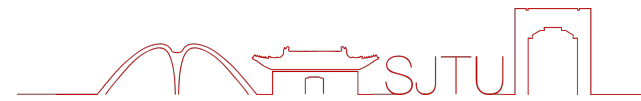




上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



人工智能研究院
Artificial Intelligence Institute



物理世界的视觉直觉学习

邓画予
上海交通大学

—— 饮水思源 · 爱国荣校 ——



邓画予，现为上海交通大学计算机系二年级博士生，导师为**杨小康**老师和**王韞博**老师，本科毕业于上海交通大学IEEE试点班。

主要研究方向为三维场景理解，直觉物理，AI for science科学计算等。部分成果发表于ICML/ICLR/CVPR。

目 录

1. 什么是视觉直觉学习？
2. 流体力学视觉推断
3. 流体力学视觉概率式
知识迁移



我们能用规则（公式求解、数值模拟、硬编码） 完整描述这个物理世界吗？

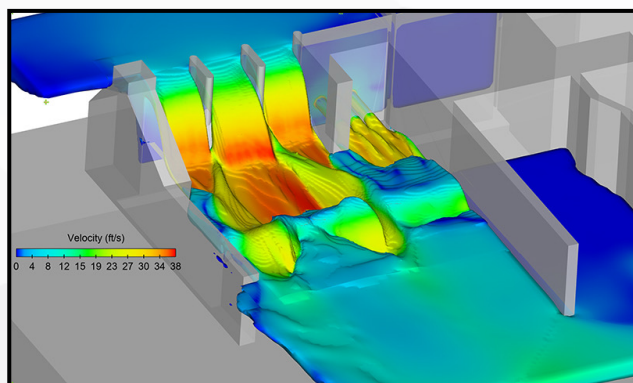
假设空间

物理规律的数学表达

$F(x)$



模型空间
(计算模拟)



观测空间



我们能用规则（公式求解、数值模拟、硬编码） 完整描述这个物理世界吗？

假设空间

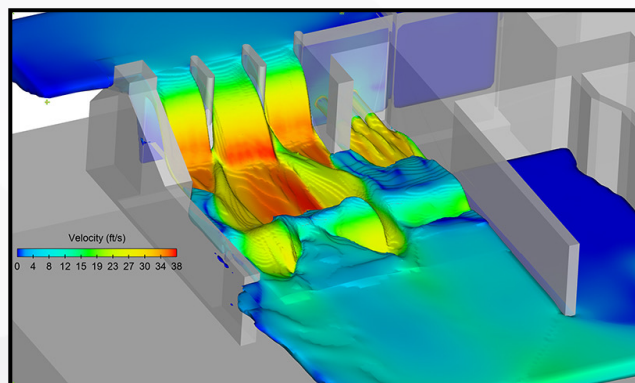
物理规律
的数学
表达

$$F(x)$$

$$F(x) + \epsilon$$

修正已有规律

物理空间
(计算模拟)

