越了扩散。

ing 的障碍，避免了直接使用概率路径而重新考虑扩散过程的需要。

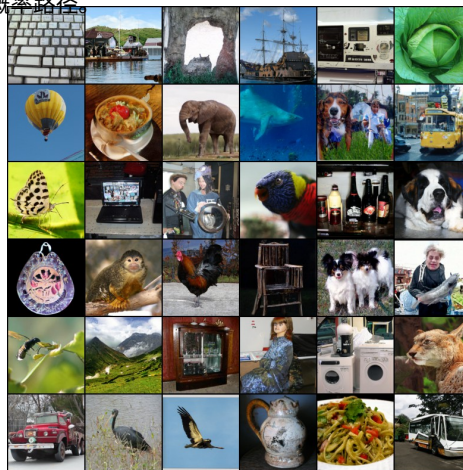


图 1：使用流量匹配和最优传输概率路径训练的 CNF 的无条件 ImageNet-128 样本。

我们特别提出了 "流匹配" 目标 (第 3 节)，这是一个简单直观的训练目标，即回归到目标向量场，从而生成所需的概率路径。我们首先证明，我们可以通过每个实例 (即条件) 的形式来构建这样的目标向量场。然后，受去噪分数匹配的启发，我们展示了一种称为条件流匹配 (Conditional Flow Matching, CFM) 的按实例训练目标，它能提供等效梯度，而且不需要关于难以处理的目标向量场的外部知识。此外，我们还讨论了可用于流匹配的每实例概率路径的一般族 (第 4 节)，它将扩散路径作为特殊实例。我们发现，即使是在扩散路径上，使用 FM 也能提供更稳健、更稳定的训练，并取得比分数匹配更优越的性能。此外，这个概率路径族还包括一个特别有趣的案例：与最优传输 (OT) 位移插值 (McCann, 1997 年) 相对应的向量场。我们发现，条件 OT 路径比扩散路径更简单，能形成直线轨迹，而扩散路径则会形成曲线轨迹。根据经验，这些特性似乎能带来更快的训练速度、更快的生成速度和更好的性能。

我们在大型、高度多样化的图像数据集 Im-ageNet 上对流量匹配和通过最优传输路径构建模型进行了实证验证。我们发现，在基于扩散的竞争方法中，我们可以轻松地训练模型，使其在似然估计和样本质量方面都取得良好的性能。此外，我们还发现，与之前的方法相比，我们的模型能更好地权衡计算成本和样本质量。图 1 展示了从我们的模型中选取的非传统 ImageNet 128×128 样本。

2 前言：连续归一化流量

让 \mathbf{R}^d 表示数据空间，数据点 $x = (x^1, \dots, x^d) \in \mathbf{R}^d$ 。本文使用的两个重要对象是：概率密度路径 $p : [0, 1] \times \mathbf{R}^d \rightarrow \mathbf{R}_{>0}$ ，它是与时间相关的⁽¹⁾概率密度函数，即 $\int p_t(x) dx = 1$ ，以及与时间相关的向量场 $v : [0, 1] \times \mathbf{R}^d \rightarrow \mathbf{R}^d$ 。向量场 v_t 可用于构建一个随时间变化的衍射图，称为流， $\varphi : [0, 1] \times \mathbf{R}^d \rightarrow \mathbf{R}^d$ ，通过常微分方程 (ODE) 定义：

$$\frac{d}{dt} \varphi_{(t)}(x) = v_t(\varphi_{(t)}(x)), \quad (1)$$

$$\varphi_0(x) = x \quad (2)$$

此前，Chen 等人 (2018) 建议用神经网络 $v_{\theta}(x; \theta)$ 对向量场 v_t 进行建模，其中 $\theta \in \mathbf{R}^p$ 是其可学习的参数，这反过来又会导致流 φ_t 的深度参数模型，称为连续归一化流 (CNF)。CNF 用于将简单的先验密度 $p_{(0)}$ (例如纯噪声) 通过前推方程重塑为更复杂的先验密度 p_1

$$p_t = [\varphi_t]_* p_{(0)} \quad (3)$$

其中，前推 (或变量变化) 算子 $*$ 的定义是

$$[\varphi_t]_* p_{(0)}(x) = p_{(0)}(\varphi_t^{-1}(x)) \det \frac{\partial \varphi_t^{-1}}{\partial x}(x). \quad (4)$$

如果向量场 v_t 的流 φ_t 满足方程 3，则称向量场 v_t 产生了概率密度路径 $p_{(t)}$ 。检验向量场是否产生概率路径的一个实用方法是使用连续性方程，它是我们证明中的一个关键组成部分，见附录 B。我们将在附录 C 中简要介绍有关 CNF 的更多信息，特别是如何计算任意点 $x \in \mathbf{R}^d$ 的概率 $p_1(x)$ 。

