SERPENTFLOW：通过共享结构分解进行生成式非配对域对齐

Julie KeislerARCHES 团队 INRIA Paris
julie.keisler@inria.fr

Anastase CharantonisARCHES 团队 INRIA Paris

Yannig GoudeEDF Lab Paris-Saclay 巴黎-萨克雷大学

Boutheina OueslatiEDF Lab Paris-Saclay

Claire MonteleoniARCHES 团队 INRIA Paris 科罗拉多大学博尔德分校

arXiv:2601.01979v1 [cs.LG] 5 Jan 2026

摘要

域对齐广义上是指学习不同域的数据分布之间的对应关系。本文聚焦于这样一种情形：尽管各域的具体表现形式不同，但它们共享潜在的结构模式。在缺少配对观测的情况下，这项任务尤其具有挑战性，因为这意味着无法获得跨域的直接监督。我们提出了一种用于非配对域对齐的生成式框架，称为 SerpentFlow（生成式域适应的共享结构分解）。SerpentFlow 将潜在空间中的数据分解为两个部分：两个域共有的共享成分，以及域特定成分。通过分离共享结构并用随机噪声替代域特定成分，我们构造了共享表示与目标域样本之间的合成训练对，从而能够使用传统上仅适用于配对情形的条件生成模型。我们将该方法应用于超分辨率任务，其中共享成分自然对应于低频内容，而高频细节则体现域特定变异。用于区分低频与高频分量的截止频率通过基于分类器的判据自动确定，从而确保分解由数据驱动并能适应不同域。通过生成保留低频结构、同时注入随机高频实现的伪配对，我们学习给定共享表示时目标域的条件分布。我们采用流匹配作为生成流程来实现 SerpentFlow，不过该框架也兼容其他条件生成方法。在合成图像、物理过程模拟和气候降尺度任务上的实验表明，该方法能够有效重建与潜在低频模式一致的高频结构，说明共享结构分解是实现非配对域对齐的有效策略。[1]

关键词 域对齐、生成模型、表示学习、非配对数据、无监督学习、共享结构分解、基于频率的表示

1 引言

生成建模旨在从仅通过示例观测到的复杂数据分布中合成逼真的样本。通常，这是通过学习从简单源域分布（例如高斯分布）到由可用数据点表示的目标分布的映射来实现的。归一化流 [1], 变分自编码器（变分自编码器） [2], 生成对抗网络（生成对抗网络） [3], 以及基于扩散的方法 [4, 5, 6]等框架为学习此类映射提供了灵活的工具，从而能够高保真地生成图像 [3, 4], 音频信号 [7],

[1]我们的代码可在此处获取： https://github.com/JulieKeisler/serpentflow

或蛋白质结构 [8, 9]。将数据生成构建为分布之间的传输，已被证明既在理论上优雅，又在实践中强大。

沿着这一视角，可以考虑两个经验域之间的传输，$\mathcal{A}$ 和 $\mathcal{B}$，每个域都关联一个概率分布，p_A 和 p_B，且同样只能通过样例观测。目标是学习一种映射，将来自 $\mathcal{A}$ 的样本转换为来自 $\mathcal{B}$ 的样本，同时遵循底层分布。当存在配对样本（$x_{\mathcal{A}}$，$x_{\mathcal{B}}$）时，条件生成方法可以直接学习 $p(x_{\mathcal{B}} \mid x_{\mathcal{A}})$，例如条件生成对抗网络 [10] 或基于扩散的条件模型 [11]。
然而，在许多现实场景中，配对数据并不可用。不同域可能对应于在不同传感器、分辨率或物理条件下采集的测量值，彼此之间没有一一对应关系。这一非配对域对齐问题催生了 CycleGAN [12], 、MUNIT [13], 、AlignFlow [14],等方法，以及较新的随机桥接或连续时间传输方法 [15, 16, 17, 18]。这些方法通过施加动态或潜在空间一致性来连接边缘分布，但往往缺乏明确的机制来识别各域真正共享的结构，从而可能限制模型的可解释性以及两个域之间的一致性。
我们提出 SerpentFlow（生成式域适应的共享结构分解），这是一种基于显式分离共享成分与域特定成分的无配对域适应框架。我们假设，来自两个域 $\mathcal{A}$ 和 $\mathcal{B}$ 的样本可以映射到一个共同的潜在表示，该表示捕获跨域共享的结构，而其余变率则是域特定的。该共享表示定义了一种可从两个域到达的潜在分布，并作为对齐的接口。SerpentFlow 利用这一潜在共享空间构造合成训练对。具体而言，将目标域样本映射到共享表示，同时以随机噪声替换域特定成分。随后训练一个条件生成模型，从这一部分指定的输入重建完整的目标域样本。该训练过程学习给定共享表示时目标域特定内容的条件分布，无需域间配对观测。在推理时，源域样本被投影到相同的共享潜在表示。随后，使用训练好的条件生成器，以该共享表示为条件，对目标域特定成分进行采样，从而生成对齐后的目标域输出。值得注意的是，SerpentFlow 不学习域之间的显式映射；相反，它通过用在目标域上学到的、与域一致的细节补全共享结构来实现对齐

在本研究中，我们将 SerpentFlow 应用于无监督超分辨率任务，其中共享表示对应粗尺度信息，域特定成分则捕捉精细尺度细节。为以数据驱动的方式识别共享成分，我们逐步移除输入中的信息，直至域分类器无法区分来自两个域的样本。该过程得到的共享部分与域特定部分分解在不同域之间保持一致。基于这一分解，我们在目标域上训练生成模型：输入中仅保留共享成分，其余内容替换为噪声，并据此重建高分辨率样本。在推理时，将源域的低分辨率样本映射到相同的共享表示，再由训练好的生成器补全，从而得到与目标域对齐的高分辨率输出。在本文考虑的超分辨率场景中，这种共享表示可通过基于频率的分解自然实现：低频分量捕捉两个域共有的大尺度结构，而高频分量则编码各域不同的精细细节。频率截断点由域分类器自动选定，确保仅保留域不变结构。

我们在三个数据集上验证了 SerpentFlow。首先是一个受控图像数据集，其中人为移除了高频分量，以模拟域偏移。其次是一个模拟物理过程的数据集：我们以两种不同的空间频率对模拟方程进行采样，从而生成高分辨率和低分辨率数据。第三个是气候降尺度应用，通过增强粗分辨率模拟来恢复精细空间模式。在所有实验中，SerpentFlow 都能稳定地重建连贯的高频结构，同时保留底层低频内容，从而确保物理一致性和统计一致性。

总而言之，我们的贡献如下：

- 我们提出了 SerpentFlow，这是一种用于非配对域对齐的流程，它将数据分解为共享成分和域特定成分，并通过在域特定部分注入随机性来构造合成配对样本。
- 我们为无监督超分辨率任务提出了该流程的一种基于频率的实例化方法：使用分类器自动确定低频/高频截止点，并使用生成模型重建高频内容。
- 与现有的非配对生成方法相比，我们的方法在受控图像、物理模拟和气候降尺度数据集上展现出更好的连贯性、鲁棒性和泛化能力。

2 背景

生成框架。 生成建模旨在学习从简单的潜在分布（通常为高斯分布）到复杂数据分布的变换，而复杂数据分布只能通过有限的样本观测到。令 $z \sim \mathcal{N}(0, I)$ 表示一个潜变量， $x \sim p(x)$ 表示一个数据样本。其目标是学习一个映射 T_θ： $z \mapsto x$ ，以捕捉 $p(x)$ 的结构。不同的框架采用不同的数学方法来实现这一理念。生成对抗网络（GAN） [3] 学习一个隐式生成器，并通过与判别器进行对抗训练；归一化流 [1] 则定义可逆映射，并能精确评估似然。神经常微分方程（Neural ODE） [19] 将流扩展到连续时间动力学，而扩散模型 [4, 5] 学习逆转随机加噪过程，从而达到最先进的保真度和稳定性。流匹配 [6] 和随机插值 [16], 等较新的方法则直接学习随时间变化的速度场，在源域分布和目标分布之间传输概率质量。这些框架共同提供了灵活且具有坚实数学基础的工具，用于定义分布之间的变换，因此特别适用于域对齐任务。

非配对域对齐。 非配对域对齐旨在解决这样一个问题：在无法获取配对观测的情况下，如何学习来自两个域 $\mathcal{D}_A$ 和 $\mathcal{D}_B$、分别服从分布 p_A 和 p_B的样本之间的对应关系。经典应用包括图像到图像转换，例如将马转换为斑马，或将夏景转换为冬景 [12], ；艺术风格迁移，例如将照片转换为梵高风格的画作 [20],；更一般地说，使用 StyleGAN 架构生成风格混合可控的图像 [21]。另一个例子是无监督超分辨率，其目标是从粗分辨率输入中推断精细尺度的细节 [22]。相关思想也应用于科学领域，包括压缩感知 [23, 24] 和气候降尺度，即根据粗分辨率模拟重建高分辨率物理场 [25, 26]。

早期方法（如 CycleGAN [12], ）依靠采用双向生成器的循环一致对抗训练，以保留不同域之间的内容。UNIT [13], 、AlignFlow [14], 和双扩散隐式桥（DDIB） [15]利用共享潜在表示（通常为高斯分布）作为两个域之间的接口。较新的随机传输框架，包括薛定谔桥流 [18] 和整流流 [27], ，对 p_A 和 p_B之间的连续轨迹进行建模，通常无需对抗训练，同时保留显式的传输动力学。在无监督超分辨率中，扩散桥形式 [26] 利用这样一种思路：通过向低分辨率数据部分添加噪声，我们便能进入生成高分辨率数据的扩散模型轨迹。

尽管这些方法在经典图像转换任务中取得了视觉上令人信服的结果，但要量化哪些结构应予以保留或修改，仍然颇具挑战。然而，在科学和物理应用中，定义明确的度量指标能够评估输出是否遵循底层域结构。这促使人们探索能够明确分离共享成分与域特定成分的方法，从而实现稳健且可解释的非配对域对齐。

3 方法

尽管我们的方法建立在现有框架之上，但它的不同之处在于，旨在建立两个域之间更明确的映射，从而清楚地指出源域的哪些方面应予以保留，哪些方面应进行转换。其基本原则是，要使域对齐有意义，两个域之间必须存在共享结构，同时数据的其他部分也存在真实差异。因此，域对齐的目标是保留这一共享成分，同时适当地修改源域中的域特定部分，并确保与那些不打算改变的部分保持一致。

3.1 具有共享结构的域对齐

考虑嵌入在同一空间 $\mathcal{X} \subset \mathbb{R}^D$中的两个域 $\mathcal{D}_A \subset \mathcal{X}$ 和 $\mathcal{D}_B \subset \mathcal{X}$。我们分别用 p_A 和p_B 表示 $\mathcal{D}_A$ 和 $\mathcal{D}_B$中样本的分布。域对齐的目标是学习一个映射，将来自 p_A 的样本转换到 p_B。要使这种对齐有意义，两个域必须至少共享少量共同结构。

形式化地说，设存在一个映射到给定潜在空间 $\mathcal{Z}$的双射映射：

$$\mu : \mathcal{X} \to \mathcal{Z}. \tag{1}$$

使得

$$\mu(\mathcal{D}_A) = \mathcal{B}_A \subset \mathcal{Z}, \qquad \mu(\mathcal{D}_B) = \mathcal{B}_B \subset \mathcal{Z}. \tag{2}$$

我们假设两个 h $\mathcal{B}_A$ 和 $\mathcal{B}_B$ 可分解为共享成分和域特定（不同的）成分 e:

$$\mathcal{B}_A = \mathcal{B}_A^S \oplus \mathcal{B}_A^D, \qquad \mathcal{B}_B = \mathcal{B}_B^S \oplus \mathcal{B}_B^D, \tag{3}$$

并假设 $\mathcal{B}_A^S = \mathcal{B}_B^S = \mathcal{B}^S$。在这种设定下，$\mathcal{B}^S$ 表示两个域共享的潜在结构，而 $\mathcal{B}_A^D$ 和 $\mathcal{B}_B^D$ 则捕捉域特定变率。

对于样本 $x_A^i \sim p_A$，我们记作

$$\mu(x_A^i) = z_A^i = z_A^{i,S} + z_A^{i,D}, \tag{4}$$

其中 $z^{i,}{}_A{}^S \in \mathcal{B}^S$ 和 $z^{i,}{}_A{}^D \in \mathcal{B}_A^D$。类似地，对于 $x_B^i \sim p_B$，

$$\mu(x_B^i) = z_B^i = z_B^{i,S} + z_B^{i,D}. \tag{5}$$

输入空间中的一个白噪声样本 $\epsilon \sim \mathcal{N}(\mathbf{0},\ \mathbf{I})$ 被映射为

$$\mu(\epsilon) = z_\epsilon = z_\epsilon^S + z_\epsilon^D \in \mathcal{B}_\epsilon^S \oplus \mathcal{B}_\epsilon^D \subset \mathcal{Z}. \tag{6}$$

然后，我们可以将目标域中的共享成分与域特定部分的一个随机实现相结合，从而构造伪配对：

$$(\tilde{x}_B^i, x_B^i) \in \big(\mu^{-1}(\mathcal{B}^S \oplus \mathcal{B}_\epsilon^D) \times \mathcal{D}_B\big), \qquad \tilde{x}_B^i = \mu^{-1}(z_B^{i,S} + z_\epsilon^D). \tag{7}$$

随后可以训练生成模型 f_θ，将这些伪配对映射为：

$$f_\theta : \mu^{-1}(\mathcal{B}^S \oplus \mathcal{B}_\epsilon^D) \to \mathcal{D}_B. \tag{8}$$

这种表述可以理解为：给定共享成分 z_B^S，学习 p_B 的域特定成分的条件分布：

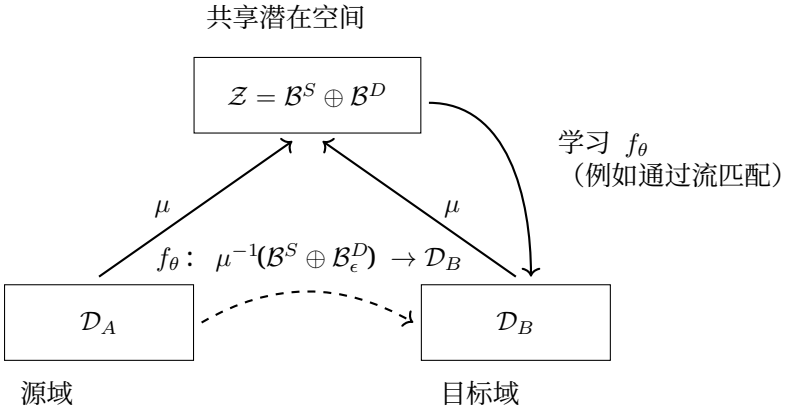
$$f_\theta : \tilde{x}_B \mapsto x_B \sim p_B(x_B \mid z_B^S). \tag{9}$$

训练完成后，可以通过以下方式将新样本 $x_A^i \in \mathcal{D}_A$ 转换到目标域 $\mathcal{D}_B$

$$f_\theta(\tilde{x}_A^i) = f_\theta\Big(\mu^{-1}(z_A^{i,S} + z_\epsilon^D)\Big). \tag{10}$$

在实践中应用这一通用框架的主要挑战在于：(i) 找到一个潜在空间 $\mathcal{Z}$，使共享成分与域特定成分能够清晰分离；(ii) 找到合适的双射映射 μ。此外，$\mathcal{B}^S$ 应尽可能大，以保留源样本 x_A^i 中的充足信息。在$\mathcal{D}_A$ 与 $\mathcal{D}_B$ 不具有任何共同结构的极限情况下，我们有 $\mathcal{B}^S = \varnothing$，这对应于无条件生成建模。

注。 在此表述中，条件化在潜在空间 $\mathcal{Z}$ 中表示为求和。原则上，也可以考虑其他条件化机制；这些替代方案将在实验部分讨论（见第 *4* 节）。



3.2 共享域假设的基于频率的实例化

所提出的框架假设存在一个潜在空间 $\mathcal{Z}$，在其中可以有效地解耦共享结构和域特定结构。在无监督超分辨率的背景下，我们将此潜在空间实例化于傅里叶域，其中每个场都分解为其空间频率分量。低频模态编码了通常在不同域之间共享的大尺度连贯结构，而高频模态则捕捉了细尺度的域特定变化，这些变化往往无法从粗分辨率观测中分辨出来。因此，这种基于频率的分解为分离和重建与无监督超分辨率相关的分量提供了自然且易于解释的基础。

傅里叶表示特别适合这一设置，因为它按照尺度和局部性组织信息。许多复杂系统都表现出尺度依赖行为，在傅里叶域中这种行为会变得更加明显[28, 29, 30]。在基于物理的仿真中，解析精细尺度特征往往会带来难以承受的计算开销，而控制算子通常会选择性地作用于不同频带 [31]。此外，实际的数据采集和采样过程实际上起到低通滤波器的作用，会衰减或截断高频内容 [32]。此前关于信号重建的研究表明，在稀疏性或凸正则化假设下，可以根据有限的低频观测恢复缺失的高频信息 [23], ，尽管这类保证通常仅适用于低维或理想化情形。综合来看，这些因素使傅里叶域成为一种原理充分且用途广泛的潜在空间，可用于在超分辨率任务中将共享的低频结构与域特定的高频细节分离开来。

傅里叶域表述。令 $\mathcal{X} \subset \mathbb{R}^D$ 表示观测空间（例如，网格化的物理场或图像）。我们将傅里叶变换及其逆变换定义为:

$$\mathcal{F}\{x\}(\boldsymbol{\xi}) = \int_{\mathbb{R}^D} x(\mathbf{r}) e^{-2\pi i \boldsymbol{\xi}\cdot\mathbf{r}}\, d\mathbf{r}, \qquad \mathcal{F}^{-1}\{y\}(\mathbf{r}) = \int_{\mathbb{R}^D} y(\boldsymbol{\xi}) e^{2\pi i \boldsymbol{\xi}\cdot\mathbf{r}}\, d\boldsymbol{\xi}. \tag{11}$$