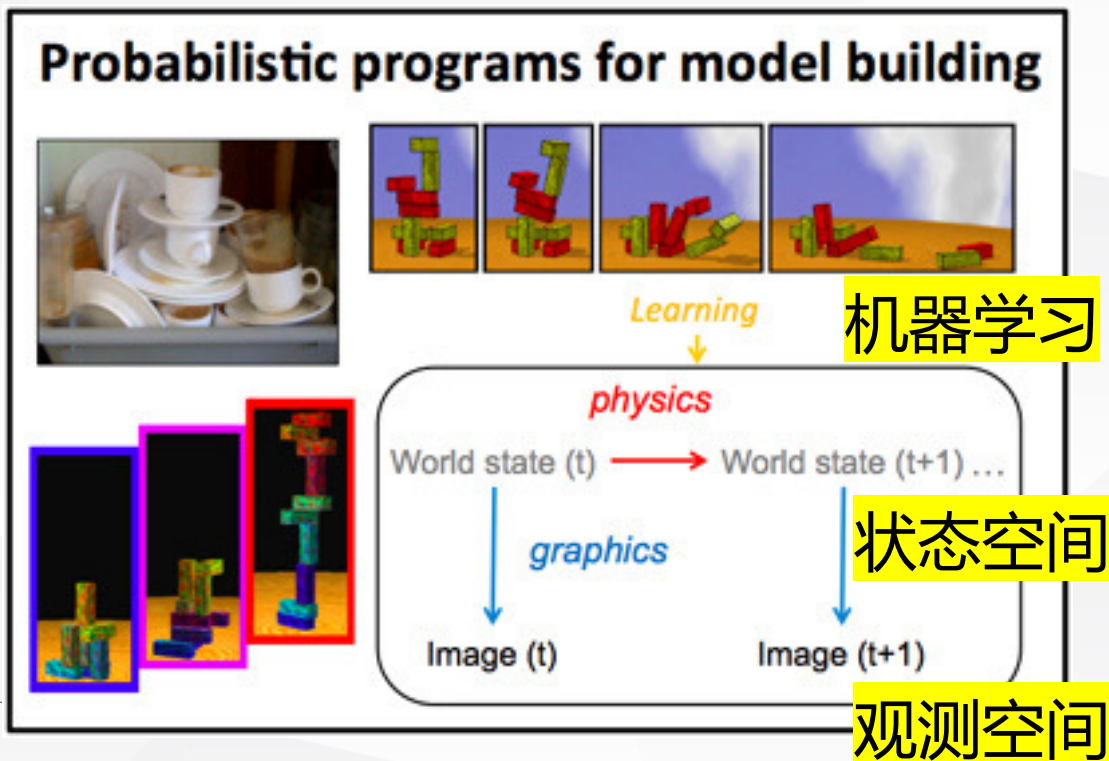


观测空间



- 通过**生成、对比、优化**，训练神经网络，建模现实世界中**难以用方程**刻画的物理规律、物理性质

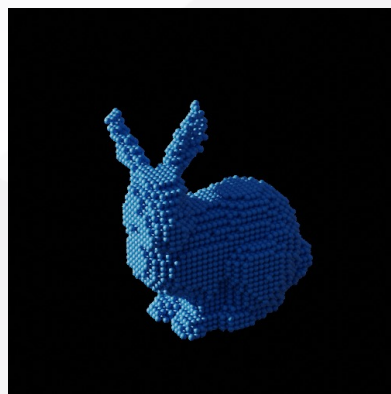
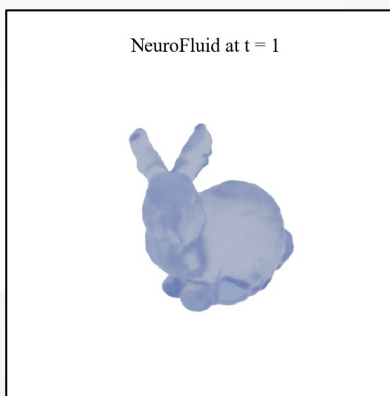
$$s_{t+1} = f_{\theta}(s_t, a_t, z_t); \quad s_t = g_{\phi}(o_t)$$



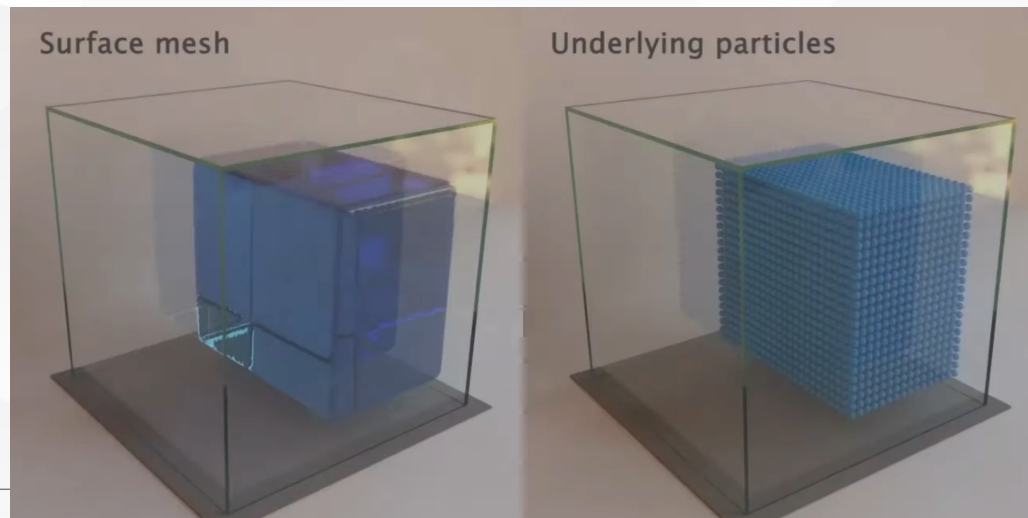
- 利用神经网络的可微分性质，通过基于梯度的误差反向传递，实现**观测—状态、状态—状态**的迭代优化