3 流量匹配

让 x_1 表示根据某个未知数据分布 $q(x_1)$ 分布的随机变量。我们假设只能获取 $q(x_1)$ 的数据样本，而无法获取密度函数本身。此外，我们假设 p 是一个概率路径，使得 $p_0 = p$ 是一个简单分布，例如标准正态分布 $p(x) = \mathcal{N}(x | 0, I)$ ，并假设 p_1 在分布上近似等于

q 。我们稍后将讨论如何构建这样一条路径。流量匹配目标就是为了匹配这条目标概率路径，使我们能从 p_0 流向 p_1 。

¹ 我们用下标表示时间参数，如 $p_t(x)$ 。

给定目标概率密度路径 $p_{\eta}(x)$ 和相应的向量场 $u_{\eta}(x)$ ，生成 $p_{\eta}(x)$ ，我们将流量匹配 (FM) 目标定义为

$$L_{\text{FM}}(\theta) = \mathbb{E}_{(u, p) \sim \mathcal{D}} \left\| v_{\eta}(x) - u_{\eta}(x) \right\|^2, \quad (5)$$

其中， θ 表示 CNF 向量场 v_{η} 的可学习参数（定义见第 2 节）， $t \sim U[0, 1]$ （均匀分布）， $x \sim p_{\eta}(x)$ 。简单地说，调频损耗用神经网络 v_{η} 对向量场 u_{η} 进行回归。当损失为零时，学习到的 CNF 模型将生成 $p_{\eta}(x)$ 。

流量匹配是一个简单而有吸引力的目标，但它本身在实际应用中却很难实现，因为我们事先并不知道什么是合适的 p_{η} 和 u_{η} 。有很多概率路径可以满足 $p_{(1)}(x) \approx q(x)$ ，更重要的是，我们通常无法获得能生成所需的 p_{η} 的封闭形式 u_{η} 。在本节中，我们将展示如何使用仅按样本定义的 *边缘* 概率路径和向量场来构建 p_{η} 和 u_{η} ，并通过适当的聚合方法得到所需的 p_{η} 和 u_{η} 。此外，这种构造还能让我们为流量匹配创建一个更易实现的目标。

3.1 从条件概率路径和向量场构建 p_{η} 、 u_{η}

构建目标概率路径的一个简单方法是混合更简单的概率路径：给定一个特定的数据样本 x_1 ，我们用 $p_t(x|x_1)$ 表示一个 *条件概率路径*，使得它在 $t=0$ 时满足 $p_0(x|x_1) = p(x)$ ，我们将 $t=1$ 时的 $p_1(x|x_1)$ 设计为一个集中在 $x=x_1$ 附近的分布，例如， $p_1(x|x_1) = \mathcal{N}(x|x_1, \sigma^2 \mathbf{I})$ ，一个具有 x_1 均值和足够小的标准差 $\sigma > 0$ 的正态分布。将 $q(x_1)$ 上的条件概率路径 *边缘化*，得到 *边缘概率路径*

$$p_{\eta}(x) = \int p_{\eta}(x|x_1) q(x_1) dx_1, \quad (6)$$

其中，特别是在时间 $t=1$ 时，边缘概率 p_1 是一个近似于数据分布 q 的混合分布、

$$p_1(x) = \int p_1(x|x_1) q(x_1) dx_1 \approx q(x). \quad (7)$$

有趣的是，我们还可以通过对条件向量场进行以下“边缘化”来定义 *边缘向量场*（我们假设 $p_{\eta}(x) > 0$ 适用于所有 t 和 x ）：

$$u_{\eta}(x) = \int u_{\eta}(x|x_1) \left\{ \frac{p_{\eta}(x|x_1) q(x_1)}{\int p_{\eta}(x|x_1) q(x_1) dx_1} \right\} dx_1, \quad (8)$$

其中 $u_{\eta}(x|x_1) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ 是产生 $p_{\eta}(x|x_1)$ 的条件向量场。这似乎并不明显，但这种条件向量场的聚合方式实际上产生了用于边缘概率路径建模的正确向量场。

我们的第一个关键观察点是

边缘向量场（方程 8）产生边缘概率路径（方程 6）。

这就在条件 VF（产生条件概率路径的 VF）和边缘 VF（产生边缘概率路径的 VF）之间建立了惊人的联系。这种联系使我们能够将未知且难以解决的边缘 VF 分解为更简单的条件 VF，而条件 VF 的定义要简单得多，因为它们只取决于单个数据样本。我们在下面的定理中正式说明了这一点。

定理 1. 给定产生条件概率路径 $p_{\eta}(x|x_1)$ 的向量场 $u_{\eta}(x|x_1)$ ，对于任意分布 $q(x_1)$ ，方程 8 中的边缘向量场 u_{η} 产生方程 6 中的边缘概率路径 p_{η} ，即 u_{η} 和 p_{η} 满足连续性方程（方程 26）。

所有定理的完整证明见附录 A。定理 1 也可以从 Peluchetti（2021 年）的扩散混合物表示定理中推导出来，该定理提供了扩散 SDE 中边缘漂移和扩散系数的计算公式。

3.2 条件流匹配

不幸的是，由于边际概率路径和 VF 定义（等式 6 和 8）中的积分很难求，计算 u_t 仍然很困难，因此，计算原始流动匹配目标的无偏估计值也很困难。相反，我们提出了一个更简单的目标，但令人惊讶的是，这个目标会产生与原始目标相同的最优结果。具体来说，我们考虑条件流量匹配 (CFM) 目标、

$$L_{\text{CFM}}(\theta) \mathbb{E}_{t \sim U[0,1], x_1 \sim q(x_1), p_t(x|x_1)} \left[\left(v_t(x) - u_t(x|x_1) \right)^2 \right] \quad (9)$$