在这一表述中，大尺度结构对应于低频分量（ $\|\boldsymbol{\xi}\|$较小），而小尺度的域特定变异对应于高频分量。引入截止频率 ω_c后，我们定义分解如下:

$$\mathcal{B}^S = \{z : \|\boldsymbol{\xi}\| < \omega_c\}, \qquad \mathcal{B}^D = \{z : \|\boldsymbol{\xi}\| \geq \omega_c\}. \tag{12}$$

因此，共享结构 $\mathcal{B}^S$ 对应于两个域共有的低频骨干，而 $\mathcal{B}^D$ 则捕获域特定变率。

自动截止频率选择。截止频率 ω_c 定义了共享内容与域特定内容之间的边界。为了以数据驱动的方式估计 ω_c，我们引入一种简单的基于判别器的准则。

对于候选 ω_c，我们对每个样本应用低通滤波器:

$$x^S = \mathcal{F}^{-1}[\chi_{<\omega_c} \cdot \mathcal{F}(x)], \tag{13}$$

其中，$\chi_{<\omega_c}$ 是频率 $\|\boldsymbol{\xi}\| < \omega_c$的示性函数。我们训练卷积神经网络分类器 D_ψ，使用标准的二元交叉熵损失来判别 x^S 源自 $\mathcal{D}_A$ 还是 $\mathcal{D}_B$:

$$\mathcal{L}_D(\psi;\omega_c) = -\mathbb{E}_{x_A\sim\mathcal{D}_A}\left[\log D_\psi(x_A^D)\right] - \mathbb{E}_{x_B\sim\mathcal{D}_B}\left[\log\left(1 - D_\psi(x_B^D)\right)\right]. \tag{14}$$

从较高的截止频率 ω_c 开始，我们逐步降低其数值。只要分类器仍能可靠地区分不同域（即准确率显著高于50%），低频表示中就仍包含域特定信息。最优截止频率 ω_c^* 定义为使判别器准确率接近随机猜测的最小值:

$$\mathrm{Acc}(D_\psi;\omega_c^*) \approx 0.5. \tag{15}$$

这一停止准则确保 $\mathcal{B}^S$ 包含最大的域不变成分，同时排除域特定特征。作为合理性检查，我们验证在未应用任何滤波时（$\omega_c \to \infty$），判别器能够实现完美区分，从而确认网络具备足够的判别能力。

实际工作流程。选定 ω_c^* 后，通过将一个域的高频内容替换为噪声来生成伪配对，并使用这些配对训练生成模型 f_θ，使其学习到到目标域的条件映射。SerpentFlow 的伪代码见算法1。

算法1 SerpentFlow：基于频率的伪配对生成、对齐与推理

要求： 源域 $\mathcal{D}_A$、目标域 $\mathcal{D}_B$、判别器 D_ψ、生成器 f_θ**1**：初始化截止频率 $\omega_c \leftarrow \omega_{\max}$

2：**重复**

3：应用低通滤波器：

$$x^S = \mathcal{F}^{-1}\big[\chi_{<\omega_c} \cdot \mathcal{F}(x)\big]$$

4: 训练并计算判别器准确率 Acc(D_ψ; ω_c)

5: 降低截止频率 ω_c

6: **直到**Acc(D_ψ; ω_c) ≈ 0.5

7: 设置 $\omega_c^* \leftarrow \omega_c$

▷ **阶段 1：伪配对生成与训练**

8: **对于**每个样本 $x_B \in \mathcal{D}_B$ **执行**

9: 采样噪声 $\epsilon \sim \mathcal{N}(\mathbf{0},\ \mathbf{I})$

10: 构造伪输入：

$$\tilde{x}_B = \mathcal{F}^{-1}\big[\chi_{<\omega_c^*}\mathcal{F}(x_B) + \chi_{\geq\omega_c^*}\mathcal{F}(\epsilon)\big] \tag{16}$$

11: 训练生成流程 f_θ 以重建 x_B：

$$\mathcal{L}_G(\theta) = \ell(f_\theta, \tilde{x}_B, x_B) \tag{17}$$

12: **结束循环**

▷ **阶段 2：在源域上进行推理** $\mathcal{D}_A$

13: **对**每个样本 $x_A \in \mathcal{D}_A$ **执行**

14: 样本 $\epsilon \sim \mathcal{N}(\mathbf{0},\ \mathbf{I})$

15: 构造伪样本：

$$\tilde{x}_A = \mathcal{F}^{-1}\big[\chi_{<\omega_c^*}\mathcal{F}(x_A) + \chi_{\geq\omega_c^*}\mathcal{F}(\epsilon)\big] \tag{18}$$

16: 生成对齐的目标域输出：

$$\hat{x}_A = \mathcal{G}(f_\theta, \tilde{x}_A) \tag{19}$$

17: **结束循环**

3.3 生成模型的选择

从理论上讲，SerpentFlow 不受 f_θ 所用生成方法选择的限制。原则上，f_θ 可以通过生成对抗网络、扩散模型、归一化流或其他生成范式来实现，只要它能够学习将伪配对分布 p_0 映射到目标域 p_B。在本研究中，我们采用流匹配方法 [6],，它提供了确定性的连续时间生成传输流程。流匹配定义了一个神经速度场 $f_\theta(x,\ t)$，该速度场根据常微分方程（ODE）控制样本的演化：

$$\frac{dx_t}{dt} = f_\theta(x_t, t), \qquad x_0 \sim p_0, \tag{20}$$

其中，f_θ 是可学习的部分，p_0 是伪样本的分布：

$$\tilde{x}_B = \mu^{-1}(z_B^S + z_\epsilon^D),$$

其中，z_B^S 是共享低频分量，z_ϵ^D 是随机高频噪声。该流将源域样本从 p_0 连续传输至目标分布 p_B。我们定义一条线性插值路径（原则上也可以使用其他路径）：

$$x_t = (1-t)\tilde{x}_B + t x_B, \qquad \dot{x}_t^* = x_B - \tilde{x}_B,$$

通过对 t进行偏斜采样，强调变率较高的区域。流匹配目标为

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_t\,\mathbb{E}_{(\tilde{x}_B, x_B)\sim(p_0,p_B)}\Big[\|f_\theta(x_t,t) - \dot{x}_t^*\|_2^2\Big] = \ell(f_\theta, \tilde{x}_B, x_B), \tag{21}$$

该目标训练 f_θ 以近似所选路径上的瞬时位移。

在推理过程中，一个源 样本 $x_A \in \mathcal{D}_A$ 被映射到傅里叶潜空间 $\mathcal{Z}$，并提取其共享成分 z_A^S i s extracted。一个伪 s 随后构造样本

$$\tilde{x}_A = \mu^{-1}(z_A^S + z_\epsilon^D).$$

通过沿常微分方程对速度场进行积分，得到生成的样本

$$\frac{dx_t}{dt} = f_\theta(x_t, t), \quad x_{t_0} = \tilde{x}_A, \tag{22}$$

使用数值求解器从 $t_0 = 1$到 $t_1 = 0$，记为 ODS：

$$\hat{x}_A = \mathrm{ODS}(\tilde{x}_A, f_\theta, t_0 = 1, t_1 = 0) = \tilde{x}_A + \int_{t_1=1}^{t_0=0} f_\theta(x_t, t)\, dt = \mathcal{G}(f_\theta, \tilde{x}_A), \tag{23}$$

生成 一个与 p_B 一致的样本，同时保留源域的低频结构 .

备注。 在基于分数的生成方法中，例如流匹配或扩散模型，将低频骨干结构 $\mathcal{B}^S$ 与随机高频分量 $\mathcal{B}^D$ 相加是一种自然的做法。用于构造从高斯先验到数据的随机路径的噪声调度，通常会先去除高频内容，再影响低频 *[26, 30],* ，这与我们的分解方式一致： $\mathcal{B}^S$ 中的低频结构得以保留，而 $\mathcal{B}^D$ 中的高频细节则以随机方式生成。因此，该模型自然地遵循底层数据结构，同时允许出现逼真的高频变率。

4 实验

我们在合成和物理场景中评估 SerpentFlow，以考察它对齐具有共同大尺度结构、但细尺度成分不同的域的能力。在本文所述实验中，我们聚焦于无监督空间超分辨率，旨在根据粗分辨率观测重建高分辨率场。一般而言，该框架也可用于时间信号或时空数据，例如插值或细化时间序列或动力场（附录D提供了一个时间序列超分辨率的简单示例）。我们使用了三个数据集：（i）一个受控的合成图像数据集 [33]，称为带背景图像的旋转MNIST（MRBI），其中高频内容已被移除；（ii）一个流体模拟数据集 [18, 26], ；以及（iii）一项现实的气候降尺度任务：将耦合模式比较计划第 6 [34], 阶段（简称 CMIP6）的全球环流模式数据映射至 ECMWF 再分析数据 v5 [35],（亦称 ERA5），即法国的风场。

4.1 实验设置

对于所有实验，生成模型 f_θ 采用基于 U-Net 的架构实例化，并在第3节所述的频率条件框架内进行训练。共享域成分对应低频带，而域特定变异则通过随机的高频实现引入。除非另有说明，模型均使用 Adam优化器进行训练，学习率为 10^{-4}，批量大小为 32，并采用偏斜时间采样$t \sim$ Skewed(0，1)，以强调分布变率较大的区域。实际中，偏斜采样实现为

$$t = \frac{1}{1+\sigma}, \quad \sigma = \exp(P_{\text{std}} \cdot \epsilon + P_{\text{mean}}), \quad \epsilon \sim \mathcal{N}(0, 1), \tag{24}$$

其中 $P_{\text{mean}} = -1.2$ 和 $P_{\text{std}} = 1.2$，且 t 裁剪至 $[10^{-4}, 1]$。这使训练集中于流匹配路径上的中间区域，在这些区域中，速度场表现出更大的变化。

在推理期间，伪样本通过学习到的速度场 f_θ 进行传输，方法是对常微分方程（式 20）进行积分。在我们的实现中，我们使用 Dormand-Prince 5(4)方法 [36] 进行自适应积分。

基线方法。 我们将我们的方法与以下用于非配对域对齐的最先进方法进行比较：

- **Diffusion Bridge** ——该方法出自《使用扩散桥对流体流进行非配对降尺度》 [26],，利用基于扩散的潜在桥进行无配对数据转换，见图10f。
- **Dual FM** ——双扩散隐式桥 [15], 利用两个扩散过程和一个中间高斯潜在空间，实现域之间的双向传输，见图10d。在其原始论文中，Dual FM 展现出了优于早期图像转换方法（如 CycleGAN[12] 或 AlignFlow [14]）的性能。

我们还注意到，其他范式——例如随机插值框架 [16] 或桥匹配算法 [17]——虽然从理论角度看颇具前景，但实践中已发现它们难以稳定，且

难以在高维生成式传输任务 [37, 38], 中可靠收敛，因此我们没有将其纳入基线方法。

所有基线方法均在流匹配设置下进行训练（因此双扩散隐式桥基线方法被命名为 Dual FM），使用相同的 UNet 架构和完全一致的超参数，包括训练轮数、批量大小、扩散路径长度和学习率调度器。对于 Diffusion Bridge 基线方法，我们使用分类器优化插值参数 t^*，以确定源域和目标域的低频分量变得无法区分的时刻。为评估鲁棒性，我们还测试了 $t^* - 0.1$ 和 $t^* + 0.1$。

4.2 动画数据集上的合成频率重建

第一个实验基于修正旋转背景图像（MRBI）数据集 [33], , 该数据集是修正美国国家标准与技术研究院（MNIST）数字 [39] 的一个变体，其中手写数字叠加在随机的自然图像背景上。每个样本都是大小为（1，28，28）的灰度图像。

设置。我们从 MRBI 数据集中构建两个不同的域：

- $\mathcal{D}_A$：低通滤波的 MRBI 图像，其中高频分量已被去除；
- $\mathcal{D}_B$：包含完整频谱的原始 MRBI 图像。

滤波在傅里叶域中进行，使用截止频率 ω_c，以分离共享的低频结构 $\mathcal{B}^S$，同时去除域特定的高频细节 $\mathcal{B}^D$。因此，来自 $\mathcal{D}_A$ 的样本仅保留数字粗略的大尺度内容，而来自 $\mathcal{D}_B$ 的样本则保留完整的频率组成，包括精细细节。

对两个域均采用随机训练集—测试集划分。生成模型 f_θ 经训练后，将来自 $\mathcal{D}_A$ 的滤波样本映射到 $\mathcal{D}_B$ 中相应的全分辨率对应图像，从而根据可用的低频信息重建缺失的高频内容。这种实验设置构成了我们框架的理想化实例，其特点是低频与高频清晰分离、傅里叶表示具有双射性，并且可以直接访问共享潜在子空间。

评估。为了定量评估重建质量，我们使用两个互补的分类器。令 $y \in \{0, \ldots, 9\}$表示数字标签。第一个分类器 h_ϕ 是一个在 ImageNet-1K 上预训练、再在 MRBI 上微调的 ResNet-18 [40] ，旨在最大化分类准确率；其训练采用交叉熵损失：

$$\mathcal{L}_{\mathrm{CE}}(\phi) = \mathbb{E}_{(x,y)\sim\mathcal{D}_B}\big[-\log h_\phi(y \mid x)\big]. \tag{25}$$

该分类器用于评估结构保留：由于 MRBI 的背景主要由高频成分构成，数字编码了必须保留的大尺度结构。我们报告生成样本 $f_\theta(\tilde{x}_A)$ 上的 h_ϕ 准确率，并将其与真实 MRBI 图像上的准确率进行比较。第二个分类器 d_ψ 是一个基于卷积编码器的网络，用于评估生成样本是否属于 $\mathcal{D}_B$。我们先训练该分类器区分来自 $\mathcal{D}_\mathrm{A}$ 和 $\mathcal{D}_\mathrm{B}$的样本，据此确定其架构和超参数；这一初步调优使准确率接近完美，并确保分类器具备足够的能力来辨别高频成分。确定架构和超参数后，我们重新训练 d_ψ，使其使用逻辑回归目标函数区分生成样本 $\mathcal{G}(f_\theta,\ \tilde{x}_A)$ 与真实 $\mathcal{D}_\mathrm{B}$ 样本：

$$\mathcal{L}_{\mathrm{disc}}(\psi) = -\mathbb{E}_{x\sim\mathcal{D}_B}[\log d_\psi(x)] - \mathbb{E}_{\hat{x}\sim\mathcal{G}(f_\theta,\mu^{-1}(\mathcal{B}_A^S\oplus\mathcal{B}_\epsilon^D))}[\log(1-d_\psi(\hat{x}))]. \tag{26}$$

在评估期间，准确率接近 50%（随机猜测水平）表明生成样本与来自 $\mathcal{D}_\mathrm{B}$ 的示例无法区分。为避免过拟合导致误导性结论，我们在初步调优后固定模型和超参数的选择。我们在表1中报告了两个分类器的准确率。这两个度量指标共同考察重建的互补方面：大尺度语义结构（数字身份）的保留情况，以及域特定精细细节的真实性。

我们在 MRBI 数据集上进行了一项小规模消融研究，其中低分辨率图像经过滤波，截止频率为 $w_c = 4$。我们考虑了几种情形：

1. **最优情形：** 训练和推理均使用经过 $w_c = 4$滤波的数据。 2. **更强的推理滤波：** 使用 $w_c = 4$ 进行训练，使用 $w_c = 3$ 进行推理。 3. **更强的训练滤波：** 使用 $w_c = 3$进行训练，使用 $w_c = 4$进行推理。 4. **两者均采用更强滤波：** 训练和推理均使用 $w_c = 3$。

这些设置使我们能够评估训练和推理过程中不匹配或更强的低频滤波对数字重建和域对齐的影响。

表1：MRBI 在数字和域分类上的基线比较。对于数字分类器，在真实 MRBI 数据集上的表现为：平均准确率 96.94%，平均置信度 98.23%。

方法	数字分类($\uparrow, h_\phi$)		域分类($\downarrow, d_\psi$)
	平均准确率（%）	平均置信度（%）	准确率（%）
Dual FM	35.85	85.44	0.90
Diffusion Bridge $t^\star = 0.6$	41.35	88.62	0.98
Diffusion Bridge $t^\star = 0.5$	36.78	88.71	0.97
Diffusion Bridge $t^\star = 0.4$	30.04	88.62	0.96
SerpentFlow $w_c = 4$，无限。 $w_c = 4$	**88.63**	**95.42**	**0.50**
SerpentFlow $w_c = 4$，无限。 $w_c = 3$	20.83	81.18	0.81
SerpentFlow $w_c = 3$，无限。 $w_c = 4$	<u>75.67</u>	<u>91.97</u>	0.60
SerpentFlow $w_c = 3$，无限。 $w_c = 3$	63.22	89.19	<u>0.53</u>
SerpentFlow 条件， $w_c = 4$	28.90	79.38	1.00

结果。 表1清晰地表明，当模型使用正确的截止频率 w_c 时，重建数字的可辨识度显著高于使用其他方法时的结果。在真实 MRBI 数字上训练的分类器，在我们的最优配置生成的样本上达到了 88% 的准确率，而最佳竞争基线方法仅为 40%，显示出显著的提升。当模型使用过于激进的频率滤波进行训练、但在全分辨率数据上进行评估时，性能有所下降，但仍远高于所有基线方法。这种鲁棒性令人鼓舞，因为在许多现实场景中，超分辨率是一个病态问题，即划分不同域的确切截止频率是未知的。相比之下，在推理时应用过度滤波会导致结果急剧恶化，这表明在生成时保留适当范围的频率至关重要。