视觉的直觉物理动力学推断

- **显式物理状态空间建模**，从图像出发，形成动力学动态的直觉推断（以**符号模型**或**神经网络**的形式表达）
- 主要神经网络技术：神经辐射场（由动力学状态空间转移到图像观测空间）、可学习的可微分动力学网络（建模动力学状态空间迭代）等

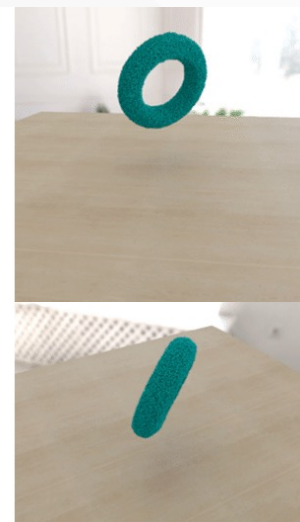
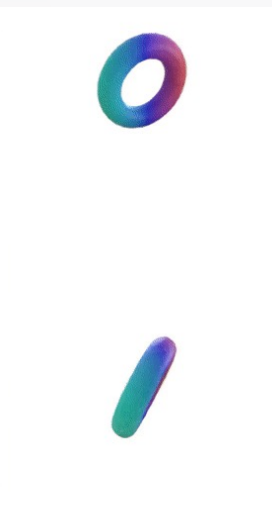
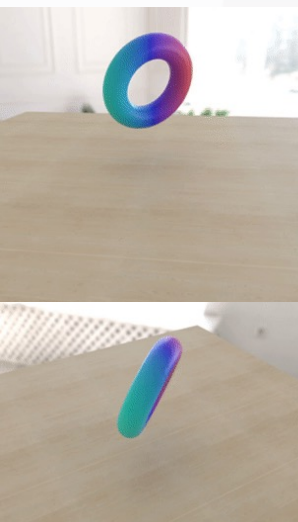
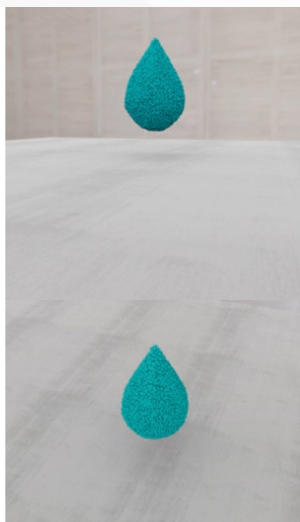
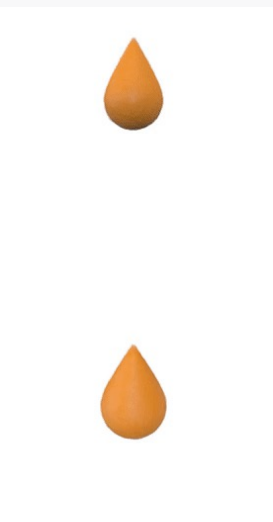
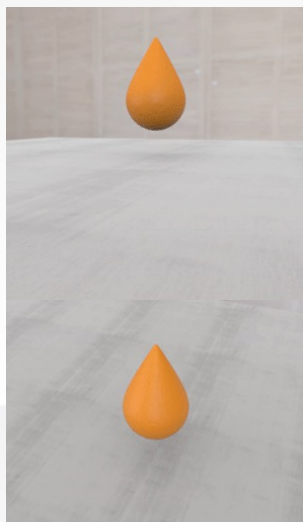


模拟流体等动力学系统



视觉观测的物理属性推断

- **显式物理状态空间建模**，从图像出发，结合**可微分仿真器（MPM）和渲染模型**反演动力学系统中的物理属性（例如Navier-Stokes方程中的牛顿流体粘滞系数）