其中 $t \sim U[0, 1]$, $x_1 \sim q(x_1)$ ，现在 $x \sim p_t(x|x_1)$ 。与 FM 目标不同，CFM 目标允许我们轻松地对无偏估计值进行采样，只要我们能有效地从 $p_t(x|x_1)$ 中采样并计算 $u_t(x|x_1)$ ，这两项工作都很容易完成，因为它们都是按样本定义的。因此，我们的第二个关键结论是

FM 目标（公式 5）和 CFM 目标（公式 9）与 θ 具有相同的梯度。

也就是说，优化 CFM 目标（在期望值上）等同于优化 FM 目标。因此，我们可以训练 CNF 来生成边际概率路径 p_t ，尤其是逼近 $t=1$ 时的未知数据分布 q ，而无需访问边际概率路径或边际向量场。我们只需设计合适的条件概率路径和向量场。我们在下面的定理中正式阐述了这一特性。

定理 2. 假设对于所有 $x \in \mathbb{R}^d$ 和 $t \in [0, 1]$, $p_t(x) > 0$ ，那么，直到一个与 θ 无关的常数， L_{CFM} 和 L_{FM} 相等。因此， $\nabla_{\theta} L_{\text{FM}}(\theta) = \nabla_{\theta} L_{\text{CFM}}(\theta)$ 。

4 条件概率路径和向量场

条件流匹配目标适用于任意选择的条件概率路径和条件向量场。在本节中，我们将讨论一般高斯条件概率路径族的 $p_t(x|x_1)$ 和 $u_t(x|x_1)$ 的构造。也就是说，我们考虑以下形式的条件概率路径

$$p_t(x|x_1) = \mathcal{N}(x | \mu_{(t)}(x_1), \sigma_{(t)}(x_1)(2)^{1/2}) \quad (10)$$

其中 $\mu : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ 是随时间变化的高斯分布均值，而 $\sigma : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}_{>0}$ 描述了随时间变化的标量标准差 (std)。我们设定 $\mu_{(0)}(x_1) = 0$ 和 $\sigma_{(0)}(x_1) = 1$ ，这样所有条件概率路径都会在 $t=0$ 时收敛到相同的标准高斯噪声分布， $p(x) = \mathcal{N}(x | 0, I)$ 。然后，我们设置 $\mu_1(x_1) = x_1$ 和 $\sigma_1(x_1) = \sigma_{\min}$ ， σ_{\min} 设置得足够小，以便 $p_1(x|x_1)$ 是以 x_1 为中心的集中高斯分布。

产生任何特定概率路径的矢量场有无数种（例如，在连续性方程中加入一个无发散分量，见方程 26），但其中绝大多数是由于存在使底层分布不变的分量--例如，当分布为旋转不变时的旋转分量--导致不必要的额外计算。我们决定使用与高斯分布的典型变换相对应的最简单向量场。具体来说，考虑流（以 x 为条件）

$$\psi_t(x) = \sigma_t(x_1)x + \mu_{(t)}(x_1) \quad (11)$$

当 x 分布为标准高斯时， $\psi_{(t)}(x)$ 是仿射变换，它映射到一个具有均值 $\mu_{(t)}(x_1)$ 和 $\text{std } \sigma_{(t)}(x_1)$ 的正态分布随机变量。也就是说，根据等式 4， ψ_t 将噪声分布 $p_0(x|x_1) = p(x)$ 推向 $p_t(x|x_1)$ ，即

$$[\psi_t]_* p_0(x) = p_t(x|x_1) \quad (12)$$

这个流提供了产生条件概率路径的矢量场：

$$\frac{d}{dt} \psi_t(x) = u_t(x|x_1) \quad (13)$$

将 $p_t(x|x_1)$ 重新参数化为 x_0 ，并将方程 13 插入 CFM 损失中，我们得到

$$L_{\text{CFM}}(\theta) \mathbb{E} = \mathbb{E}_{t \sim U[0,1], x_1 \sim q(x_1), x_0 \sim p(x)} \left[\left(v_t(\psi_t(x_0)) - u_t(\psi_t(x_0)|x_1) \right)^2 \right] \quad (14)$$

由于这是一个简单（可逆）仿射映射，我们可以利用方程 13 以封闭形式求解 $u_{(t)}$ 。让 f 表示随时间变化的函数 f 的导数，即 $f' = \frac{df}{dt}$ 。

定理 3. 设 $p_{(t), x(1)}$ 是等式 10 中的高斯概率路径, $\psi_{(t)}$ 是等式 11 中的相应流图。那么, 定义 $\psi_{(t), BS}$ 唯一矢量场的形式为

$$u_t(x|x(1)) = \frac{\sigma'(x_1)}{\sigma_{(t), x(1)}} (x - \mu_{(t), x(1)}) + \mu^{(t)}(x(1)). \quad (15)$$

因此, $u_t(x|x(1))$ 产生高斯路径 $p_{(t), x(1)}$ 。

4.1 高斯条件概率路径的特殊实例

对于任意函数 $\mu_{(t), x(1)}$ 和 $\sigma_{(t), x(1)}$ 而言, 我们的公式是完全通用的, 我们可以将它们设置为满足所需边界条件的任意可微分函数。我们首先讨论恢复与先前使用的扩散过程相对应的概率路径的特殊情况。由于我们直接使用概率路径, 因此我们可以直接放弃对扩散过程的推理。因此, 在下面的第二个例子中, 我们直接根据瓦瑟斯坦-2 最佳运输解作为一个有趣的实例来制定概率路径。