图 1 进一步细化了这些观测结果。从定性角度来看，与 Dual FM 或 Diffusion Bridge 不同，我们的方法能够保留原始图像的低频结构；后两者往往会改变全局结构，有时甚至会错误生成数字本身。此外，我们的方法成功重建了 MRBI 数据集中典型的高频背景纹理，而 Dual FM 和 Diffusion Bridge 则无法做到这一点。这一特性与许多应用尤为相关，因为在这些应用中，信息通常分布于整个空间域，而非集中在局部区域（自然图像数据集往往就是如此）。因此，准确建模前景和背景内容对于下游分析至关重要。

最后，我们尝试了该方法的一个条件变体，其中生成器的输入由两个通道组成：一个包含白噪声，另一个包含滤波数据。这种配置的表现显著逊于其他版本所采用的加性组合，这表明，在重建缺失频率时，显式条件化不如我们的隐式融合策略有效。

噪声添加的影响。为了探究随机性对高频重建的影响，我们通过采样独立的高斯噪声分量 z_ϵ^D，为每个低通输入生成多个预测结果。对于每个输入，我们生成五个添加非零噪声的预测结果，以及第六个将噪声设为零的预测结果。我们展示生成的图像及其相对于真实值的偏差（偏差）。

添加噪声使模型能够探索多种合理的高频细节，这反映出在现实应用中，高频内容可能不存在唯一的真实值，而是服从某种统计分布。在最优配置下（模型使用 $w_c = 4$ 训练并使用 $w_c = 4$ 推理，见图 2a 和图 2b），不同噪声实现生成的图像之间差异很小，偏差图显示其与真实值仅有细微偏差。对于模型使用 $w_c = 3$ 训练并使用 $w_c = 4$ 推理的情况（见图 2c 和图 2d），生成图像的多样性要高得多：一个预测结果可能与真实值非常相似，而其他结果则可能存在较大偏差。无噪声预测的背景完全平滑，凸显了随机采样对于重建逼真的高频细节至关重要。

4.3 流体模拟数据集

我们考虑 [26], 引入的流体模拟数据集，该数据集基于理想化的二维平流–凝结模型 [41]。该数据集由未配对的低分辨率（64×64）和高分辨率（512×512）场组成，表示两种物理变量：涡量和过饱和度。高分辨率场取决于

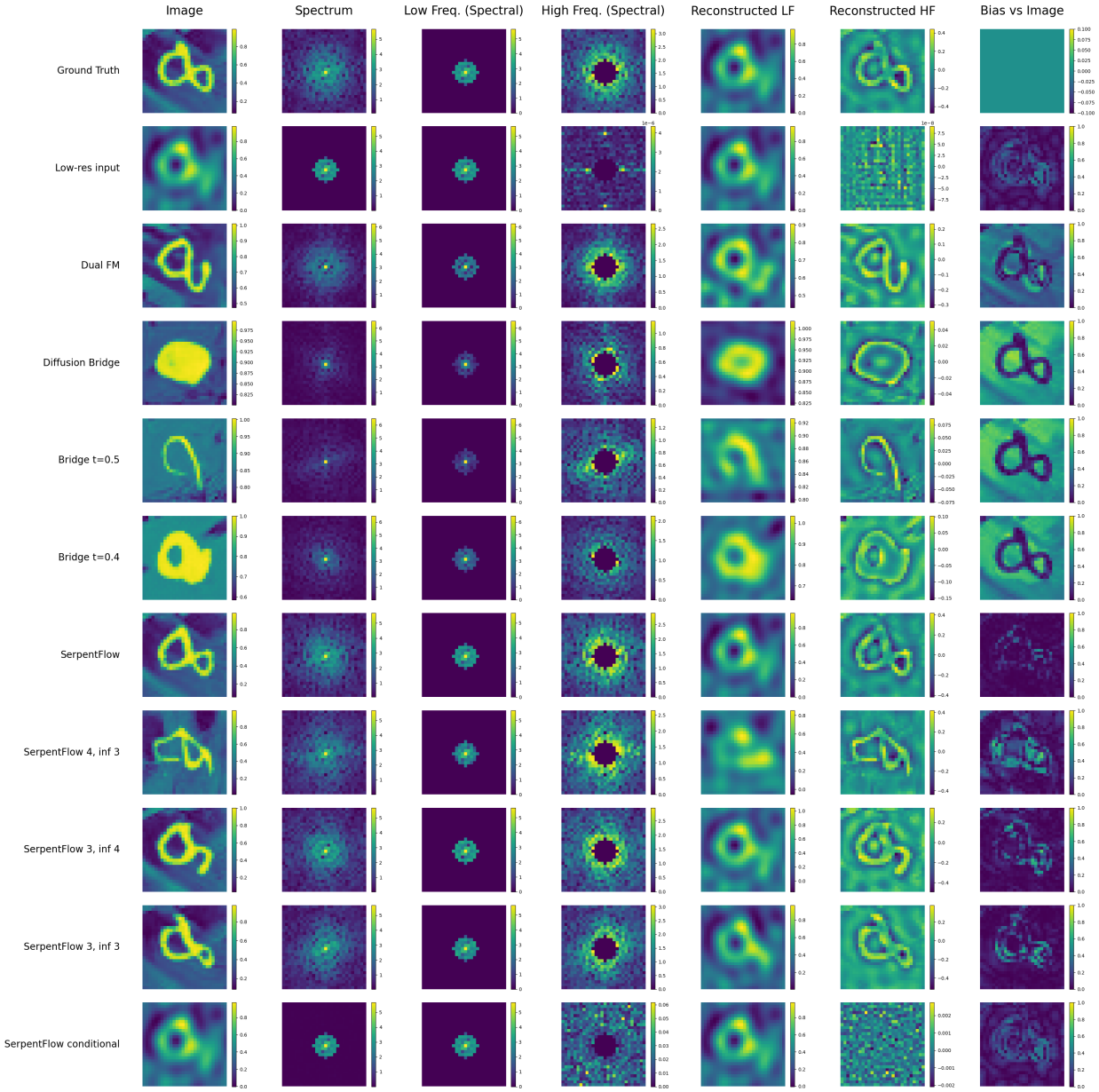
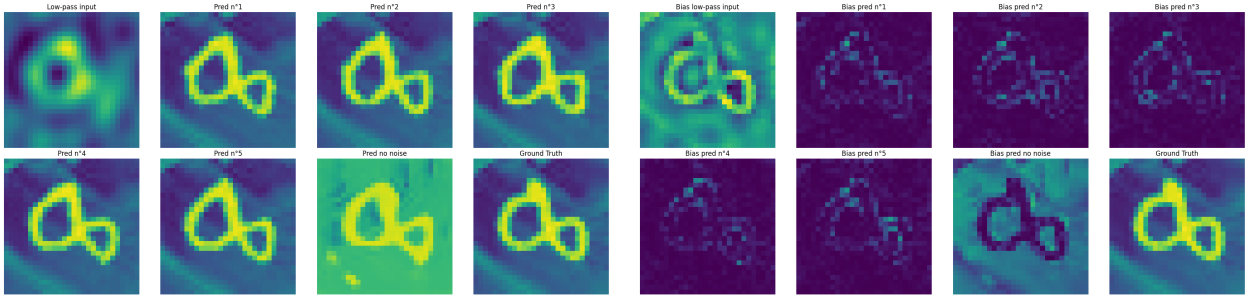
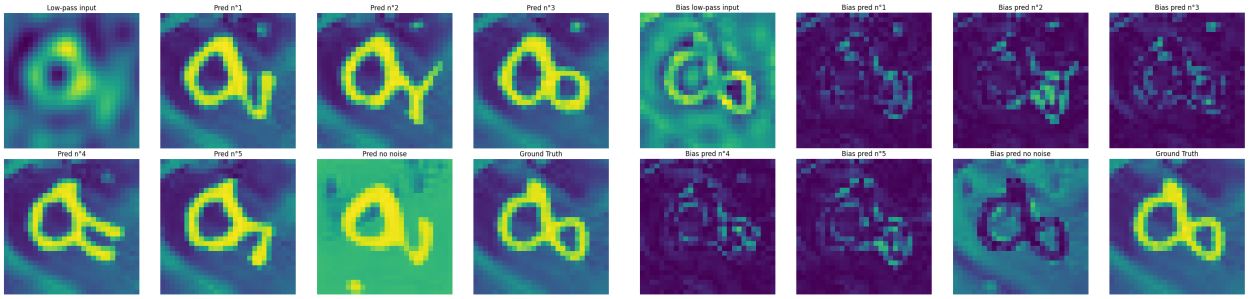


图 1：MRBI 重建比较，针对一张标记为 8 的图像，含有 $w_c = 4$。



(a) 模型 $w_c = 4$，推理 $w_c = 4$

(b) 偏差图，模型 $w_c = 4$，推理 $w_c = 4$



(c) 模型 $w_c = 3$，推理 $w_c = 4$

(d) 偏差图，模型 $w_c = 3$，推理 $w_c = 4$

图 2：随机高频分量对 MRBI 重建的影响。上排：最优配置（使用 $w_c = 4$ 训练的模型，使用 $w_c = 4$ 进行推理），下排：使用更强的低频滤波进行训练：$w_c = 3$，但使用 $w_c = 4$ 进行推理。左侧：生成的图像，包括多个带噪预测结果和一个不带噪声的预测结果。右侧：与真实值对应的偏差。

波数 $k_x = k_y \in \{2.0, 4.0, 8.0, 16.0\}$。低分辨率数据集和各个版本的高分辨率数据集均包含约 2，000 次迭代。

设置。 对于每个波数 k，我们根据这些模拟定义两个域：

- $\mathcal{D}_A =$ 低分辨率模拟；
- $\mathcal{D}_{B_k} =$ 波数为 k.的高分辨率模拟

重要的是， $\mathcal{D}_A$ 并不是通过简单地下采样高分辨率场得到的。相反，它来自经过最近邻插值和低通滤波器上采样的低分辨率模拟。因此， $\mathcal{D}_A$ 的频谱内容大致覆盖 $\mathcal{D}_{B_k}$的低频子空间。这一预先存在的设置与我们的方法天然契合：可以从 $\mathcal{D}_A$ 中提取共享的低频结构，并结合域特定高频分量，在 $\mathcal{D}_{B_k}$中生成逼真的高分辨率样本。由于计算限制，我们采用了缩小后的规模：原始 512×512 场经过平均处理，得到适用于低分辨率和高分辨率数据集的 64×64 。引入到涡量谱中的波数 k 会产生一个频率强度峰值，其强度与 k 成正比；而在未以 k为参数化的低分辨率数据中并不存在该峰值。因此，为选择截止频率 w_c，我们取能够切除该频率强度峰值的最大半径，以尽可能保留更多低频。该数据集也提供了 $k \in [0, 1]$，但由于该峰值与数据集中的最低频率过于接近，我们几乎不得不摧毁全部内容才能保留它。在附录 C.2 中，我们展示了截取得较少时的方法结果：低频数据的结构会丢失；而截取得较多时，该峰值无法重建。

评估。 图 3 展示了 $k = 8.0$ 的基线性能的初步概览。尽管三种基线方法生成的重建结果看起来都与高分辨率数据（$\mathcal{D}_{B_8}$）颇为相似，但就这个示例而言，从定性角度看，只有 SerpentFlow 能够在涡量和过饱和度两方面始终保留输入信号的整体结构。

为了更精确地定量分析性能，我们考察了重建结果的时间演化，以及场的分布和频谱性质。

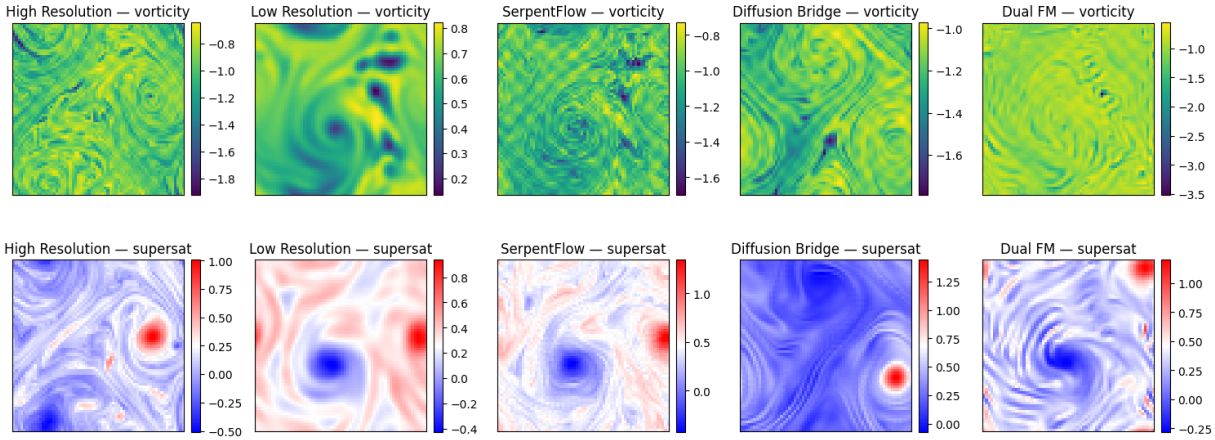
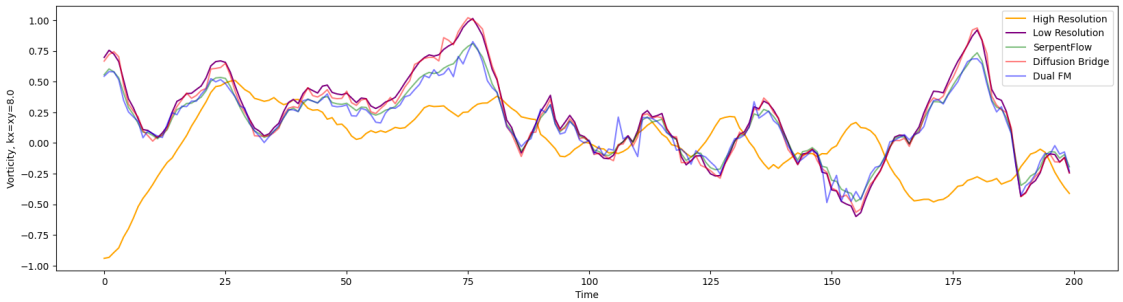
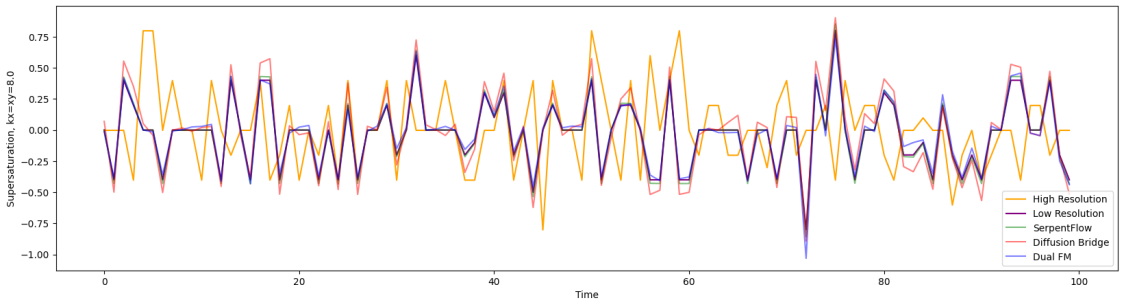


图 3：涡量和　　给定迭代次数下的过饱和度场，适用于 $\mathcal{D}_{B_8}$、$\mathcal{D}_{\mathrm{A}}$，以及基线方法的重建结果　　.



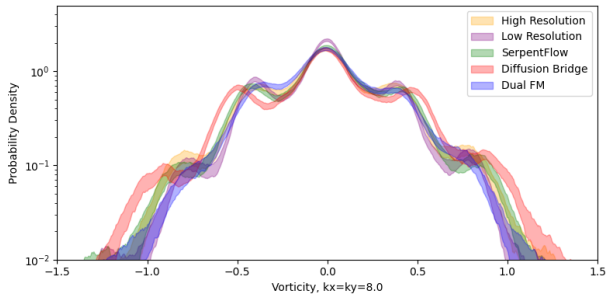
(a) 前 200 个样本的涡量时间演化（$k_x = k_y = 8$）。



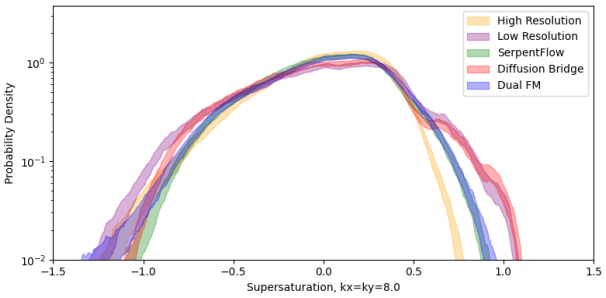
(b) 前 100 个样本的过饱和度时间演化（$k_x = k_y = 8$）。

Figure 4：　$k_x = k_y = 8$ 低分辨率场和高分辨率场的空间平均时间演化　　.