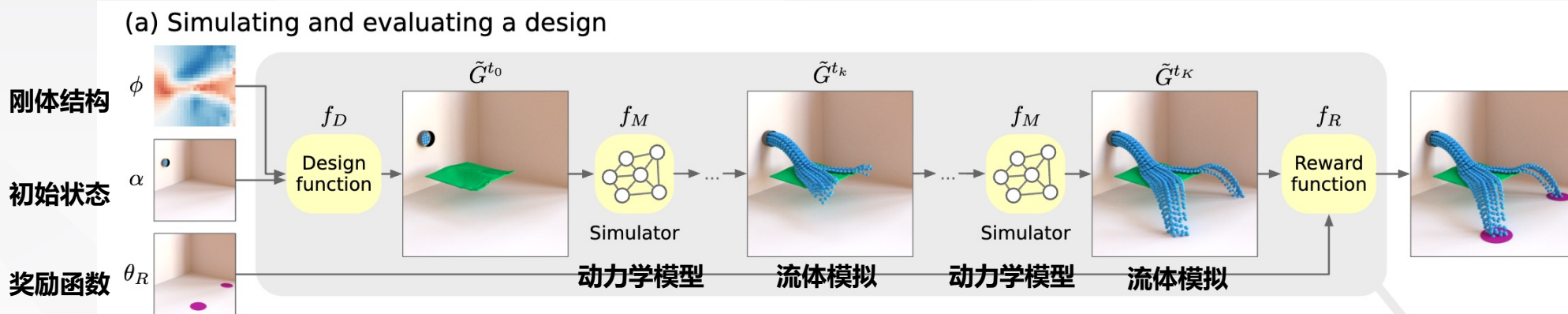


牛顿流体的粘滞系数及对应的仿真结果

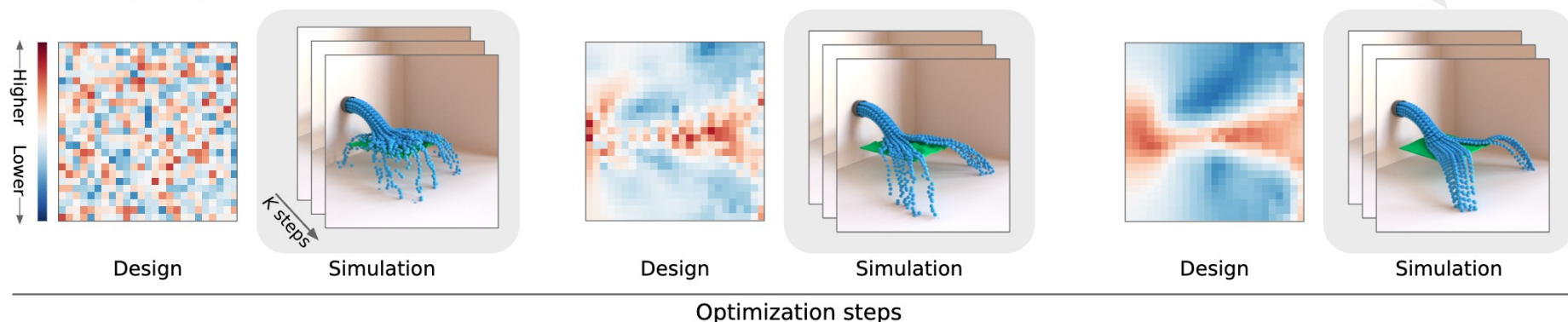
弹性体的刚度和泊松系数及对应的仿真结果

智能体决策的模拟器用于结构设计

- 引入**动作空间**，支持智能体在世界模型内部进行策略推演和试错
- 主要神经网络技术：变分自编码器、状态空间模型、深度强化学习等



(b) Example optimization



基于神经网络流
体模拟，优化固
体结构设计

视觉观测的物理属性与动态推理问答

- 将视觉感知模块与物理的动力学模型和大语言模型相结合形成视觉物理推断问答引擎，进行场景理解。

Introduction: An array of pulleys, including both **movable** and **fixed** types, along with anchor points, is arranged on a wall. Ropes are configured with their ends connected to **pulleys**, **loads**, or **anchor points**, and can be wound around the pulleys. These loads possess varying **masses**, interacting with other forces in the system, leading to the emergence of distinct motion patterns. | The primary objective of the model is to identify the **tension distributions** within this elementary rope system. Additionally, it is tasked with recognizing **potential correlations or constraints among objects** in motion, such as the coordinated movement of loads and the rotation of pulleys on a single rope. Moreover, the model is expected to infer **relationships** between the loads' masses.