示例 I: 扩散条件 VF. 扩散模型从数据点开始, 逐渐添加噪声, 直到接近纯噪声。这些模型可以表述为随机过程, 在任意时间 t 时, 为了得到闭式表示, 这些过程有严格的要求, 结果是高斯条件概率路径 $p_{(t), x(1)}$, 具有特定选择的均值 $\mu_{(t), x(1)}$ 和 std $\sigma_{(t), x(1)}$ (Sohl-Dickstein 等人, 2015 年; Ho 等人, 2020 年; Song 等人, 2020 年b)。例如, 反向 (噪声 → 数据) 方差爆炸 (VE) 路径的形式为

$$p_t(x) = N(x | x(1), \sigma_1^2 t - I), \quad (16)$$

其中, σ_t 是递增函数, $\sigma_0 = 0$, $\sigma_1 = 1$ 。接下来, 公式 16 提供了 $\mu_t(x_1) = x_1$ 和 $\sigma_t(x_1) = \sigma_1 - \sigma_0$ 。将这些插入定理 3 的公式 15, 我们得到

$$u_t(x|x(1)) = -\frac{\sigma'(x_1)}{\sigma_1 t} (x - x(1)). \quad (17)$$

反向 (噪声 → 数据) 方差保持 (VP) 扩散路径的形式为

$$p_t(x|x(1)) = N(x | \alpha_1 - x_1, (1 - \alpha^2) t - I), \text{ where } \alpha = e^{-\int_0^t \beta(s) ds}, T(t) = \int_0^t \beta(s) ds, \quad (18)$$

和 β 是噪声标度函数。等式 18 提供了 $\mu_t(x_1) = \alpha_1 - x_1$ 和

$\sigma_t(x_1) = \alpha_1 - \alpha^2$ 。将这些插入定理 3 的公式 15, 我们得到

$$u_t(x|x(1)) = \frac{\alpha'(1-t)}{1 - \alpha^2} (\alpha_1 - x_1 - x) = -\frac{T'(1-t)}{2} \frac{e^{-\int_0^t \beta(s) ds} (x - e^{-\int_0^t \beta(s) ds} x_1)}{1 - e^{-\int_0^t \beta(s) ds}}. \quad (19)$$

事实上, 我们构建的条件 VF $u_{(t), x(1)}$ 与之前用于确定性概率流的向量场 (Song 等人 (2020b), 公式 13) 在限制这些条件扩散过程时确实是重合的; 详见附录 D。尽管如此, 将扩散条件 VF 与流量匹配目标相结合提供了一种有吸引力的训练替代方案—我们在实验中发现这种方案比现有的分数匹配方法更稳定、更稳健。

另一个重要的观察结果是, 由于这些概率路径之前是作为扩散过程的求解推导出来的, 它们实际上并不能在有限的时间内达到真正的噪声分布。在实践中, $p_{(0), x}$ 只是被一个合适的高斯分布近似, 用于采样和可能性评估。相反, 我们的结构提供了对概率路径的完全控制, 我们可以直接设置 μ_t 和 $\sigma_{(t)}$, 就像我们接下来要做的。

例二: 最佳运输条件 VF. 可以说, 条件概率路径更自然的选择是定义均值和 std 随时间线性变化, 即

$$\mu_t(x) = tx_1, \quad \sigma_t(x) = 1 - (1 - \sigma_{\min})t. \quad (20)$$

根据定理 3, 这条路径由 VF 生成

$$u(x|x(1)) = \frac{(x(1) - (1 - \sigma_{\min}))x}{1 - \sigma_{\min}}, \quad (21)$$

$$r = 1 - (1 - \sigma_{\min})t$$

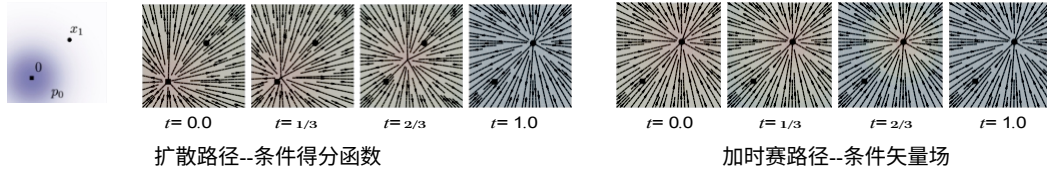


图 2: 与扩散路径的条件得分函数相比, 加时赛路径的条件矢量场在时间上具有恒定的方向, 可以说更容易用参数模型拟合。请注意, 蓝色表示较大的幅度, 红色表示较小的幅度。

与扩散条件 VF (等式 19) 不同的是, 它是针对所有 $t \in [0, 1]$ 定义的。与 $u_t(x|_{x(1)})$ 相对应的条件流为

$$\psi_t(x) = (1 - (1 - \sigma_{\min})t)x + t\sigma_{\min}x(1), \quad (22)$$

在这种情况下, CFM 损失 (见等式 9、14) 的形式为

$$L_{CFM}(\theta) = \mathbb{E}_{x \sim p(x_0)} \left[\frac{1}{2} \left(\frac{\psi_t(x) - x}{1 - \sigma_{\min}} \right)^2 \right] \quad (23)$$