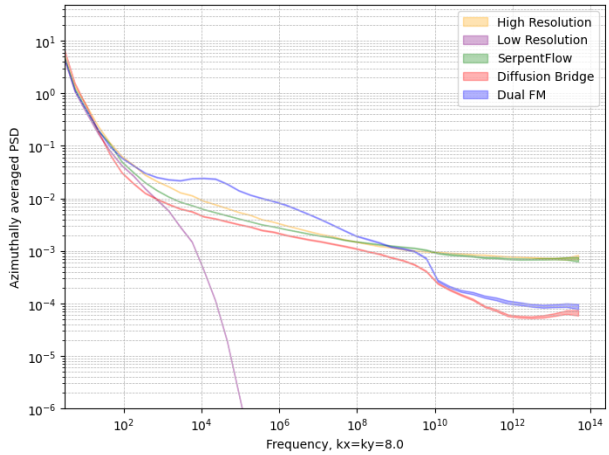
如图4所示，尽管所有方法总体上都遵循低分辨率数据的时间动态，但我们的重建结果与原始信号最为接近，证实了我们的方法能够保留时间动态。



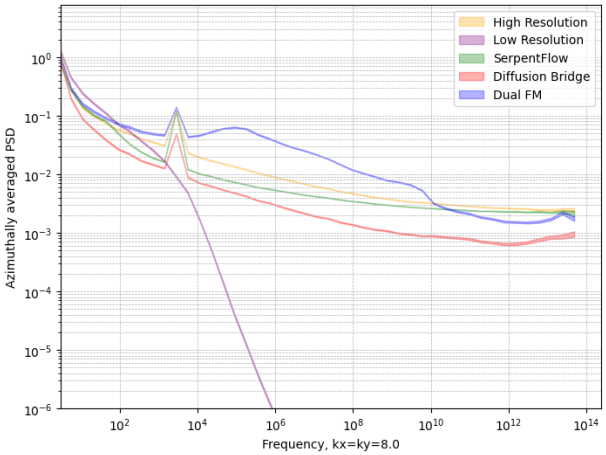
(a) 涡量场密度估计。



(b) 过饱和度场密度估计。



(c) 涡量场谱密度。



(d) 过饱和度场谱密度。

图5：低分辨率场和高分辨率场及基线方法的密度与谱密度比较（$k_x = k_y = 8$）。

关于向 $\mathcal{D}_{B_8}$ 的转换，图5展示了各基线方法对高分辨率统计特征的再现程度。图5a和图5b表明，Dual FM 和我们的重建结果与高分辨率场的分布非常接近，而 Bridge Matching 基线方法产生的尾部过重。在谱密度方面（图5c和图5d），所有方法恢复的高频内容都与高分辨率场相当。不过，对于涡量，Diffusion Bridge 和 Dual FM 会产生一些伪影，而我们的方法则忠实地遵循高分辨率谱。对于过饱和度，我们的方法仍优于低分辨率输入，但在高频范围内仍存在一些差异。

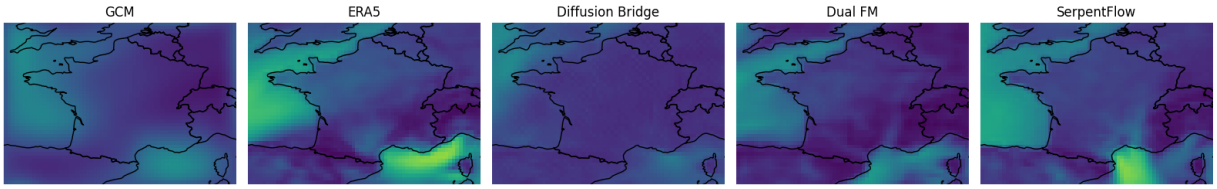
4.4 气候降尺度：从 CMIP6 到 ERA5 风场

数据呈现。全球环流模式（GCM）模拟气候变量，同时再现其关键统计特征，例如主导模态、变率，以及极端事件的发生频率或重现期。GCM 尤其适合生成时空一致的大尺度气候统计特征；然而，其计算成本使其无法在精细空间分辨率下运行。因此，要获取区域或局地尺度的有用信息，就需要采用降尺度技术。由于 GCM 确实包含关键统计特征，降尺度的目标并非预测精确的局地数值，而是对其统计特征进行降尺度。为此设计的方法称为概率降尺度方法（PDM）。评估 PDM 时需与观测进行比较，而观测应被理解为 GCM 动力学的一种可能实现，而非绝对的“真实”状态。

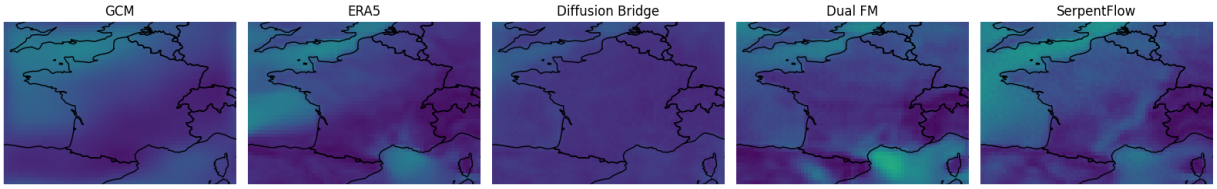
在此，我们考虑一个GCM，即第六阶段耦合模式比较计划（CMIP6 [34]）中的ACCESS地球系统模式 [42], , 该模式以日时间分辨率和粗空间分辨率提供气候模拟。作为观测参考，我们使用ERA5再分析数据 [35], , 其空间分辨率更高，频率为小时。为确保与GCM的时间一致性，我们计算ERA5场的日平均值。在本研究中，我们仅关注法国的每日风强度。我们的目标是对GCM输出进行空间降尺度

将其从原生分辨率（1.875×1.25 度）转换至ERA5网格（0.25×0.25 度），同时保持每天一个值。

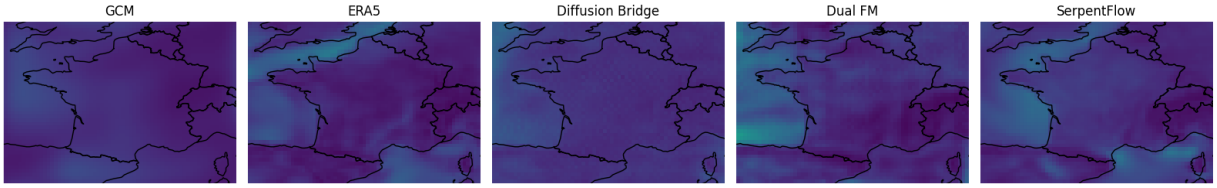
实验设置。 虽然 GCM 和 ERA5 数据表示相同的物理量，但它们在网格尺度上并未对齐。由于构造方式不同，即使两者采用相同的分辨率，GCM 也比观测到的 ERA5 数据具有更强的空间相关性和更平滑的场。这意味着，尽管两个域之间的大尺度低频结构大体相同，小尺度高频变率却有所不同，并且各自具有域特异性。我们的低频／高频分解特别适用于这种情形：可以从 GCM 中提取并保留低频骨干。相比之下，可以通过随机重建高频分量，使其与 ERA5 中观测到的变率相匹配。这使我们无需显式的配对数据即可进行降尺度，并确保生成的高分辨率风场既符合物理规律，又与 ERA5 观测在统计上保持一致。模型使用 1981年至2001年的数据进行训练，而验证数据涵盖2002年至2022年，确保验证期为二十年，以便计算气候统计量。我们训练了第3.2节中提到的分类器，以选择 $w_c^\star$。分类器返回的值为 1，因此可以将尺度大于约 1200 km 的影响保留为低频。附录B进一步讨论了该分类器。



(a) 日期：2004年4月11日。



(b) 日期：2011年8月4日。

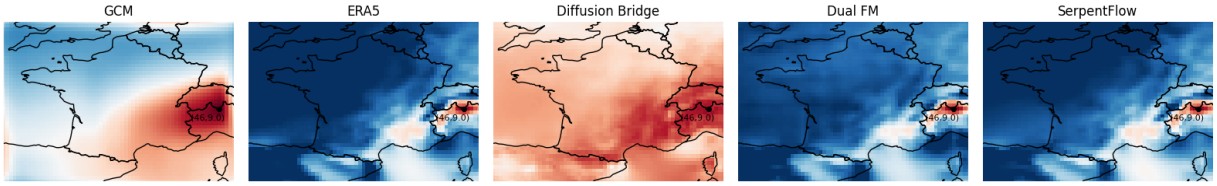


(c) 日期：2018年6月10日。

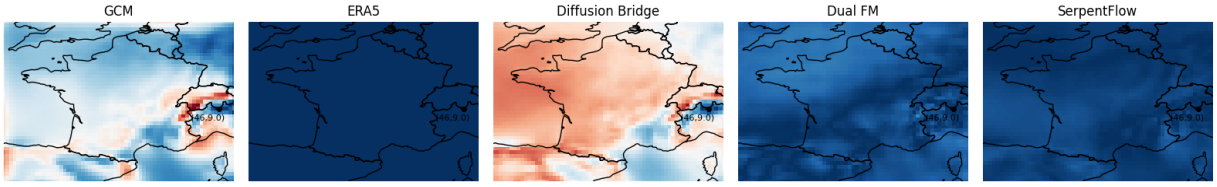
图6：展示测试集中多个日期的基线方法降尺度结果。

评估。 图6初步直观展示了基线方法的降尺度性能。Dual FM 和 SerpentFlow 似乎改进了 GCM 的风场，而 Diffusion Bridge 则似乎进一步平滑了气候模型中本已高度相关的数据。

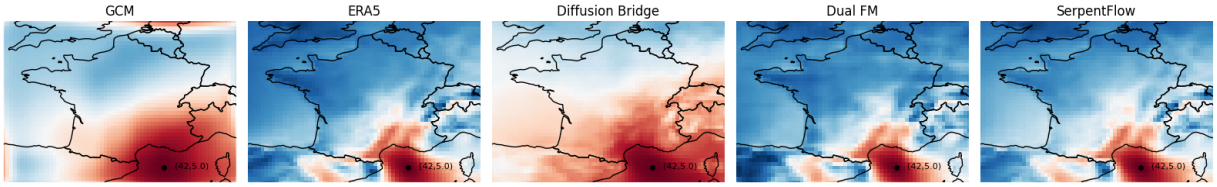
为评估我们的降尺度所实现的空间细化程度，我们比较了验证期内计算得到的相关性图与 ERA5 的相关性图。对于一个参考网格点（纬、经），通过计算其20年时间序列与数据集中所有其他点的时间序列之间的相关性，即可生成相关性图。图7a和7c分别展示了阿尔卑斯山和地中海某个点的相关性图示例，而图7b和7d则显示了相对于 ERA5 的对应偏差。ERA5 与 GCM 之间的差异十分显著：ERA5 的图清晰地反映了地形，而 GCM 的场过度相关且过于平滑，因此这些图可作为评估降尺度方法的有力判别依据。Dual FM 和我们的 SerpentFlow 都生成了与 ERA5 十分相似的相关性图，捕捉到了地形引起的空间结构，而 Diffusion Bridge 的输出未能重现这种结构。在偏差图中，尤其是对于地中海的参考点，Dual FM 似乎比 SerpentFlow 略接近 ERA5。通过对所有网格点的相关性偏差取平均，我们得到表2中报告的总体相关性得分。Dual FM 的得分高出



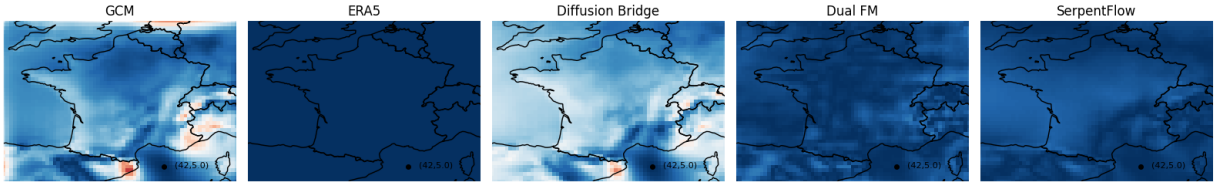
(a) 相关性图，坐标 (46,9)。



(b) 相对于 ERA5 的偏差相关性图，坐标 (46,9)。



(c) 相关性图，坐标 (42,5)。



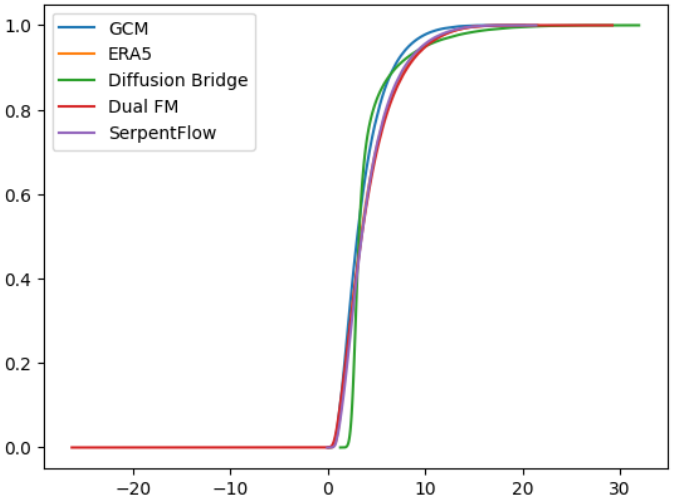
(d) 相对于 ERA5 的偏差相关性图，坐标 (42,5)。

图 7：风强度场的相关性图。我们选取一个关注点，并计算该时间段内相应时间序列与其他网格点时间序列之间的相关性。对于每种方法，我们将相关性图与 ERA5 的相关性图进行比较。图 7a 和 7b 中，参考网格点位于阿尔卑斯山。图 7c 和 7d 中，参考网格点位于地中海。

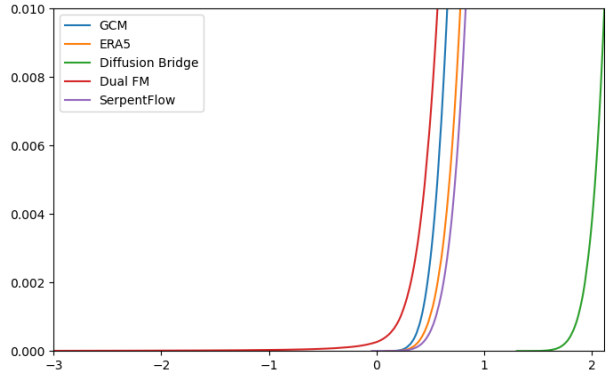
15%，优于 SerpentFlow，这证实了它在重建目标域空间变异性方面的能力更强 $\mathcal{D}_B$。

我们还从图8中观察到 Dual FM 和 SerpentFlow 方法具有出色的重建表现。图中展示了基线方法与 ERA5 和 GCM 的累积分布函数（CDFs）。回顾一下，随机变量 X 的累积分布函数 $F(x)$ 定义为 X 取值小于或等于 x 的概率，即 $F(x) = \mathbb{P}(X \leq x)$。图8a显示，Dual FM 和 SerpentFlow 均能成功调整 GCM 分布，使其接近 ERA5 的分布，两者的累积分布函数几乎重合。相比之下，Diffusion Bridge 与目标分布的偏差更大。观察下尾部（图8b）可见，Dual FM 产生了少量负伪影，而 SerpentFlow 与 ERA5 的吻合度更高。在极端上尾部（图8c），SerpentFlow 的表现略逊一筹，而 Dual FM 仍与 ERA5 高度吻合。表2报告的 Kolmogorov–Smirnov（KS）得分体现了这一性能差异，该得分衡量两个累积分布函数之间的最大距离。Dual FM 的 KS损失略低于 SerpentFlow，表明其对目标分布的整体重建效果更好。

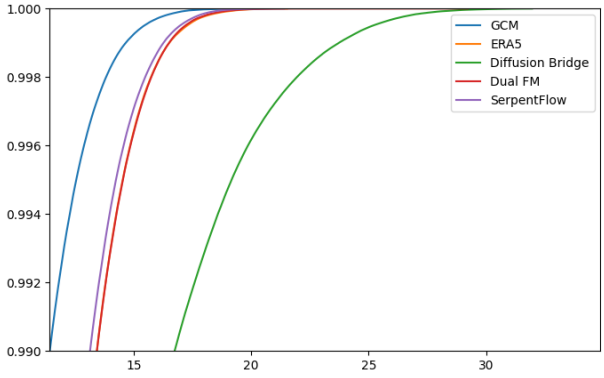
到目前为止，我们只评估了基线方法重建属于 $\mathcal{D}_B$ 的数据的能力，尚未评估它们保留 $\mathcal{D}_A$ 结构的能力。虽然这种保留能力不一定能通过直接观察立即显现，但图6让我们能够比较 GCM 与基线方法的空间平均的归一化时间动态。在图 9a 中，我们展示了基线方法、ERA5 和 GCM 的归一化空间年度均值，以呈现年际变异性。目标是尽可能紧密地跟随 GCM 信号。我们观察到，Diffusion Bridge



(a) 累积分布函数（累积分布函数）。



(b) 分位数放大图 < 0.01。



(c) 分位数放大图 > 0.99。

图8：基线方法与 GCM 和 ERA5 的风强度累积分布函数。

和 Dual FM 难以跟踪该信号，尤其是在 2005 年前后，而 SerpentFlow 几乎完美地跟随了该信号。

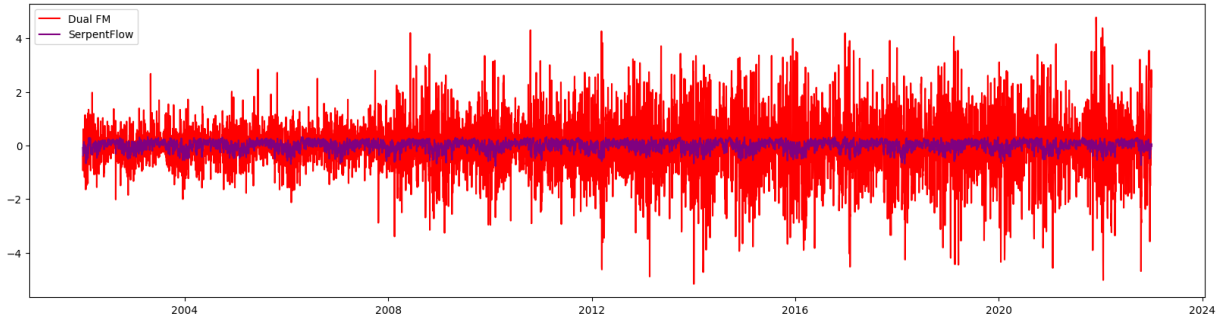
考察日信号时，这一差异更加显著。图9b展示了原始GCM信号与 Dual FM 和 SerpentFlow 重建值之间的偏差。显然，在所有时间步上，Dual FM 生成的数据都比 SerpentFlow 与原始信号相差更大。我们通过计算这些信号之间的均方根误差（RMSE）来量化这一差异，结果见表2；就这一指标而言，SerpentFlow 的误差比 Dual FM 低一个数量级。除 RMSE 外，我们还报告 Nash–Sutcliffe效率指数（NSE，[43]），这是一种地球物理建模中的标准指标，用于评估重建结果相对于参考序列气候平均值对时间变异性的再现程度。NSE 为 1 表示完美重建，而接近 0 或为负的值则意味着其表现不优于——甚至劣于——用序列的时间平均值替代该序列。就这一指标而言，SerpentFlow 的表现也明显更好，得分从 Dual FM 的 0.29 提升至 0.98。因此，尽管 Dual FM 在重建 in $\mathcal{D}_B$ 数据方面略胜 SerpentFlow，但这是以破坏原有气候结构为代价的。在单变量降尺度应用中，保留 GCM 动力学至关重要，因为这些动力学体现了此类模拟的主要科学价值。