场景描述片段



视频观测序列

Physical Property Questions

物理属性问答

Mass I

Is the **mass** of the **purple cube** greater than twice that of the **brown cube**?

- a) Yes
- b) No
- c) Cannot Judge

Mass II

Is the **mass** of the **sphere** less than that of the **brown cube**?

- a) Yes
- b) No
- c) Cannot Judge

Tension

Is the **tension** in the **black rope** approximately equal to half that in the **red rope** (i.e. the short rope linking purple cube)?

- a) Yes
- b) No
- c) Cannot Judge

Dynamics Questions

动力学推断问答

Counterfactual

If the **purple cube** were far much heavier, which direction would the **blue movable pulley** rotate?

- a) Anti-clockwise
- b) Clockwise
- c) Not affected

Goal-Driven

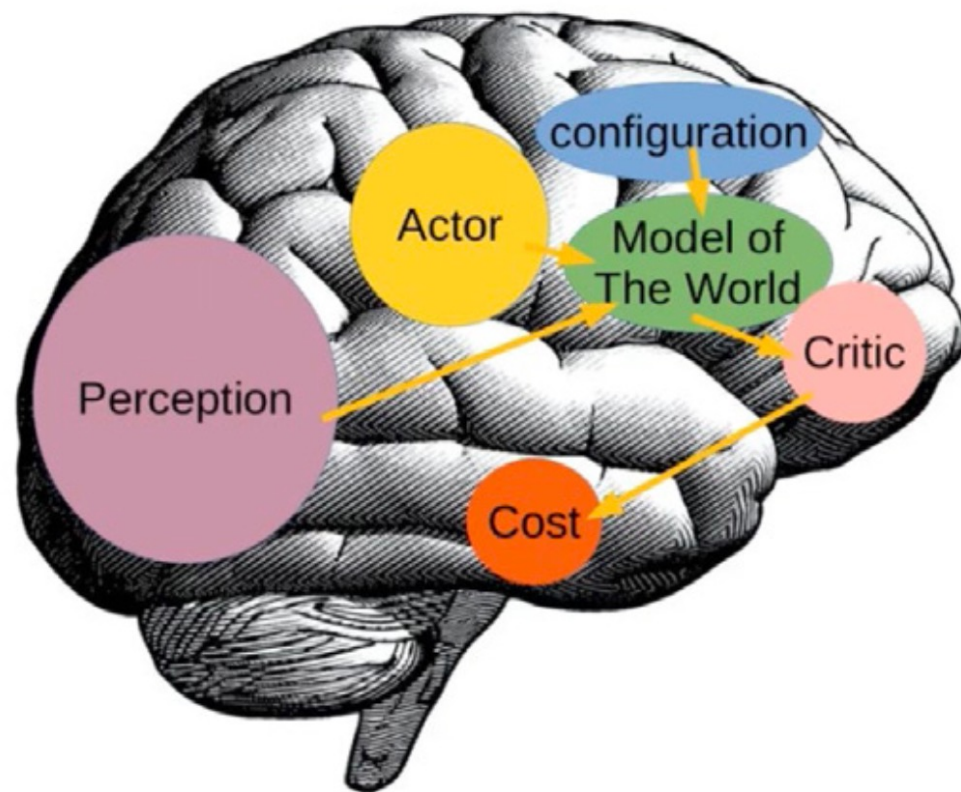
What can we do to rotate the **blue fixed pulley** clockwise?

- a) Decrease gray cube mass
- b) Decrease **brown cube** mass
- c) Increase sphere mass
- d) None of the above works

Goal-Driven

What can we do to lift the **red cube** upwards?