允许均值和 std 线性变化不仅可以得到简单直观的路径, 而且实际上也是以下意义上的最优路径。条件流 $\psi_t(x)$ 实际上是两个高斯 $p_{(0)}(x|x(1))$ 和 $p_{(1)}(x|x(1))$ 之间的最优传输 (OT) 位移图。作为概率路径的 OT 内插值定义如下 (见 McCann (1997) 中的定义 1.1) :

$$p_t = [(1-t)\text{id} + t\psi]_* p_0 \quad (24)$$

其中 $\psi: \mathbb{R}^d \rightarrow \mathbb{R}^d$ 是将 p_0 推向 p_1 的 OT 映射, id 表示身份映射, $\text{id}(x) = x$, $(1-t)\text{id} + t\psi$ 称为 OT 位移映射。McCann (1997) 中的例 1.7 表明, 在两个高斯 (其中第一个是标准高斯) 的情况下, OT 位移图的形式为方程 22。

直观地说, OT 位移图下的粒子总是以恒定的速度沿直线轨迹移动。图 3 描述了扩散和 OT 条件 VF 的采样路径。有趣的是, 我们发现从扩散路径采样的轨迹可能会

会 "偏离" 最终样本, 导致不必要的回溯, 而 OT 路径则能保证保持直线。

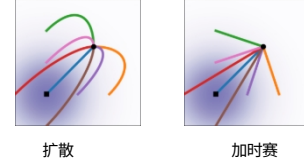


图 3: 扩散和加时赛轨迹。

图 2 比较了扩散条件分值函数 (扩散条件分值函数的再现) 和扩散条件分值函数 (扩散条件分值函数的再现)。

OT $\log p_t(x|x(1))$, p_t 的定义如式 18 所示, OT 条件 VF (式 21)。开始高斯 (p_0) 和结束高斯 (p_1) 两个例子中的高斯完全相同。一个有趣的现象是, OT 条件 VF 在时间上具有恒定的方向, 这可以说导致回归任务更加简单。这一特性也可以直接从等式 21 中验证, 因为 VF 可以写成 $u_t(x|x(1)) = g(t)h(x|x(1))$ 的形式。附录中的图 8 显示了扩散 VF 的直观图。最后, 我们注意到, 虽然条件流是最优的, 但这并不意味着边界 VF 是最优的传输解。不过, 我们希望边界矢量场保持相对简单。

5 相关工作

连续归一化流在 (陈等人, 2018) 中作为归一化流的连续时间版本被引入 (参见 Kobyzev 等人 (2020) ; Papamakarios 等人 (2021) 的概述)。最初, CNFs 采用最大似然目标进行训练, 但这涉及昂贵的前向和后向传播 ODE 仿真, 由于 ODE 仿真的顺序性, 导致时间复杂度较高。尽管一些研究证明了 CNF 生成模型在图像合成方面的能力 (Grathwohl 等人, 2018 年), 但将其扩展到非常高维的图像本身就很困难。一些研究试图对 ODE 进行正则化处理, 使其更容易求解, 例如使用增强 (Dupont 等人, 2019 年)、添加正则化项 (Yang & Karniadakis, 2019 年; Finlay 等人, 2020 年; Onken 等人, 2021 年; Tong 等人, 2020 年; Kelly 等人, 2020 年) 或随机采样积分区间 (Du 等人, 2022 年)。这些研究只是为了正则化 ODE, 并没

有改变基本的训练算法。

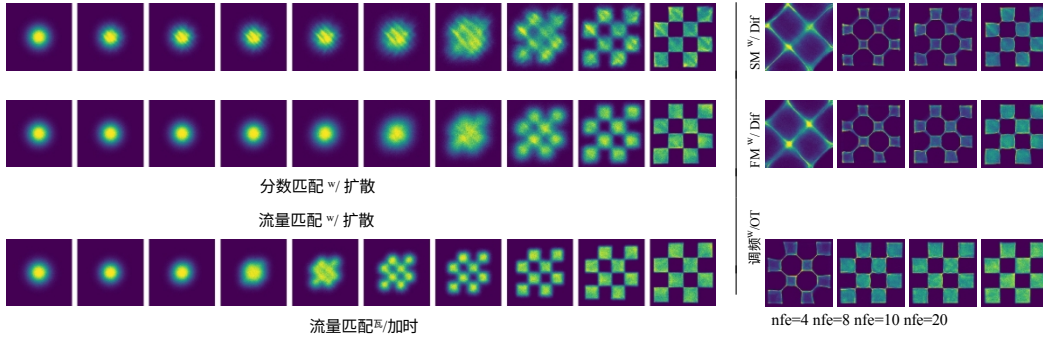


图 4: (左) 在二维棋盘数据上使用不同目标训练的 CNF 的轨迹。加时赛路径更早引入棋盘图案, 而调频路径的训练结果更稳定。(右图) 使用中点方案求解的带有加时赛的调频能带来更高效的采样。

为了加快 CNF 训练, 一些研究通过明确设计目标概率路径和动态, 开发了免模拟 CNF 训练框架。例如, Rozen 等人 (2021 年) 考虑了先验和目标密度之间的线性插值, 但涉及在高维度下难以估计的积分; 而 Ben-Hamu 等人 (2022 年) 考虑了与本研究类似的一般概率路径, 但在随机迷你批次机制中存在梯度偏差。相比之下, "流匹配" 框架允许无模拟训练, 梯度无偏, 并可轻松扩展到非常高的维度。