5 结论

在本研究中，我们介绍了 SerpentFlow（生成式域适应的共享结构分解），这是一种用于非配对域对齐的生成式框架，通过将数据进行合理分解，划分为共享成分和域特定成分。通过分离共享结构，并将域特定内容替换为随机的



(a) 归一化年均值。



(b) 相对于归一化空间平均 GCM 的时间偏差

图 9：基线方法相较于气候模型动力学的时间一致性示意图。

表2：气候数据集上的降尺度性能评估。KS：生成分布与目标分布之间的 Kolmogorov–Smirnov 统计量；相关性得分：相对于 ERA5 的平均相关性偏差；时间 RMSE：验证期内空间平均日风强度的均方根误差；NSE：基于空间平均每日序列计算的 Nash–Sutcliffe效率指数（1 = 完美， 0 = 均值预测器）。

方法	**KS** (↓)	**相关性得分**(↓)	**时间RMSE** (↓)	**NSE** （↑）
Diffusion Bridge	0.25119	0.293	1.779	-0.029
Dual FM	**0.02495**	**0.045**	1.226	0.291
SerpentFlow	0.02525	0.052	**0.031**	**0.982**

变异，SerpentFlow 生成伪配对，使得即使没有真实配对数据，也能使用条件生成模型。该分解完全自适应且由数据驱动，无需预先了解应如何分离结构，因此适用于各种不同类型的域。

我们在三种不同的数据集上，将 SerpentFlow 应用于无监督超分辨率任务：合成图像（MRBI）、模拟物理过程（流体模拟）以及气候降尺度（从 CMIP6 到 ERA5风场）。在这些任务中，共享成分对应低频内容，而高频细节则体现域特定变率。用于区分低频分量和高频分量的截止频率，通过基于分类器的判据从数据集中自动确定，从而确保分解过程完全自动化且稳健。通过保留共享结构并以随机方式建模域特定变异，SerpentFlow 能够成功重建域特异性特征，同时保持与共享模式的一致性。在所有实验中，它在保真度、统计一致性和物理可解释性方面均优于现有的非配对生成方法。

除了这些实验之外，SerpentFlow 还提供了一个适用范围广泛的框架。共享–特异性分解可扩展至时间域或时空域，从而实现时间序列、视频或其他结构化信号的高分辨率重建。重要的是，如果缺乏关于应如何进行分解的明确先验知识，使用 SerpentFlow 将无法奏效（见附录 C.1），而 Dual FM 等最先进的方法仍是强有力的基线方法。

未来的工作可以探索其他 μ 函数，例如模糊或加噪策略，以及其他分解策略，例如基于小波的多尺度分析，或包括主成分分析、独立成分分析和学习得到的嵌入在内的数据驱动方法，从而有望提升灵活性和可解释性。总而言之，SerpentFlow 提供了一套稳健、全自动且通用的非配对域对齐流程，将自适应共享–特异性分解与条件生成建模相结合，在兼顾全局结构和局部变化的同时，实现不同域之间的桥接。

参考文献

[1] George Papamakarios、Eric Nalisnick、Danilo Jimenez Rezende、Shakir Mohamed 和 Balaji Lakshminarayanan。用于概率建模与推理的归一化流。机器学习研究期刊，22(57):1–64，2021年。[2] Diederik P Kingma 和 Max Welling。自编码变分贝叶斯。*arXiv* 预印本 *arXiv:1312.6114*，2013年。[3] Ian J Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。生成对抗网络。神经信息处理系统进展，27，2014年。[4] Jonathan Ho、Ajay Jain 和 Pieter Abbeel。去噪扩散概率模型。神经信息处理系统进展，33:6840–6851，2020年。[5] Yang Song、Jascha Sohl-Dickstein、Diederik P Kingma、Abhishek Kumar、Stefano Ermon 和 Ben Poole。通过随机微分方程进行基于分数的生成建模。载于国际学习表征会议，2021年。[6] Yaron Lipman、Ricky TQ Chen、Heli Ben-Hamu、Maximilian Nickel 和 Matthew Le。用于生成建模的流匹配。载于第十一届国际学习表征会议，2023年。[7] Zhiyao Kong、Wei Ping、Jing Huang、Kexin Zhao 和 Bryan Catanzaro。Diffwave：一种用于音频合成的通用扩散模型。载于国际学习表征会议（*ICLR*），2021年。[8] John Ingraham、Vijay Garg、Regina Barzilay 和 Tommi Jaakkola。用于基于图的蛋白质设计的生成模型。载于神经信息处理系统进展，第32卷，2019年。[9] Bowen Jing、Stefan Eismann、Prafulla Suriana、Richard Townshend 和 Ron Dror。利用扩散模型学习功能性蛋白质序列。*arXiv* 预印本 *arXiv:2210.09238*，2022年。[10] Mehdi Mirza 和 Simon Osindero。条件生成对抗网络。*arXiv* 预印本 *arXiv:1411.1784*，2014年。[11] Lvmin Zhang、Anyi Rao 和 Maneesh Agrawala。为文生图扩散模型添加条件控制。载于*IEEE/CVF* 国际计算机视觉会议论文集，第3836–3847页，2023年。[12] Jun-Yan Zhu、Taesung Park、Phillip Isola 和 Alexei A Efros。使用循环一致性对抗网络进行无配对图像到图像转换。载于*IEEE* 国际计算机视觉会议论文集，第2223–2232页，2017年。[13] Ming-Yu Liu、Thomas Breuel 和 Jan Kautz。无监督图像到图像转换网络。神经信息处理系统进展，30，2017年。

[14] Aditya Grover、Christopher Chute、Rui Shu、Zhangjie Cao 和 Stefano Ermon。Alignflow：通过归一化流从多个域进行循环一致学习。载于AAAI人工智能大会论文集，第34卷，第4028–4035页，2020年。[15] Xuan Su、Jiaming Song、Chenlin Meng 和 Stefano Ermon。用于图像到图像转换的双扩散隐式桥。载于第十一届国际学习表征会议，2023年。[16] Michael S Albergo、Nicholas M Boffi 和 Eric Vanden-Eijnden。随机插值：统一流与扩散的框架。arXiv预印本 arXiv:2303.08797，2023年。[17] Yuyang Shi、Valentin De Bortoli、Andrew Campbell 和 Arnaud Doucet。扩散薛定谔桥匹配。神经信息处理系统进展，36:62183–62223，2023年。[18] Valentin De Bortoli、Iryna Korshunova、Andriy Mnih 和 Arnaud Doucet。用于非配对数据转换的薛定谔桥流。神经信息处理系统进展，37:103384–103441，2024年。[19] Ricky TQ Chen、Yulia Rubanova、Jesse Bettencourt 和 David K Duvenaud。神经常微分方程。神经信息处理系统进展，31，2018年。[20] Leon A Gatys、Alexander S Ecker 和 Matthias Bethge。使用卷积神经网络进行图像风格迁移。载于IEEE计算机视觉与模式识别会议论文集，第2414–2423页，2016年。[21] Tero Karras、Samuli Laine 和 Timo Aila。用于生成对抗网络的基于风格的生成器架构。载于IEEE/CVF计算机视觉与模式识别会议论文集，第4401–4410页，2019年。[22] Yuan Yuan、Siyuan Liu、Jiawei Zhang、Yongbing Zhang、Chao Dong 和 Liang Lin。使用循环中循环生成对抗网络进行无监督图像超分辨率。载于IEEE计算机视觉与模式识别会议研讨会论文集，第701–710页，2018年。[23] Emmanuel J Candès、Justin Romberg 和 Terence Tao。稳健不确定性原理：从高度不完整的频率信息中精确信号重建。IEEE信息论汇刊，52(2):489–509，2006年。[24] Ashish Bora、Ajil Jalal、Eric Price 和 Alexandros G Dimakis。使用生成模型进行压缩感知。载于国际机器学习会议，第537–546页。PMLR，2017年。[25] Brian Groenke、Luke Madaus 和 Claire Monteleoni。Climalign：通过归一化流对气候变量进行无监督统计降尺度。载于第十届国际气候信息学会议论文集，第60–66页，2020年。[26] Tobias Bischoff 和 Katherine Deck。使用扩散桥对流体流进行非配对降尺度。地球系统人工智能，3(2):e230039，2024年。[27] Xingchao Liu、Chengyue Gong 和 Qiang Liu。让流动笔直且快速：学习通过校正流生成和转换数据。载于第十一届国际学习表征会议（ICLR），2023年。[28] John P Boyd。《切比雪夫与傅里叶谱方法》。Dover，2001年。[29] Claudio Canuto、M Yousuff Hussaini、Alfio Quarteroni 和 Thomas A Zang。《谱方法：单区域基础》。Springer，2006年。[30] Fabian Falck、Teodora Pandeva、Kiarash Zahirnia、Rachel Lawrence、Richard Turner、Edward Meeds、Javier Zazo 和 Sushrut Karmalkar。从傅里叶空间视角看扩散模型。arXiv预印本 arXiv:2505.11278，2025年。[31] Nathan Doumèche、Francis Bach、Gérard Biau 和 Claire Boyer。物理信息引导的核学习。机器学习研究期刊，26(1):1–39，2025年。arXiv:2409.13786。[32] Claude E Shannon。噪声环境下的通信。IRE会议录，37(1):10–21，1949年。[33] Hugo Larochelle、Dumitru Erhan、Aaron Courville、James Bergstra 和 Yoshua Bengio。对具有众多变化因素的问题中的深度架构进行实证评估。载于第二十四届国际机器学习会议论文集，第473–480页，2007年。[34] Brian C O’Neill、Claudia Tebaldi、Detlef P Van Vuuren、Veronika Eyring、Pierre Friedlingstein、George Hurtt、Reto Knutti、Elmar Kriegler、Jean-Francois Lamarque、Jason Lowe 等。CMIP6情景模式比较计划（ScenarioMIP）。地球科学模型开发，9(9):3461–3482，2016年。[35] Hans Hersbach、Bill Bell、Paul Berrisford、Shoji Hirahara、András Horányi、Joaquín Muñoz-Sabater、Julien Nicolas、Carole Peubey、Raluca Radu、Dinand Schepers 等。ERA5全球再分析。英国皇家气象学会季刊，146(730):1999–2049，2020年。[36] John R Dormand 和 Peter J Prince。一类嵌入式龙格－库塔公式。计算与应用数学杂志，6(1):19–26，1980年。

[37] Kaiqi Chen、Eugene Lim、Kelvin Lin、Yiyang Chen 和 Harold Soh。从零开始并非必要：通过基于插值过程的策略扩散实现行为精炼。*arXiv* 预印本 *arXiv:2402.16075*，2024 年。[38] Siyi Chen、Yixuan Jia、Qing Qu、He Sun 和 Jeffrey A Fessler。Flowdas：一种基于随机插值过程的数据同化框架。*arXiv* 预印本 *arXiv:2501.16642*，2025 年。[39] Yann LeCun。手写数字的 MNIST 数据库。*http://yann.lecun. co m/exdb/mnist/*，1998 年。[40] 何恺明、张祥雨、任少卿和孙剑。用于图像识别的深度残差学习。载于*IEEE* 计算机视觉与模式识别会议论文集，第 770–778 页，2016 年。[41] P. A. O'Gorman 和 T. Schneider。均匀湍流中水汽输送与凝结运动学的随机模型。大气科学杂志，63：2992–3005，2006 年。[42] Tilo Ziehn、Matthew A Chamberlain、Rachel M Law、Andrew Lenton、Roger W Bodman、Martin Dix、Lauren Stevens、Ying-Ping Wang 和 Jhan Srbinovsky。澳大利亚地球系统模型：ACCESS-ESM1.5。南半球地球系统科学杂志，70(1)：193–214，2020 年。[43] J Eamonn Nash 和 Jonh V Sutcliffe。通过概念模型进行河流流量预报：第一部分——原理探讨。水文学杂志，10(3)：282–290，1970 年。[44] Alan V Oppenheim、Alan S Willsky 和 Hamid Nawab。信号与系统。普伦蒂斯·霍尔，1999 年。

最新技术进展

以下部分回顾了基于非配对分布的域对齐相关既有工作。除非另有说明，我们将待对齐的两个域记为 $\mathcal{D}_A$ 和 $\mathcal{D}_B$，其中的单个样本分别记为 $x_A^i \in \mathcal{D}_A$ 和 $x_B^i \in \mathcal{D}_B$。假设 $\mathcal{D}_A$ 和 $\mathcal{D}_B$ 中的样本分别服从分布 p_A 和 p_B。潜在表示通常记为 $\mathcal{Z}$，其中 $z_A^i \in \mathcal{Z}$对应于 x_A^i，$z_B^i \in \mathcal{Z}$对应于 x_B^i。对于双射函数 f，其逆函数记为 f^{-1}；参数为 θ的神经网络记为 f_θ。

CycleGAN CycleGAN [12]，（见图10a）是一种图像到图像转换框架，旨在通过对抗目标和循环一致性目标，学习两个非配对域之间的映射。该框架训练两个双向映射 $\mathrm{G}^{\mathcal{D}_\mathrm{A}\to\mathcal{D}_\mathrm{B}}$ 和 $\mathrm{G}^{\mathcal{D}_\mathrm{B}\to\mathcal{D}_\mathrm{A}}$，以及与之对应的判别器 $C_{\mathcal{D}_\mathrm{B}}$ 和 $C_{\mathcal{D}_\mathrm{A}}$。映射对 $(\mathrm{G}^{\mathcal{D}_\mathrm{A}\to\mathcal{D}_\mathrm{B}}, C_{\mathcal{D}_\mathrm{B}})$ 学习从 $\mathcal{D}_\mathrm{A}$ 生成 $\mathcal{D}_\mathrm{B}$ 中的样本，使其无法与真实数据区分；而 $(\mathrm{G}^{\mathcal{D}_\mathrm{B}\to\mathcal{D}_\mathrm{A}}, C_{\mathcal{D}_\mathrm{A}})$ 执行逆向转换。每个生成器接收来自源域的输入图像，并试图欺骗目标域的判别器。为了确保两个映射之间的一致性，循环一致性损失确保将图像转换到另一个域后再转换回来时，能够恢复原始输入，即：

$$\tilde{x}_A = \mathrm{G}_{\mathcal{D}_\mathrm{B}\to\mathcal{D}_\mathrm{A}}\left(\mathrm{G}_{\mathcal{D}_\mathrm{A}\to\mathcal{D}_\mathrm{B}}(x_A)\right) \approx x_A, \quad \tilde{x}_B = \mathrm{G}_{\mathcal{D}_\mathrm{A}\to\mathcal{D}_\mathrm{B}}\left(\mathrm{G}_{\mathcal{D}_\mathrm{B}\to\mathcal{D}_\mathrm{A}}(x_B)\right) \approx x_B.$$

该约束对抗训练进行正则化，并防止生成器产生不保留底层内容的任意映射。

UNIT UNIT（无监督图像到图像转换网络） [13], 见图10b，通过引入共享潜在空间假设扩展了非配对域转换。该方法假设两个域 $\mathcal{D}_\mathrm{A}$和 $\mathcal{D}_\mathrm{B}$ 中的对应图像可以通过域特定编码器 $E_{\mathcal{D}_\mathrm{A}\to\mathcal{Z}}$ 和 $E_{\mathcal{D}_\mathrm{B}\to\mathcal{Z}}$映射到共同的潜在表示 $z \in \mathcal{Z}$，使得 $E_{\mathcal{D}_\mathrm{A}\to\mathcal{Z}}\#p_A \approx E_{\mathcal{D}_\mathrm{B}\to\mathcal{Z}}\#p_B \approx \mathrm{p_Z}$，其中 # 表示推前测度。该模型采用两个变分自编码器（VAEs），并结合GAN损失，以确保重建质量和域真实性。形式上，每个域都有一个编码器–解码器对（$E_{\mathcal{D}_\mathrm{A}\to\mathcal{Z}}, G_{\mathcal{Z}\to\mathcal{D}_\mathrm{A}}$）和（$E_{\mathcal{D}_\mathrm{B}\to\mathcal{Z}}, G_{\mathcal{Z}\to\mathcal{D}_\mathrm{B}}$）。从$\mathcal{D}_\mathrm{A}$ 到 $\mathcal{D}_\mathrm{B}$ 的转换通过经由相反生成器进行编码和解码来实现：$\tilde{x}_B = G_{\mathcal{Z}\to\mathcal{D}_\mathrm{B}}\left(E_{\mathcal{D}_\mathrm{A}\to\mathcal{Z}}(x_A)\right)$。