- a) Decrease **red cube** mass
- b) Increase **purple cube** mass
- c) Increase sphere mass
- d) None of the above works



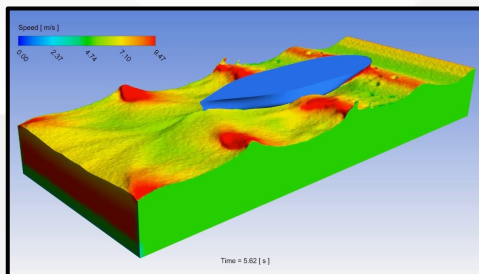
关键科学问题是，如何通过视觉生成及环境交互，构建物理世界的神经网络模拟引擎？

从视觉观测推断物理空间的**状态转移机理**

支撑智能体系统高效**理解、决策、设计**

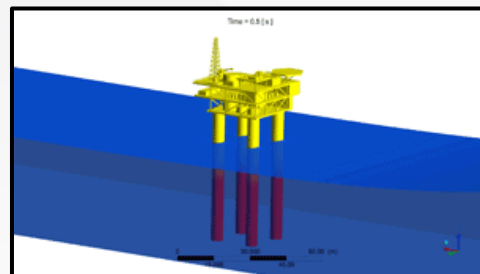
From *Deep learning, reinforcement learning, and world models* by Y. LeCun et al.

■ 计算流体力学重大需求



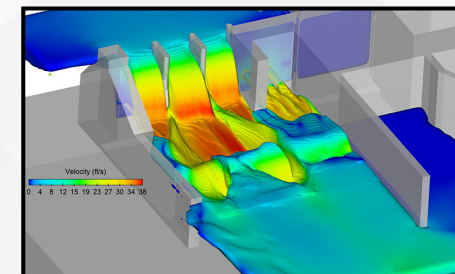
船舶

水上高速载具设计



海上平台

受风浪侵袭分析



水利

消能及分流设施仿真

■ 传统计算流体力学建模：对流体控制方程的数值求解

NAVIER-STOKES EQUATION

$$\rho g_x - \frac{\partial p}{\partial x} + \mu \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) = \rho \frac{Du}{Dt}$$

$$\rho g_y - \frac{\partial p}{\partial y} + \mu \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \right) = \rho \frac{Dv}{Dt}$$

$$\rho g_z - \frac{\partial p}{\partial z} + \mu \left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} \right) = \rho \frac{Dw}{Dt}$$