另一种无模拟训练方法依赖于构建一个扩散过程来间接定义目标概率路径 (Sohl-Dickstein 等, 2015; Ho 等, 2020; Song & Ermon, 2019)。Song 等人 (2020b) 的研究表明, 扩散模型是利用去噪分数匹配 (Vincent, 2011) 进行训练的, 这是一种条件目标, 可提供与分数匹配目标相关的无偏梯度。条件流匹配从这一结果中汲取灵感, 但直接推广到匹配向量场。由于易于扩展, 扩散模型受到越来越多的关注, 并产生了各种改进, 如损失缩放 (Song 等人, 2021 年)、增加分类器指导以及架构改进 (Dhariwal 和 Nichol, 2021 年) 和学习噪声时间表 (Nichol 和 Dhariwal, 2021 年; Kingma 等人, 2021 年)。不过, (Nichol & Dhariwal, 2021 年) 和 (Kingma 等人, 2021 年) 只考虑了由简单扩散过程定义的高斯条件路径的限制性设置, 其中只有一个参数, 尤其不包括我们的条件 OT 路径。在另一项研究中, (De Bortoli 等人, 2021 年; Wang 等人, 2021 年; Peluchetti, 2021 年) 通过扩散桥理论提出了有限时间扩散结构, 解决了无限时间去噪结构产生的近似误差问题。虽然现有研究利用了扩散过程与具有相同概率路径的连续归一化流之间的联系 (Maoutsa 等人, 2020b; Song 等人, 2020b; 2021), 但我们的研究使我们能够超越简单扩散建模的概率路径类别。有了我们的工作, 就有可能完全避开扩散过程的构建, 直接使用概率路径进行推理, 同时仍然保留高效的训练和对数似然评估。最后, 与我们的工作同时进行的 (Liu 等人, 2022; Albergo & Vanden-Eijnden, 2022) 也为 CNF 的无模拟训练提出了类似的条件目标, 而 Neklyudov 等人 (2023) 则在假定 u_t 是梯度场时推导出了隐式目标。

6 实验

我们在分辨率为 32、64 和 128 的 CIFAR-10 (Krizhevsky 等人, 2009 年) 和 ImageNet (Chrabaszcz 等人, 2017 年; Deng 等人, 2009 年) 图像数据集上探索了使用流量匹配的实证优势。我们还消解了流匹配 (Flow Matching) 中扩散路径的选择, 尤其是标准方差保存扩散路径和最优传输路径之间的选择。我们讨论了如何通过直接参数化生成矢量场和使用流匹配目标来改进样本生成。最后, 我们展示了流动匹配也可用于条件生成设置。除非另有说明, 我们使用 `dopri5` (Dormand 和 Prince, 1980 年) 对模型的似然和样本进行评估, 绝对和相对容差均为 $1e-5$ 。生成的样本见附录, 所有实施细节见附录 E。

模型	CIFAR-10			ImageNet 32× 32			ImageNet 64× 64			模型	ImageNet 128× 128		
	NLL↓	FID↓	NFE↓	NLL↓	FID↓	NFE↓	NLL↓	FID↓	NFE↓		NLL↓	FID↓	
消融													
DDPM		3.12	7.48	274	3.54	6.99	262	3.32	17.36	264	MGAN (Hoang 等人, 2018 年)	-	58.9
得分匹配		3.16	19.94	242	3.56	5.68	178	3.40	19.74	441	PacGAN2 (Lin 等人, 2018 年)	-	57.5
ScoreFlow		3.09	20.78	428	3.55	14.14	195	3.36	24.95	601	Logo-GAN-AE (Sage et al., 2018)	-	50.9
我们的													
调频 ω / 扩散	3.10		8.06	183	3.54	6.37	193	3.33	16.88	187	自身条件GAN (Luc'ic' et al., 2019)	-	41.7
调频 ω /加时		2.99	6.35	142	3.53	5.02	122	3.31	14.45	138	未知BigGAN (Luc'ic' et al., 2019)	-	25.3
											PGMGAN (Armandpour 等人, 2021 年)	-	21.7
调频 ω /加时													
												2.90	20.9

表 1: 用不同方法训练的同一模型的概率 (BPD)、生成样本的质量 (FID) 和评估时间 (NFE)。

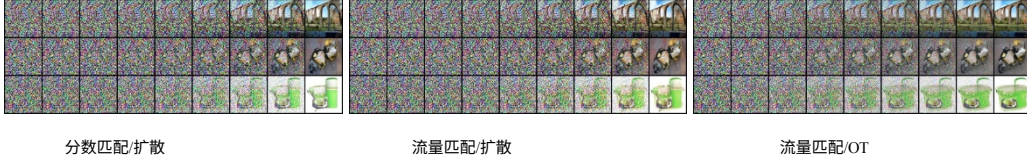


图 6: 在 ImageNet 64×64 上训练的模型在相同初始噪声下的路径样本。OT 路径大致以线性方式减少噪音，而扩散路径仅在路径末端明显去除噪音。还请注意生成图像之间的差异。