AlignFlow AlignFlow [14]，（见图10c）沿用了 UNIT 的潜在空间理念，通过将生成对抗网络的生成器替换为（可逆）归一化流，改进了 CycleGAN；这些流映射到共享潜在分布 $\mathrm{p_Z}$（通常为高斯分布，不过归一化流并不要求使用高斯先验）。从 $\mathcal{D}_\mathrm{A}$ 到 $\mathcal{D}_\mathrm{B}$ 的转换通过从 $\mathcal{D}_\mathrm{A}$ 到 $\mathcal{Z}$的正向生成器 $\mathrm{G}^{\mathcal{D}_\mathrm{A}\to\mathcal{Z}}$，再结合从 $\mathcal{D}_\mathrm{B}$ 到 $\mathcal{Z}$的生成器 $\mathrm{G}^{\mathcal{D}_\mathrm{B}\to\mathcal{Z}^{-1}} = G_{\mathcal{Z}\to\mathcal{D}_\mathrm{B}}$的逆变换来完成。利用生成器的逆向过程，可以将一个元素从 $\mathcal{D}_\mathrm{B}$ 转移到 $\mathcal{D}_\mathrm{A}$。归一化流的可逆特性确保了循环一致性。

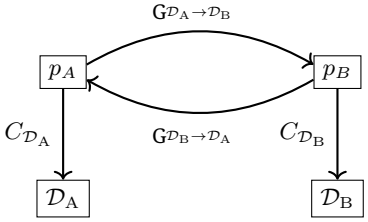
双扩散隐式桥 双扩散隐式桥 [15], 见图10d，它通过移除对抗训练的需求改进了 AlignFlow。为了从 p_A 生成数据到 p_B，需要为每个域训练一个扩散过程。随后，通过从 $\mathcal{D}_\mathrm{A}$ 进行前向扩散，将样本 $x_A \sim p_A$ 传输到 p_B，并将其推送至中间潜在空间 $\mathrm{p_Z}$，再从 $\mathrm{p_Z}$ 进行反向扩散，将其映射到 p_B，从而完成样本迁移。逆变换则可以通过从 p_B 到 $\mathrm{p_Z}$ 的前向扩散，以及从 $\mathrm{p_Z}$ 到 p_A的反向扩散来完成。

Bridge Matching、随机插值、薛定谔桥流 Bridge Matching [17], 随机插值 [16] 和薛定谔桥流 [18],见图 10e, 指的是一种类似的思路：使用基于分数的生成模型，学习从未知分布 p_A 到另一个未知分布 p_B的传输映射；对应的域可能对齐，也可能未对齐。它们的差异可能在于如何定义从 $\mathcal{D}_{\mathrm{A}}$ 中的一个点到 $\mathcal{D}_{\mathrm{B}}$ 中的一个点之间的路径，也可能在于模型实际学习的目标。与 AlignFlow 或 DualDiffusion 不同，这些框架不需要显式的潜变量。整个变换被学习为随机配对之间的动态流，模型会自然地收敛到 p_A 与 p_B之间的双向映射。

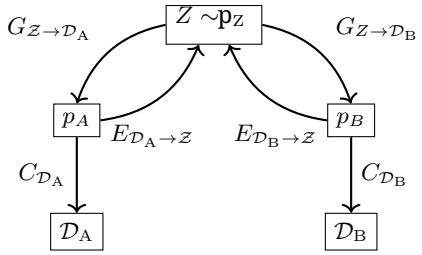
整流流 基于这一类随机桥，整流流 [27] 将 p_A 与 p_B 之间的传输计划表示为一个常微分方程，其流具有良好的不交叉性质。通过递归应用其流框架，他们获得了 $\mathcal{D}_{\mathrm{A}}$ 与 $\mathcal{D}_{\mathrm{B}}$中的数据之间的直线路径。这些直线路径使他们能够在 $\mathcal{D}_{\mathrm{B}}$内生成距离 x_A 最近的点。

用于非配对降尺度的扩散桥 该方法专为物理过程的非配对低分辨率（$\mathcal{D}_{\mathrm{A}}$）与高分辨率（$\mathcal{D}_{\mathrm{B}}$）数据集之间的匹配而设计， [26], 参见图10f。该方法首先训练一个扩散模型，将高斯分布映射到 p_B，然后对 p_B 中的数据加噪，直至优化后的时间步长 $t^\star$ ，再使用扩散模型进行去噪。其思路是，加噪会优先抹去高频，使 $t^\star$ 处的中间空间成为一个域，在该域中，来自 $\mathcal{D}_{\mathrm{A}}$ 和 $\mathcal{D}_{\mathrm{B}}$ 的数据无法区分。

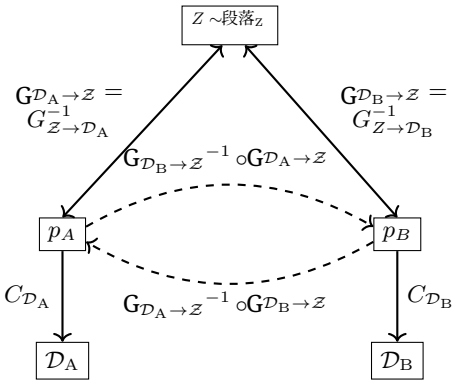
备注。 *SerpentFlow* 与扩散桥相似，但即使在 μ 函数中使用噪声，两者仍有两个主要区别。*(i)* 扩散桥中使用的噪声会影响所有频率，因此无法保证低频得到保留。*(ii) SerpentFlow* 方法更加直接，因为生成流程直接从数据变得无法区分时的共同状态进行训练。而扩散桥使用的是从纯噪声预训练的流程。通过向低分辨率数据添加噪声，我们希望进入扩散路径，但这并无保证。



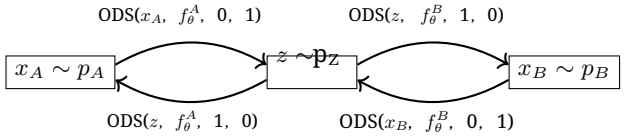
(a) CycleGAN [12]。



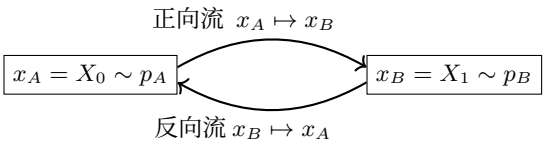
(b) UNIT [13]。



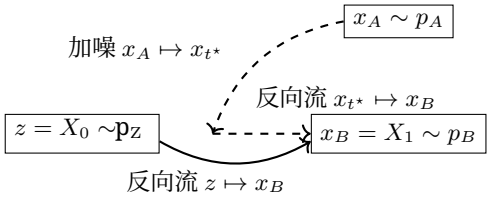
（c）AlignFlow [14]。



（d）双扩散隐式桥 [15], ODS 显式公式如下公式 23。



(e) 随机插值 [16, 18, 17]。



(f) 用于非配对降尺度的扩散桥 [26]。

图10：非配对分布的不同域对齐框架概览。

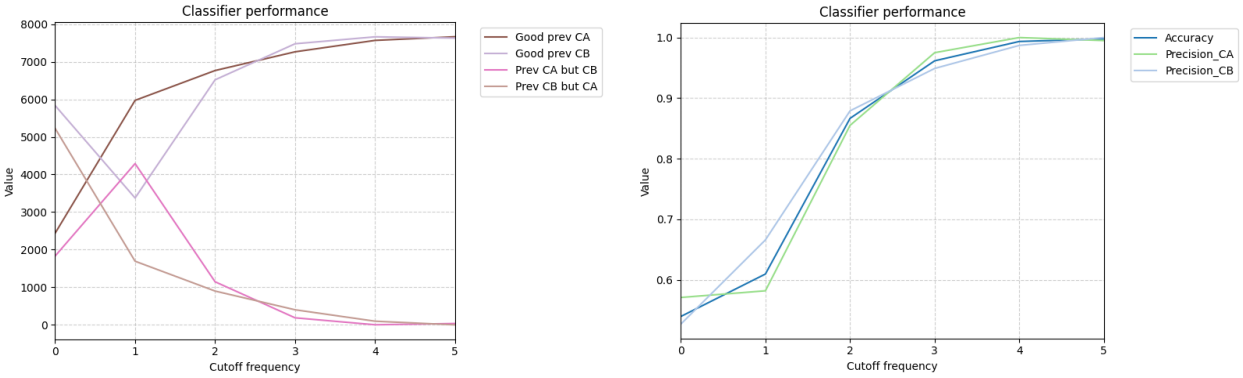
B 关于基于分类器的截止频率选择有效性的讨论

本附录为第3.2节所述的基于判别器的截止频率选择过程提供实证依据。我们以 GCM → ERA5 降尺度实验为例，说明分类器准确率能够可靠地指示频率阈值 ω_c，从而区分共享的大尺度结构与域特定的高频内容。

对于每个候选截止频率 ω_c，我们按照公式 (13) 对所有样本进行低通滤波，并使用公式 (14) 的损失训练域判别器 D_ψ。图11报告了所得分类准确率随 ω_c变化的情况。
我们观察到一个明显的趋势:

- 对于较高的截止频率（$\omega_c \geq 2$），分类器准确率较高，这表明滤波后的样本中仍然存在域特定信息；
- 随着 ω_c 降低，准确率逐渐下降；
- 一旦达到 $\omega_c \leq 1$，分类器准确率便接近 60%，并稳定在接近随机猜测的水平。

按照公式 (15) 的判据，这一现象表明，$\omega_c = 1$ 是使 GCM 和 ERA5 的低频表示变得几乎无法区分的最小截止频率。因此，主要实验采用这一数值。



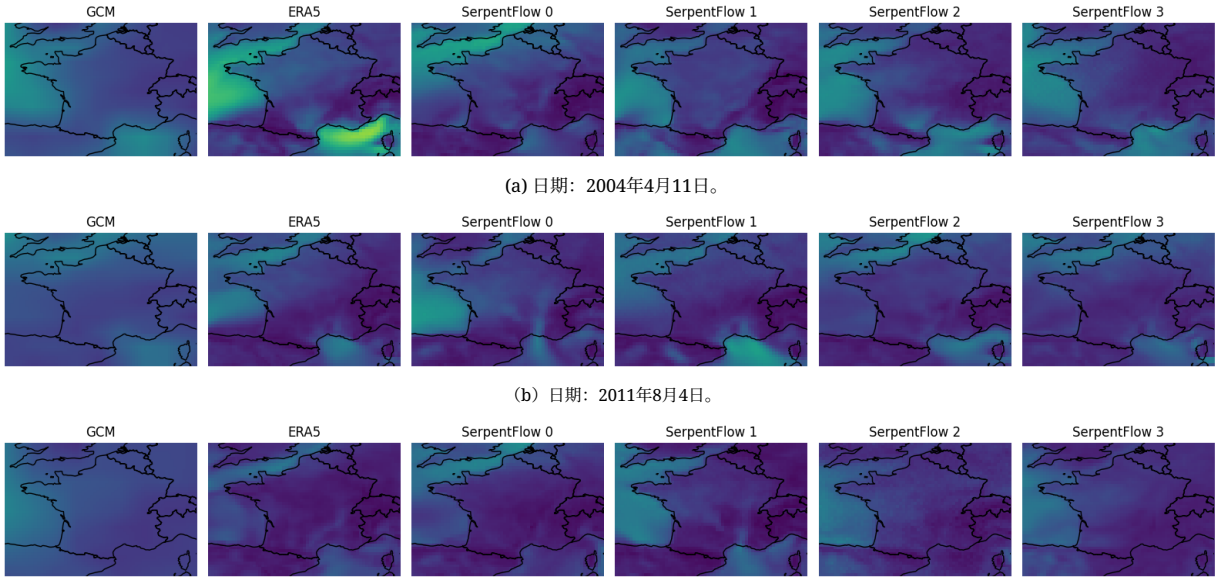
(a) 分类器的错误和正确预测率　　(b) 分类器度量指标

图11：针对不同截止频率，GGCM（数据集A，CA）与ERA5（数据集B，CB）之间的分类器性能。

为验证这一选择是否合理，我们针对不同的截止频率重复完整的降尺度流程

$$\omega_c \in \{0, 1, 2, 3\}$$

并分析生成的重建结果。



(a) 日期：2004年4月11日。

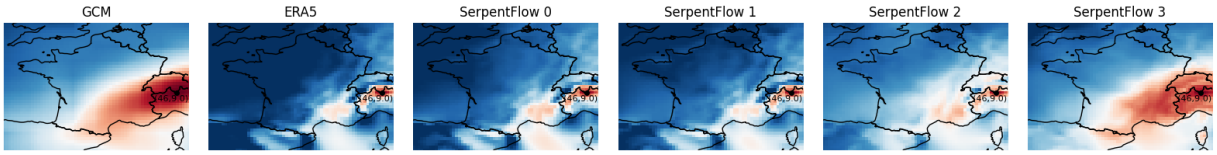
(b) 日期：2011年8月4日。

(c) 日期：2018年6月10日。

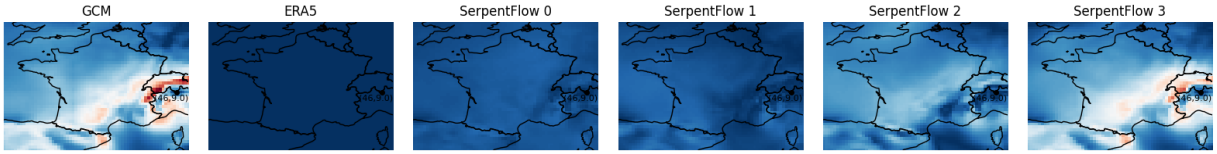
图12：测试集中不同日期的降尺度输出，适用于截止频率为 $\in \{0, 1, 2, 3\}$ 的 SerpentFlow 流程。

截止频率过高（$\omega_c = 2$**或** 3）。在此状态下，假定的共享成分仍包含来自 ERA5 的域特定高频结构。因此，输入 tflow模型的伪输入仍保留不匹配的精细尺度模式。如图13所示，生成样本无法重现 ERA5 的关键空间特征，相关性图与 ERA5 的相关性图不匹配。这证实，过高的截止频率会使本应由模型生成的高频成分通过。

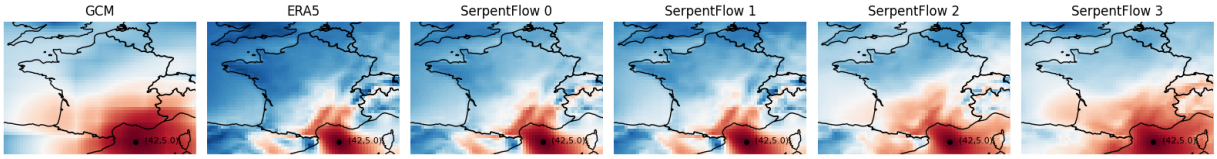
截止频率过低（$\omega_c = 0$）。在此情况下，几乎所有频率都被视为域特定噪声。这会削弱模型保留 GCM 全局时间动态的能力，如图14所示，却无法改善 ERA5 精细尺度细节的重建（见图13）。这表明，过度激进的滤波去除的是信号，而非将不变结构与变率分离。



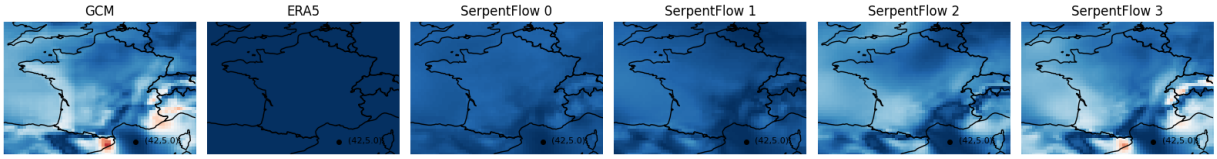
(a) 相关性图，坐标 (46,9)。



(b) 相对于 ERA5 的偏差相关性图，坐标 (46,9)。



(c) 相关性图，坐标 (42,5)。



(d) 相对于 ERA5 的偏差相关性图，坐标 (42,5)。

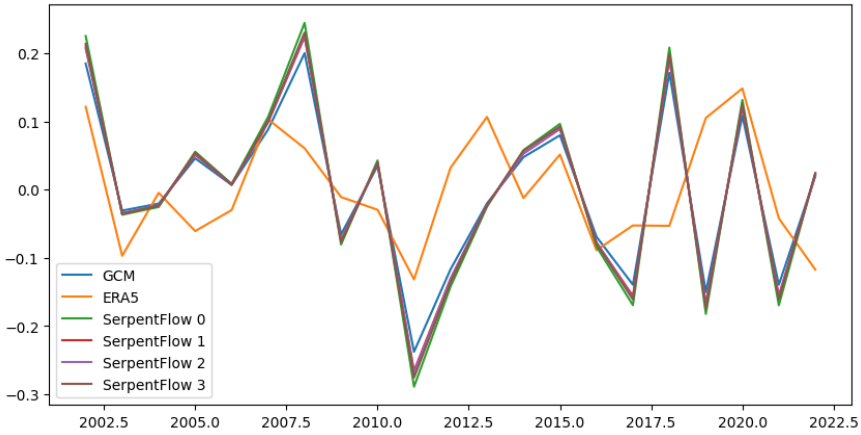
图13：风强度场的相关性图。我们选取一个感兴趣点，并计算该点在所比较时段内对应的时间序列与其他网格点时间序列之间的相关性。对于每种方法，我们都将其相关性图与 ERA5 的相关性图进行比较。图 13a 和图 13b 中，参考网格点位于阿尔卑斯山。图 13c 和图 13d 中，参考网格点位于地中海。

最优截止频率（$\omega_c = 1$）。在此取值下，伪输入保留了 GCM 正确的低频骨干，同时将高频变率交由生成模型处理。重建场呈现出清晰的轮廓、逼真的类 ERA5 小尺度模式，以及一致的大尺度结构。这些结果与基于分类器的判据完全一致，证明了该方法的有效性。表 3 中报告的度量指标也支持这一定性分析。