Navier-Stokes Equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{V}) = 0$$

Continuity



Claude-Louis Navier
1785-1836



Sir George Stokes
1819-1903

数值模拟的局限

- 物理建模难题 (假设/边界)
- 数值计算难题

机器学习方法能超越传统流体模型的物理建模能力吗？

刻画流速、密度、压强、黏度、温度等变量间的特定关系，由对应形式的微分方程组成，给予特定边界条件才可求解

目 录

1. 什么是视觉直觉学习？
2. **流体力学视觉直觉学习**
3. 流体力学视觉概率式知识迁移

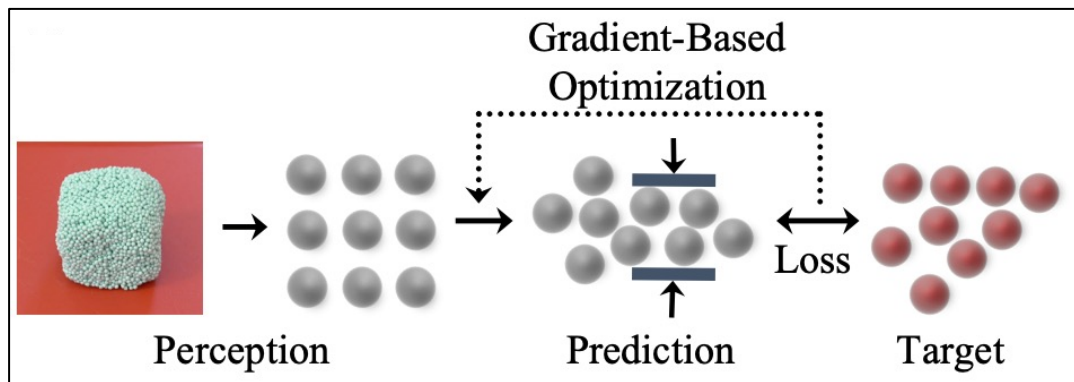


NeuroFluid: Fluid Dynamics Grounding with Particle-Driven Neural Radiance Fields

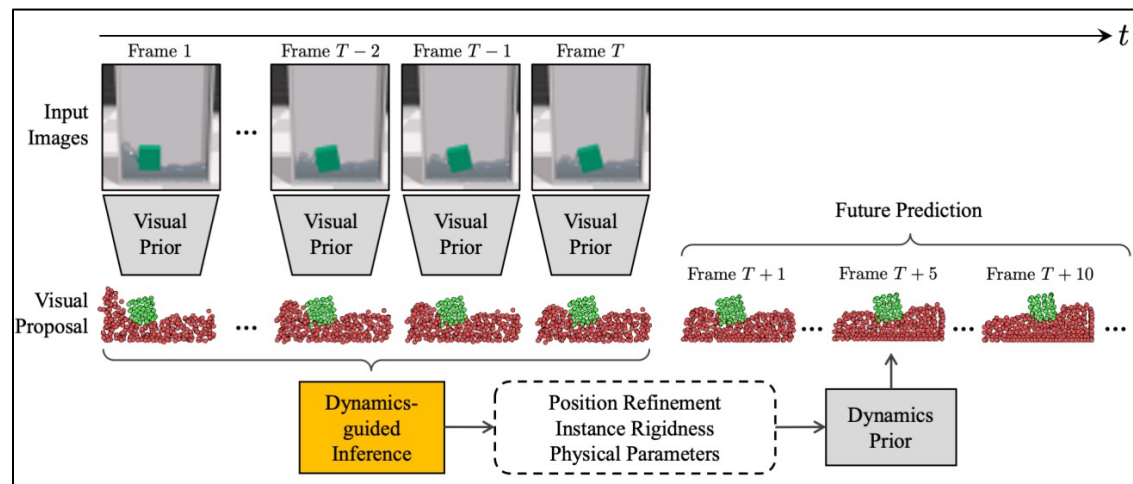
Shanyan Guan, Huayu Deng, Yunbo Wang[†], Xiaokang Yang