6.1 IMAGENET 上的密度建模和样本质量

我们首先比较了在 CIFAR-10 和 Dhariwal 和 Nichol (2021 年) 的 U-Net 架构 (改动极小)，以及不同流行的基于扩散的损失: DDPM 来自 (Ho 等人, 2020 年)，Score Matching (SM) 来自 (Song 等人, 2020 年b)，Score Flow (SF) 来自 (Song 等人, 2021 年)；具体细节见附录 E.1。表 1 (左) 总结了我们与这些基线模型的研究结果，报告了以每维比特 (BPD) 为单位的负对数似然 (NLL)、以弗雷谢特起始距离 (FID; Heusel 等, 2017 年) 衡量的样本质量，以及自适应求解器达到预设数值容差所需的平均函数评估次数 (NFE) (50k 个样本的平均值)。所有模型都使用相同的架构、超参数值和训练迭代次数进行训练，其中基线允许更多的迭代次数以获得更好的收敛性。请注意，这些都是无条件模型。在 CIFAR-10 和 ImageNet 上，与其他竞争方法相比，FM-OT 在所有定量指标上都始终获得最佳结果。与之前的研究 (Ho 等人, 2020; Song 等人, 2020b; 2021) 相比，我们注意到 CIFAR-10 的 FID 性能高于通常水平，这可能是因为我们使用的架构没有针对 CIFAR-10 进行优化。

其次，表 1 (右) 比较了在分辨率为 128×128 的 ImageNet 上使用 "流匹配" 训练的模型与 OT 路径。除 IC-GAN (Casanova 等人, 2021 年) 使用自监督 ResNet50 模型进行调节外，我们的 FID 是最先进的，因此未列入本表。附录中的图 11、图 12 和图 13 显示了这些模型的非校正样本。

更快的训练。虽然现有研究在训练扩散模型时需要很高的迭代次数 (例如 Score Flow 和 VDM 报告的迭代次数分别为 130 万次和 1000 万次)，但我们发现 Flow Matching 通常收敛得更快。图 5 显示了 Flow Matching 和所有基线在 ImageNet 64×64 的训练过程中的 FID 曲线；与其他方法相比，FM-OT 能够更快、更大程度地降低 FID。对于

对于 ImageNet-128, Dhariwal 和 Nichol (2021 年) 以 256 的批量大小训练了 436 万次迭代，而 FM (模型大 25%) 以 1.5k 的批量大小训练了 500k 次迭代，图像数量减少了 33%。

具体细节见表 3。此外，在进行分数匹配训练时，从模型中取样的成本会发生急剧变化，而使用流量匹配训练时，取样成本则保持不变 (附录图 10)。

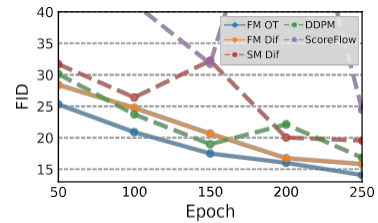


图 5: 训练期间的图像质量，ImageNet 64×64。

6.2 采样效率

采样时，我们首先抽取一个随机噪声样本 $x_0 \sim N(0, I)$ ，然后在区间 $t \in [0, 1]$ 上使用 ODE 求解器用训练后的 VF v_t 求解方程 1，计算 $\varphi(1)(x_{(0)})$ 。而扩散

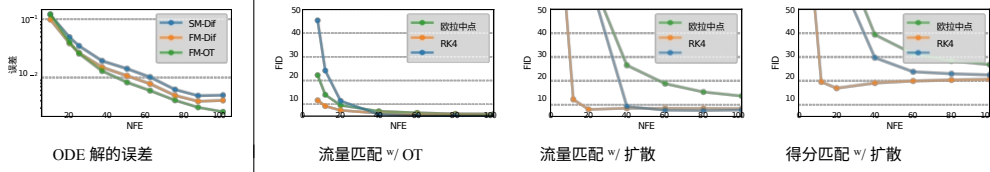


图 7: 流量匹配（尤其是在使用 OT 路径时）允许我们使用更少的评估进行采样，同时保持相似的数值误差（左）和采样质量（右）。结果显示的是在 ImageNet 32×32 上训练的模型，数值误差为 midpoint 方案的数值误差。

虽然也可以通过 SDE 公式对模型进行采样，但这种方法效率很低，许多提出快速采样器的方法（如 Song 等人 (2020a); Zhang & Chen (2022)）都直接从 ODE 角度进行采样（见附录 D）。部分原因是 ODE 求解器效率更高--在计算成本相近的情况下误差更小（Kloeden 等, 2012 年）--以及可用的 ODE 求解器方案众多。与我们的消融模型相比，我们发现，无论使用哪种 ODE 求解器，使用流量匹配和 OT 路径训练的模型总是能产生最高效的采样器。

采样路径。我们首先对扩散和 OT 在采样路径上的差异进行了定性分析。图 6 显示了使用相同随机种子的 ImageNet-64 模型的样本，我们发现 OT 路径模型比扩散路径模型更早开始生成图像，而扩散路径模型的图像直到最后一个时间点才受到噪声的影响。我们还描绘了二维生成棋盘图案的概率密度路径，如图 4（左），发现了类似的趋势。