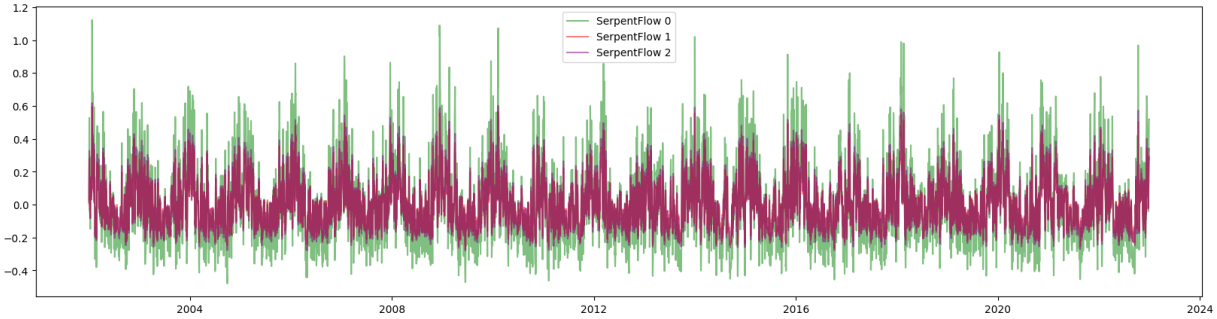
表 3：SerpentFlow 在气候数据集上针对不同截止频率（ω_c）的降尺度性能评估。KS：生成分布与目标分布之间的 Kolmogorov–Smirnov统计量；相关性得分：相对于 ERA5 的平均相关性偏差；时间 RMSE：验证期内空间平均日风强度的均方根误差；NSE：根据空间平均每日序列计算的 Nash–Sutcliffe效率指数（1 = 完美，0= 均值预测器）。

方法	KS (↓)	相关性得分(↓)	时间 RMSE (↓)	NSE (↑)
$\omega_c = 0$	**0.017**	0.084	0.061	0.953
$\omega_c = 1$	0.025	**0.045**	**0.018**	**0.986**
$\omega_c = 2$	0.020	0.088	0.023	0.982
$\omega_c = 3$	0.045	0.119	0.033	0.975

分类器准确率可作为可靠且易于解释的指标，用于识别两个域在何种频带上变得统计上不可区分。这为确定截止频率 ω_c 提供了一种有原则的方法，无需手动调参。GCM →ERA5 任务上的实证结果证实了该方法的有效性：通过判别器选取的截止频率所得到的重建结果明显优于临时选定的替代方案，而指定错误的截止频率则会导致系统性且易于解释的失效模式。



(a) 归一化年均值。



(b) 相对于归一化空间平均 GCM 的时间偏差

图14：与气候模型动力学相比，SerpentFlow 的时间一致性示意图（ $\omega_c \in \{0, 1, 2\}$ ）。

C 局限性：具有不兼容低频的超分辨率

本节展示 SerpentFlow 在低频不起作用时进行超分辨率的局限性。第一个示例基于 MNIST 数据集，第二个示例则基于我们的物理模拟数据集，采用的波数值过低。

C.1 MNIST 数据集

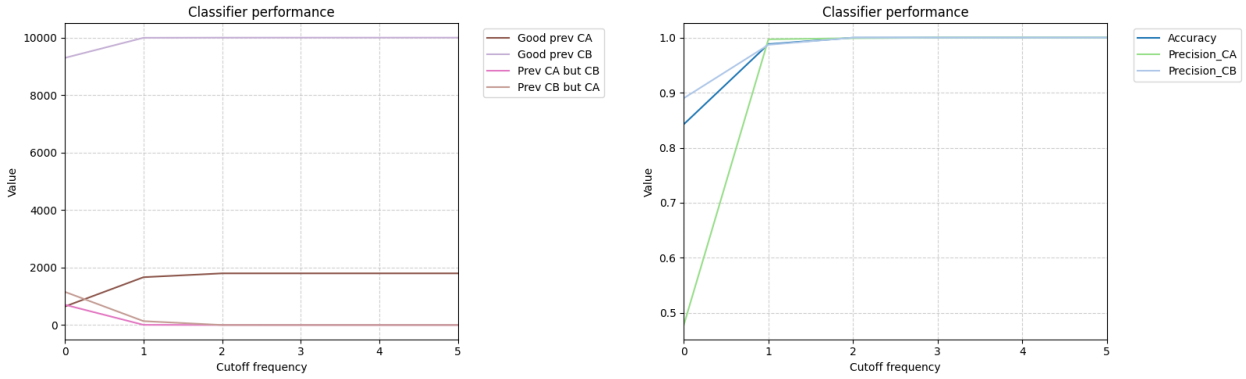
为说明第3节中讨论的频谱兼容性假设的重要性，我们展示了一个在 MNIST 上的受控失败案例。我们将标准 MNIST 数据集（28×28）视为高分辨率域 $\mathcal{D}_B$，并使用来自 sklearn 的 8×8 灰度 MNIST 变体作为低分辨率域 $\mathcal{D}_{\mathrm{A}}$。关键在于， 8×8 图像并非通过对 28×28 数字进行低通滤波获得，而是源自不同的采集与预处理流程。因此，它们的频谱存在根本性的不匹配。

插值基线方法。 我们使用以下三种标准插值方法将 8×8 图像上采样至 28×28 分辨率：

- 最近邻插值，
- 双线性插值，
- 谱插值（在傅里叶域中进行零填充）。

基于分类器的评估。我们训练了论文中讨论的基于卷积的分类器，以区分 MNIST (28×28) 数字及其插值结果，从而找到最优截止频率。即使在分类前对两个域都进行强力低通滤波，判别器在所有

插值方法中的准确率也从未低于 80%。我们在图15中展示了线性插值在不同截止频率下的表现。这种较差的表现表明，即使去除高频成分，插值后的图像在统计特征上仍与真实的 MNIST 样本有很大差异。



(a) 分类器的错误和正确预测率　　(b) 分类器度量指标

图15：对线性插值的 MNIST（数据集A，CA）与 MNIST（数据集B，CB）之间的分类器性能进行评估，截止频率取不同值。

视觉分析。 插值样本呈现出若干伪影：

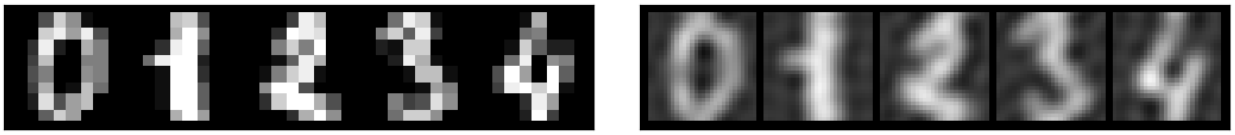
- 模糊且不一致的数字轮廓，
- 伪影背景噪声，
- 粗糙纹理，与 MNIST 特有的笔画几何形状明显不同。

具有代表性的重建结果与 MNIST分布并不相似，这证实了 sklearn-MNIST 中的空间下采样并不等同于对原始数字进行频率截断。

这项实验表明，当低分辨率域不符合高分辨率域的低频结构时，我们基于频率的分解方法会受到限制。当这种兼容性缺失时（如 MNIST 与 sklearn-MNIST 的情况），问题便会变得不适定：没有任何生成模型能够推断出与原始数据集有实质相似性的高分辨率数字，因为粗分辨率观测并不对应于目标分布的低频测量值。

C.2 流体模拟数据集

当 $k_x = k_y$ 过低时 图17展示了当低频谱不匹配时我们的方法所存在的局限。当 k_x，k_y 的值过低时，谱峰会出现在非常靠近中心的位置（即频率非常低处）。在带有 $k_x = k_y = 0$ 的图 17b 和图 17a 中可以看到这一点。此时我们可以选择像图 17a 那样只截去少量内容，但在这种情况下，这些谱峰不会出现在重建结果中，使结果偏离 $\mathcal{D}_{B_0}$ 更远。或者，我们也可以像图 17b 那样在这些谱峰下方进行截断。在这种情况下，剩余数据过少，最终导致重建结果偏离原始数据。另一方面，当这些谱峰离中心足够远时，例如 $k_x = k_y = 8$ （图 17c）和 $k_x = k_y = 16$ （图 17d ），重建结果能够很好地融入这些谱峰，同时保持与原始数据结构接近。



(a) 最近邻插值

(b) 对最近邻插值数据应用低通滤波器



(c) 最近邻插值的预测



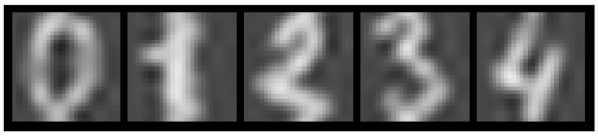
(d) 线性插值



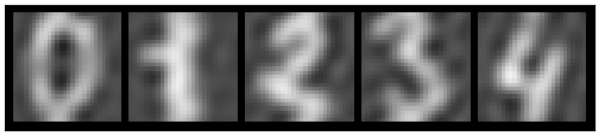
(e) 对线性插值数据应用低通滤波器



(f) 线性插值的预测



(g) 谱插值

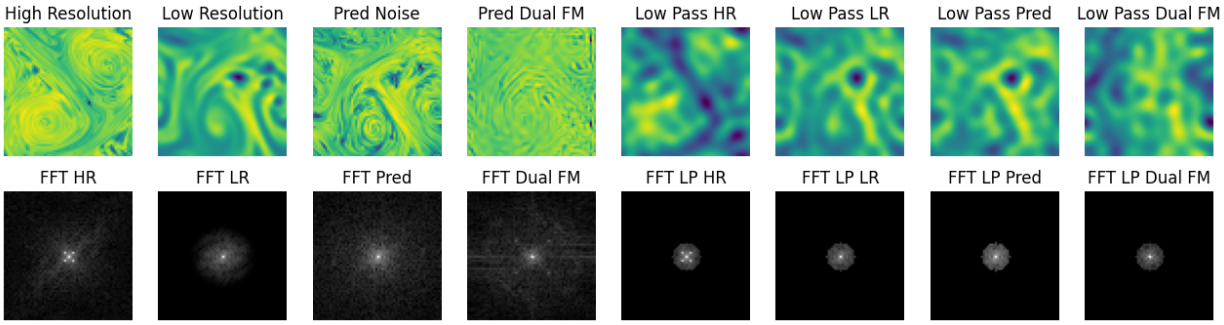


（h）对谱插值数据应用低通滤波器

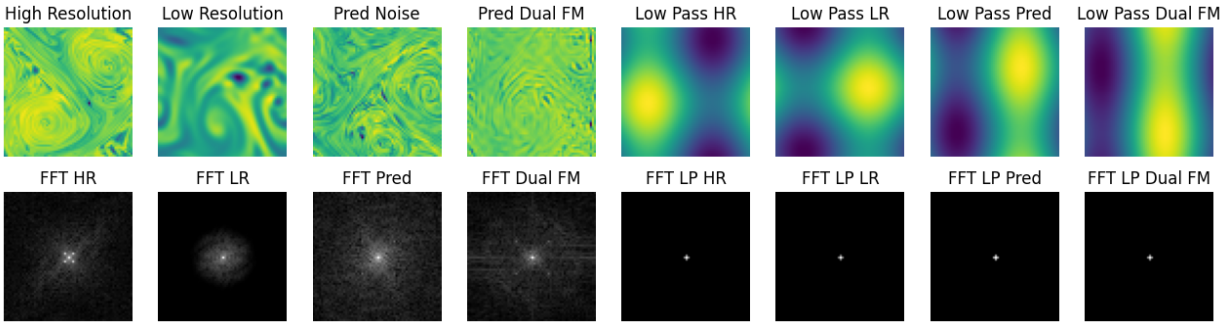


(i) 谱插值的预测

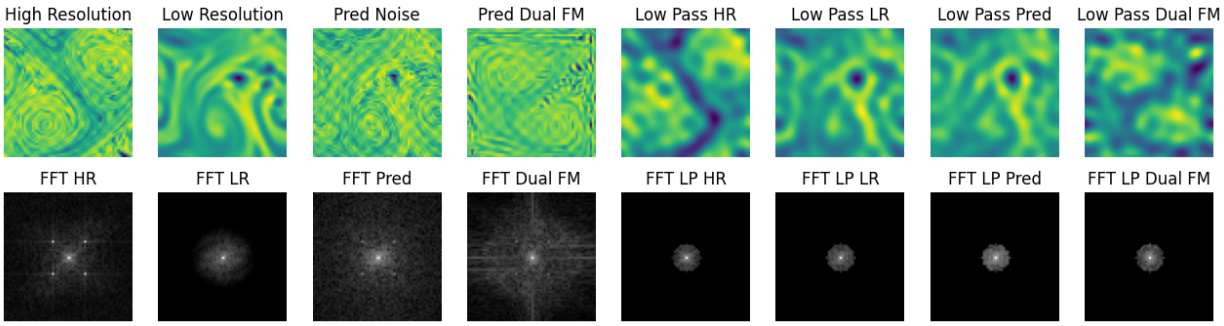
图 16：可视化 sklearn-MNIST 数据集从分辨率（$9times9$）到（28×28）的不同插值、应用低通滤波器后的结果，以及使用 SerpentFlow 转换到 MNIST 域的预测结果。



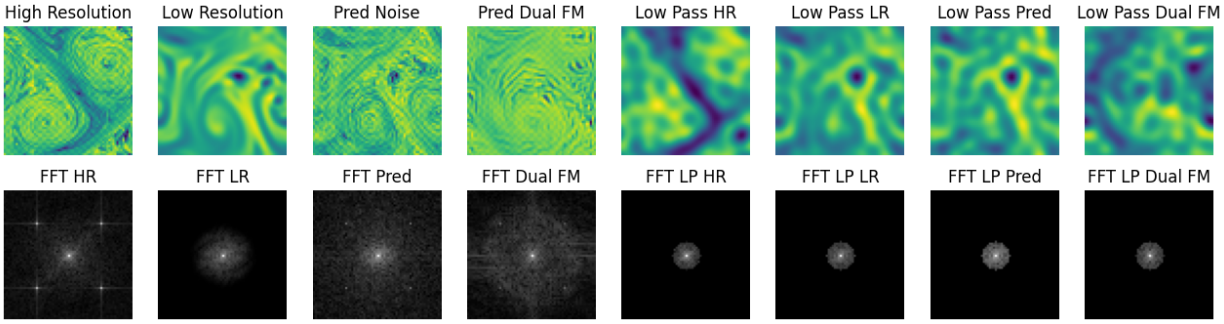
(a) 当 $w_c = 7$ 时， $k_x = k_y = 0$ 的涡量场。



(b) 当 $w_c = 1$ 时， $k_x = k_y = 0$ 的涡量场。



(c) 当 $w_c = 7$ 时， $k_x = k_y = 8$ 的涡量场。



(d) 当 $w_c = 7$ 时， $k_x = k_y = 16$ 的涡量场。

图17：不同 k_x， k_y 和 w_c 值下涡量场的重建结果及其重建结果的谱。

D 时间序列应用

时间超分辨率问题与主论文中的空间场景相呼应：目标是从稀疏、含噪且退化的测量值中重建高分辨率信号。从概念上说，这与经典插值和压缩采集有关，因为模型只能观测到部分时间信息，必须推断缺失的精细尺度结构。

下面，我们将详细介绍合成数据集的构建方式，以及用于模拟低质量传感器的退化算子。

底层信号的定义。我们通过组合低频振荡之和、调幅高频分量以及加性随机噪声，生成连续时间过程 $x(t)$。令

$$x_{\text{low}}(t) = \sum_{i=1}^{N_{\text{low}}} A_i \sin\bigl(2\pi f_i t + \phi_i\bigr), \quad f_i \in [0.5, 20] \text{ Hz},$$

wh其中振幅 A_i 和相位 ϕ_i 均为均匀抽取。我们构造一个归一化包络

$$e(t) = \frac{x_{\text{low}}(t) - \min x_{\text{low}}}{\max x_{\text{low}} - \min x_{\text{low}}}$$

并定义一个调制

$$m(t) = 0.3 + 0.7\, e(t).$$

高频分量随后为

$$x_{\text{high}}(t) = \sum_{j=1}^{N_{\text{high}}} m(t) \sin\bigl(2\pi f'_j t + \phi'_j\bigr), \quad f'_j \in [30, f_{\text{HR}}/2].$$

最后，添加少量随机噪声：

$$x(t) = x_{\text{low}}(t) + x_{\text{high}}(t) + \eta(t), \qquad \eta(t) \sim \mathcal{N}(0, \sigma^2).$$

高分辨率采样。 高质量传感器以 $f_{\text{HR}} = 512$赫兹的采样率对 $x(t)$ 进行采样：

$$X_{\text{precise}}[n] = x\left(\frac{n}{f_{\text{HR}}}\right), \qquad n = 0, \ldots, N-1.$$

该序列模拟了一台精确的高端测量设备。

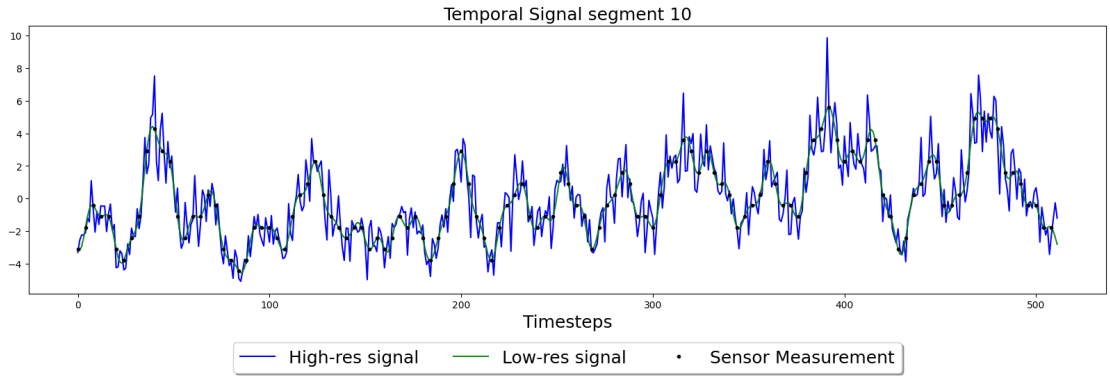
低质量传感器的构造。 遵循信号处理领域采用的标准退化模型 [44], ，我们模拟一个低质量传感器。给定以 $f_{\text{HR}} = 512$赫兹采样的高分辨率信号 $X_{\text{precise}} \in \mathbb{R}^N$，我们应用：

1. **轻度带宽限制**。 一阶低通算子 H，截止频率为 $f_c = 50$赫兹，表示低成本传感器的模拟响应。2. **采样率降低**。 滤波信号按 $S = f_{\text{HR}}/f_{\text{BR}} = 10$ 倍降采样，得到采样率为 $f_{\text{BR}} = 51.2$赫兹的观测。3. **电子噪声和量化**。 添加独立测量噪声 $n \sim \mathcal{N}(0, \sigma^2_{\text{meas}})$，其 $\sigma_{\text{meas}} = 0.01$，随后使用 b位均匀量化器进行量化（$b = 4$，动态范围根据每个片段调整）。

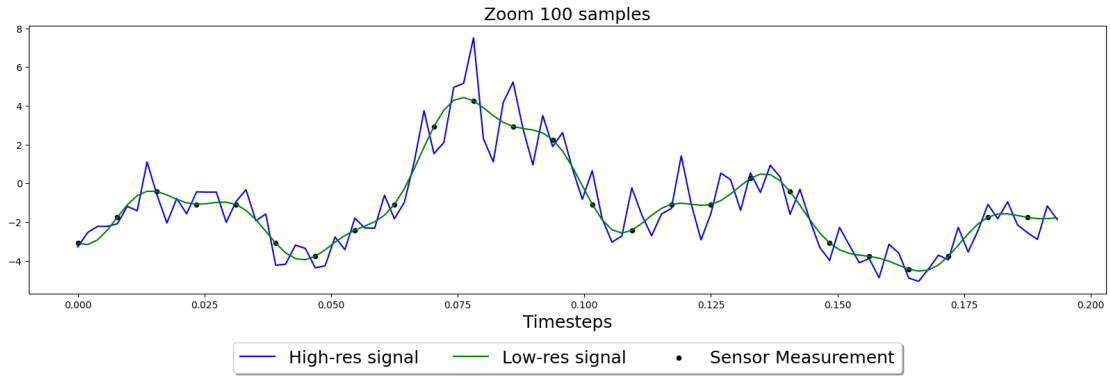
完整的前向模型可以简洁地写为

$$X_{\text{cheap}} = Q(S\, H\, X_{\text{precise}} + n)\,.$$

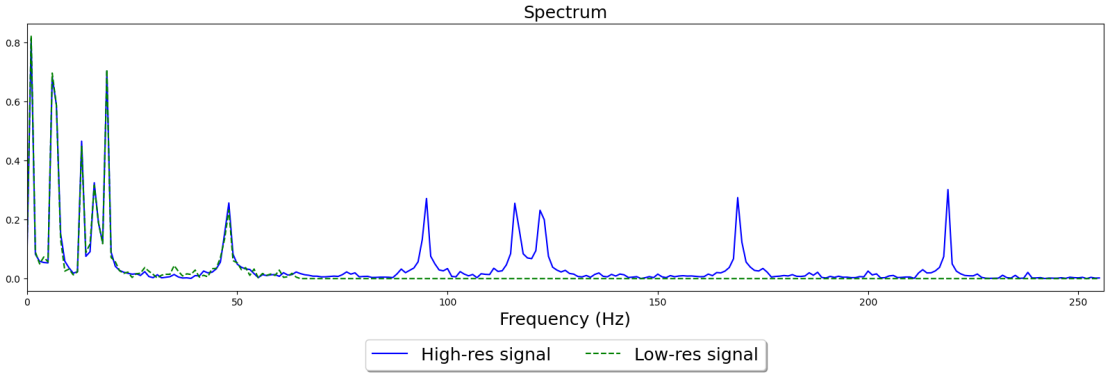
最后，使用基于 FFT 的零插入ˆ过程将稀疏测量 X_{cheap} 插值回原始网格，生成与 X_{precise}对齐的稠密退化信号 X_{cheap} 。随后，将两个信号都划分为ˆ个片段，每个片段包含 N 个时间步，满足 $N = 512$。来自 X_{cheap} 的片段对应域 $\mathcal{D}_{\text{A}}$，而来自 X_{precise} 的片段对应域 $\mathcal{D}_{\text{B}}$。为消除域之间的对齐，仅保留 $\mathcal{D}_{\text{A}}$ 的前 M 个片段，并移除 $\mathcal{D}_{\text{B}}$ 的前 M 个片段。这模拟了使用两个不同传感器采集时间信号的过程。过去使用的传感器不够可靠，我们试图利用之后由更精确的传感器采集的数据，改进这台性能较差的传感器所获得的旧数据。图18展示了一个属于第一个



(a) 片段 10 的时间信号

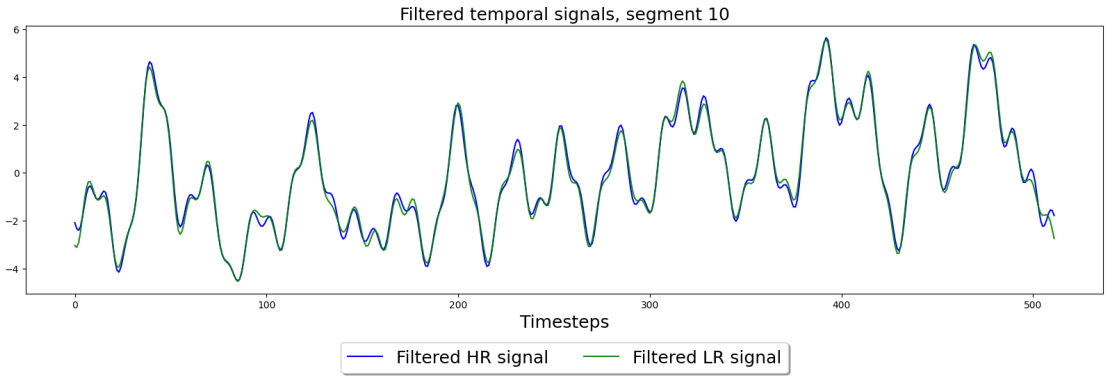


(b) 放大查看前100个时间步

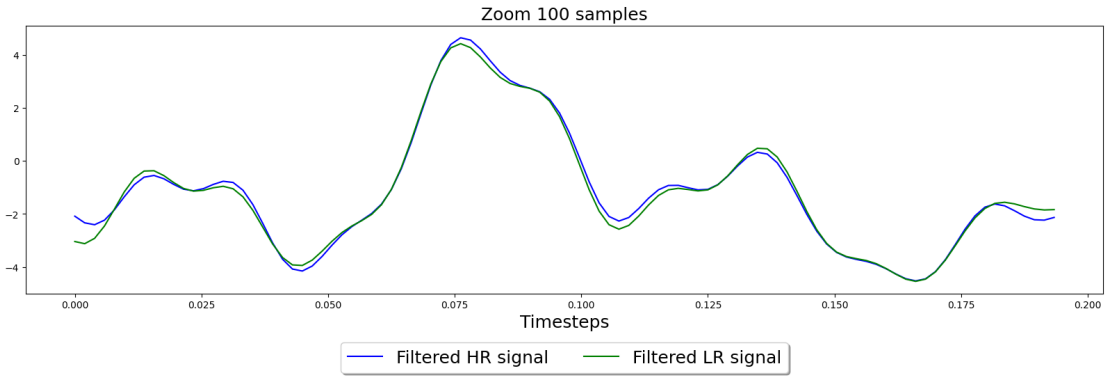


(c) 频谱

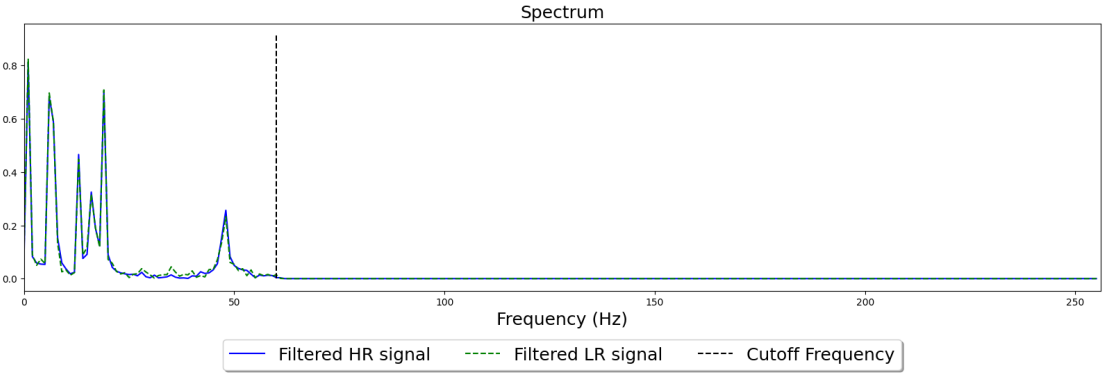
图18：由优质（高分辨率信号）和劣质（低分辨率信号）传感器采集的信号片段可视化。



（a）片段 10 的滤波后的时域信号



（b）放大查看前100个时间步



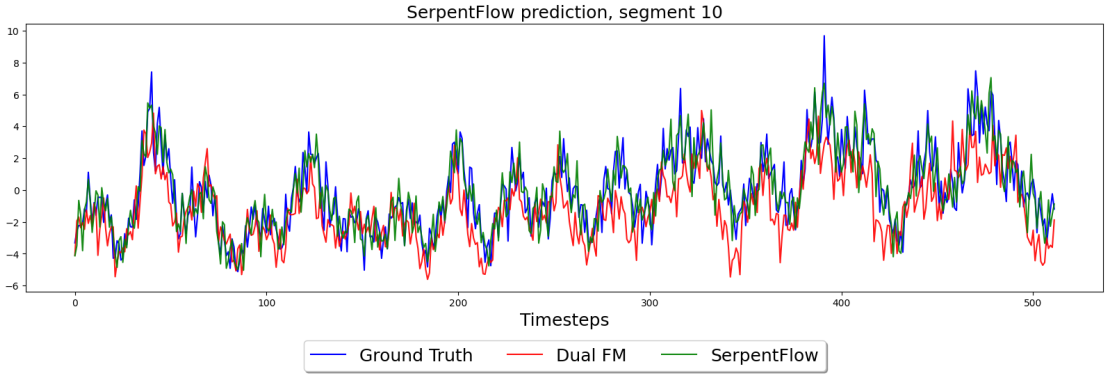
（c）频谱

图19：由质量良好的（高分辨率信号）和质量较差的（低分辨率信号）传感器采集，并经过截止频率为 $w_c = 60$ 赫兹的低通滤波器滤波的信号片段可视化。

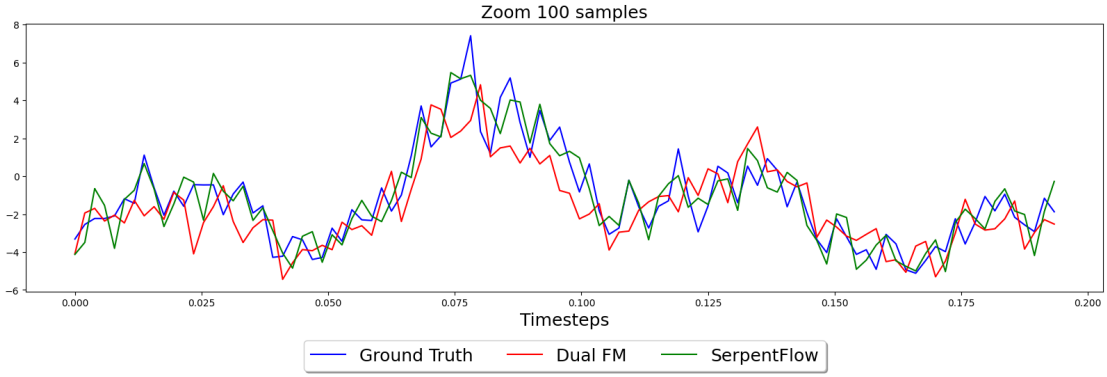
M 元素（在从 $\mathcal{D}_{\mathrm{B}}$中移除之前）。频谱表示显示， f_c以下的低频相匹配，因此频谱分解是合理的。

根据构造，我们已知截止频率为 f_c，因此无需使用分类器来确定它。我们留出较小的余量，以完整涵盖频谱上可见的最高公共频率，并应用截止频率为 $w_c = 60$ 赫兹的低通滤波器，如图19所示。可以看到，经过这一变换后，信号及其频谱几乎完全一致。

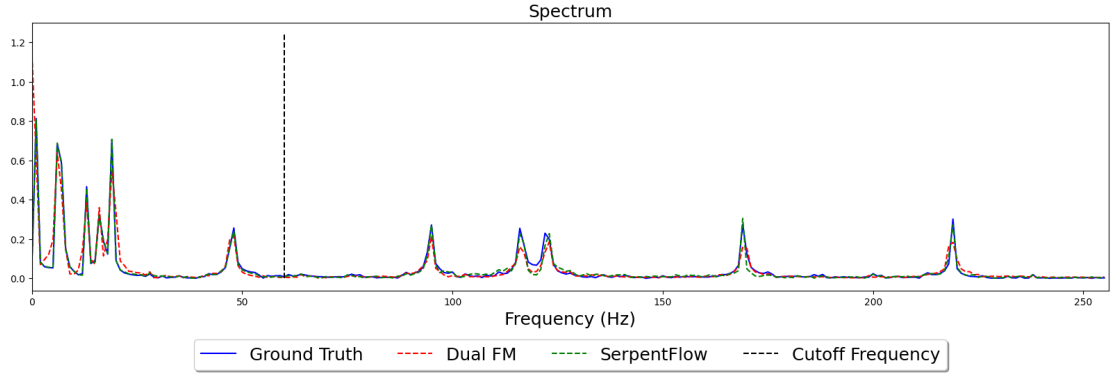
随后，我们将 SerpentFlow 和 Dual FM 应用于这一时间超分辨率问题。对于两种方法，我们都使用了配备 FiLM 条件化的一维 U-Net，并采用与其他实验相同的流匹配算法。第十个片段的重建结果见图20。可以看到，尽管 SerpentFlow 更接近原始信号，但两种方法都较好地重建了频谱。然而，在时域中，我们可以看到 Dual FM 存在明显的信号缺失，而 SerpentFlow 似乎保留了信号的整体形状（由低频决定）。重建出的高频看起来一致，但存在相移。应用相位校正后得到图21，清楚地表明 SerpentFlow 对信号的重建效果远胜 Dual FM。若未来要将 SerpentFlow 应用于时间信号，则需要考虑这种相位校正。



(a) 时间信号

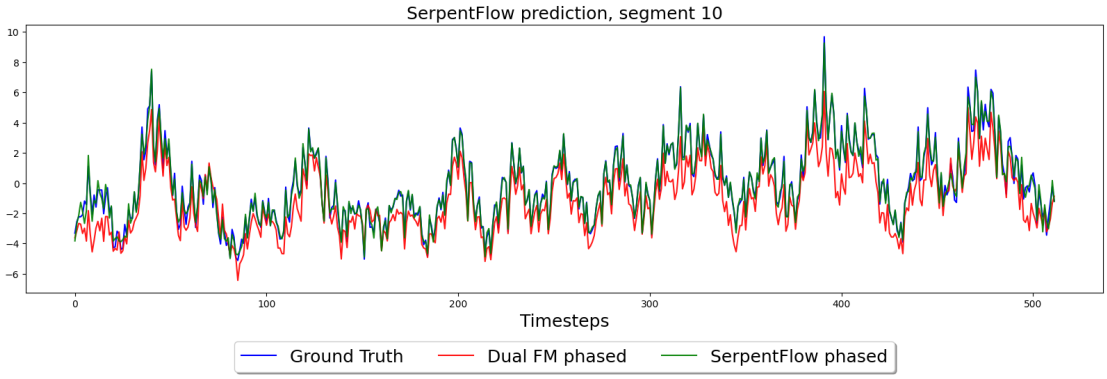


(b) 前100个时间步的放大图

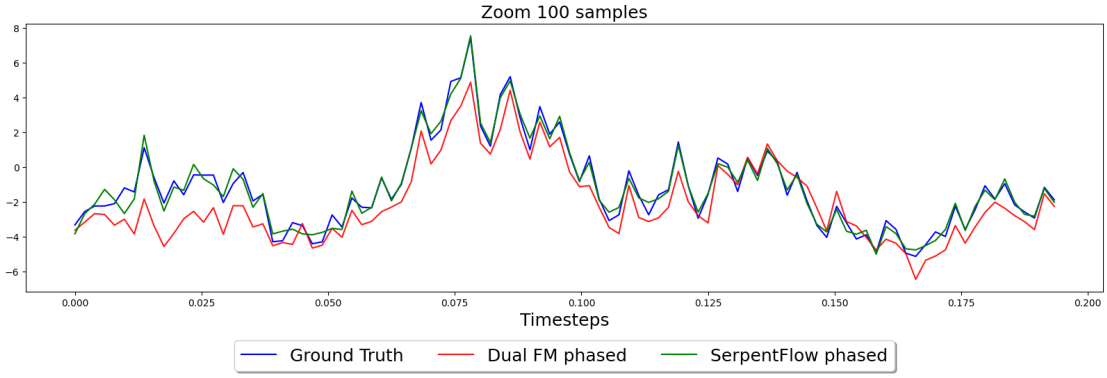


(c) 频谱

Figu图20：第十个片段以及 SerpentFlow 和 Dual F 的预测结果可视化 M



（a）时间信号



（b）放大查看前100个时间步

图21：第十个片段以及经过分阶段移位校正后 SerpentFlow 和 Dual FM 预测结果的可视化。