低成本样本。接下来，我们切换到固定步长求解器，并比较使用表 1 中 ImageNet-32 模型计算的低 NFE 样本 (≤ 100)。在图 7（左）中，我们比较了低 NFE 解算的每像素 MSE 与 1000 NFE 解算的每像素 MSE（我们使用 256 个随机噪声种子），并注意到就计算成本而言，带 OT 模型的调频产生的数值误差最好，大约只需要 60% 的 NFE 就能达到与扩散模型相同的误差阈值。其次，图 7（右）显示了 FID 随计算成本的变化而变化的情况，我们发现即使在 NFE 值非常低的情况下，带 OT 的调频模型也能达到很好的 FID，与烧蚀模型相比，它能更好地权衡样本质量和成本。图 4（右）显示了二维棋盘实验的低成本采样效果。

6.3 从低分辨率图像中进行条件采样

最后，我们尝试使用流匹配技术生成条件图像。特别是将图像从 64×64 提升到 256×256。我们按照（Saharia 等人, 2022 年）中的评估程序，计算上采样验证图像的 FID；基线包括参考（原始验证集的 FID）和回归。结果

见表 2。升采样图像样本见附录中的图 14 和图 15。FM-OT 的 PSNR 和 SSIM 值与（Saharia et al.

模型	FID↓	IS↑	PSNR↑	SSIM↑
参考	1.9	240.8	-	-
回归	15.2	121.1	27.9	0.801
SR3 (Saharia 等人, 2022 年)	5.2	180.1	26.4	0.762
调频 ^{OT} /加时	3.4	200.8	24.7	0.747

表 2: ImageNet 验证集上的图像超分辨率

正如（Saharia 等人, 2022 年）所认为的那样，FM-OT 更好地反映了生成质量。

7 结论

我们介绍了“流匹配”（Flow Matching）--一种用于训练连续归一化流模型的全新免模拟框架，它依赖于条件构造，可以毫不费力地扩展到非常高的维度。此外，FM 框架还为扩散模型提供了另一种视角，并建议放弃随机/扩散结构，转而更直接地指定概率路径，使我们能够构建更快采样和/或改进生成的路径。我们在实验中展示了使用流量匹配框架进行训练和采样的简便性，未来，我们希望 FM 能为多种概率路径（如非各向异性高斯或更通用的核）打开大门。

8 社会责任

除了许多积极的应用之外，图像生成也可能被用于有害的目的。使用内容控制的训练集和图像验证/分类有助于减少这些用途。此外，训练大型深度学习模型的能源需求正在快速增长（Amodei 等人，2018 年；Thompson 等人，2020 年），专注于能够使用较少梯度更新/图像吞吐量进行训练的方法，可以大大节省时间和能源。

参考文献

- Michael S Albergo 和 Eric Vanden-Eijnden. 用随机间距构建归一化流。 *arXiv 预印本 arXiv:2209.15571*, 2022.
- Dario Amodei, Danny Hernandez, Girish Sastry, Jack Clark, Greg Brockman 和 Ilya Sutskever. Ai and compute. <https://openai.com/blog/ai-and-compute/>, 2018.
- Mohammadreza Armandpour, Ali Sadeghian, Chunyuan Li, and Mingyuan Zhou. 分区引导的甘斯。 In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.5099-5109, 2021.
- Heli Ben-Hamu, Samuel Cohen, Joey Bose, Brandon Amos, Aditya Grover, Maximilian Nickel, Ricky T. Q. Chen 和 Yaron Lipman. 流形上的匹配归一化流和概率路径。 *arXiv 预印本 arXiv:2207.04711*, 2022.
- Arantxa Casanova, Marlene Careil, Jakob Verbeek, Michal Drozdal, and Adriana Romero Soriano. 实例条件矢量。 *神经信息处理系统进展*，34: 27517-27529, 2021.
- Ricky T. Q. Chen. torchdiffeq, 2018. URL <https://github.com/rtqichen/torchdiffeq>.
- Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. 神经常微分方程。 *神经信息处理系统进展*，2018 年第 31 期。
- Patryk Chrabaszcz, Ilya Loshchilov 和 Frank Hutter. imagenet 的下采样变体作为 cifar 数据集的替代品。 *arXiv 预印本 arXiv:1707.08819*, 2017.
- Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. (arXiv:2106.01357), Dec 2021. doi: 10.48550/arXiv.2106.01357. URL <http://arxiv.org/abs/2106.01357>. arXiv:2106.01357 [cs, math, stat].
- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: 大规模分层图像数据库。 In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. doi: 10.1109/CVPR.2009.5206848.
- Prafulla Dhariwal 和 Alexander Quinn Nichol. 扩散模型在图像合成中击败 GANs。 见 A. Beygelzimer, Y. Dauphin, P. Liang 和 J. Wortman Vaughan (编)，《神经信息处理系统进展》，2021 年。 URL <https://openreview.net/forum?id=AAWuCVzaVt>.
- 约翰-R-多尔曼和彼得-J-普林斯. 嵌入式 runge-kutta 公式族。 *计算与应用数学杂志*, 6 (1) : 19-26, 1980.
- Shian Du, Yihong Luo, Wei Chen, Jian Xu 和 Delu Zeng. To-flow: 带有移动速度的时序优化邻接的高效连续流, 2022 年。 URL <https://arxiv.org/abs/2203.10335>.
- Emilien Dupont, Arnaud Doucet 和 Yee Whye Teh. Augmented neural odes. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL <https://proceedings.neurips.cc/paper/2019/file/21be9a4bd4f81549a9d1d241981cec3c-Paper.pdf>.