ICML 2022 Spotlight

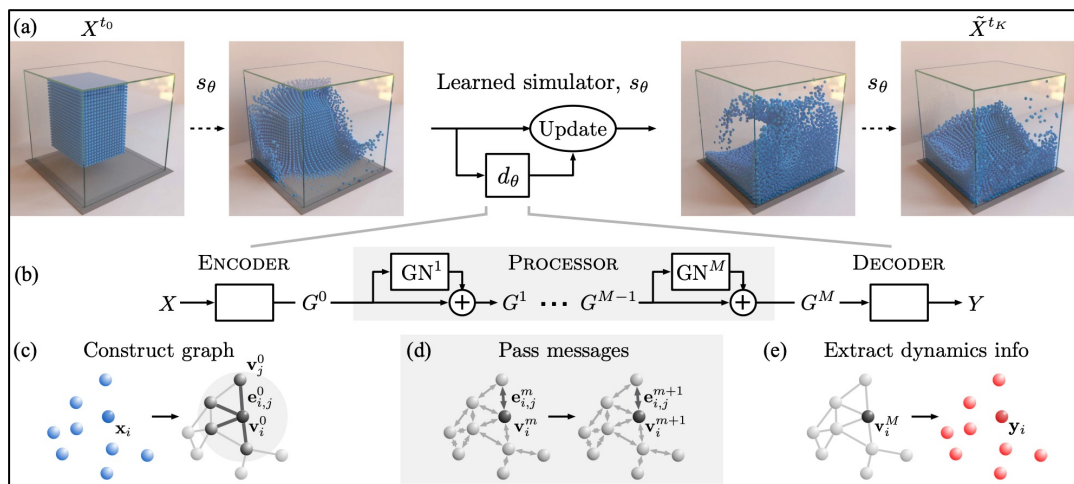
基于拉格朗日流体粒子的动力学模型



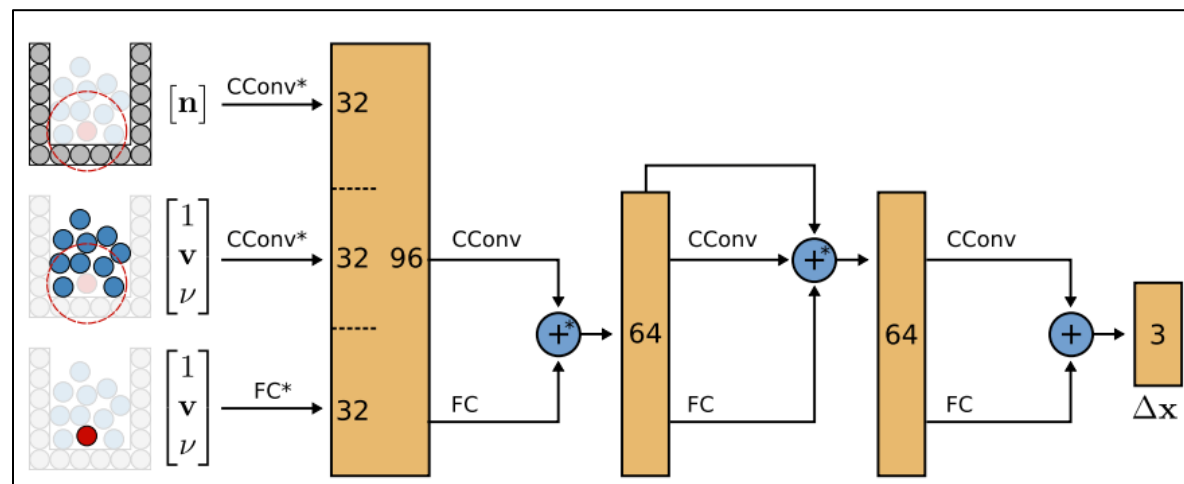
DPI-Net, Li, et al. [ICRL 2019]



VGPL, Li, et al. [ICML 2020]



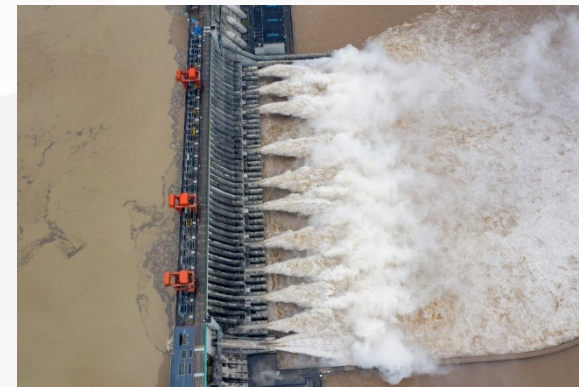
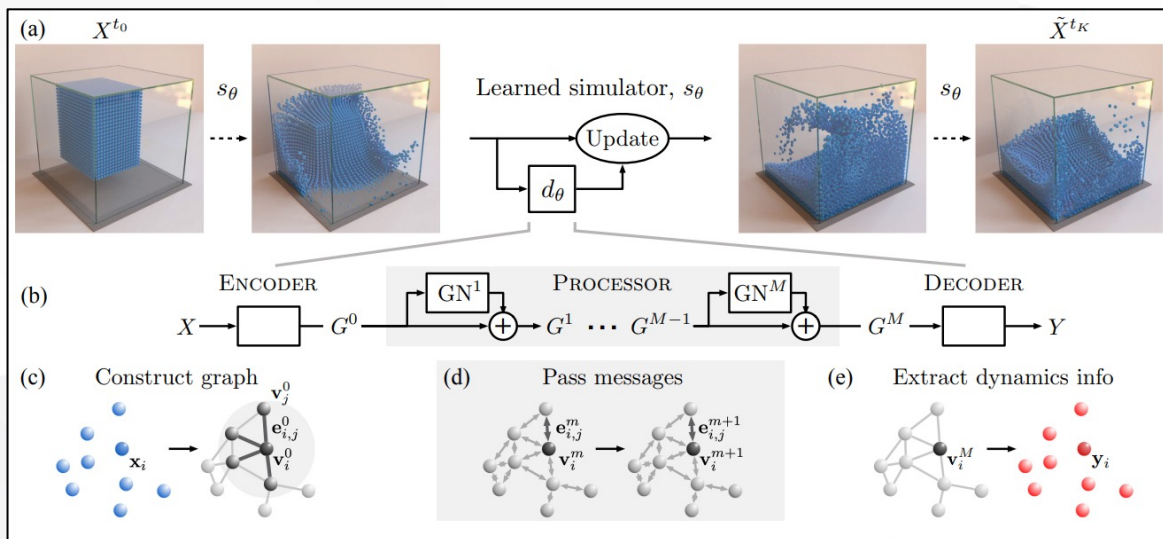
GNS, Sanchez-Gonzalez, et al. [ICML 2020]



DLF, Ummenhofer, et al. [ICRL 2020]

现有方法的缺陷：需要良好的动力学先验

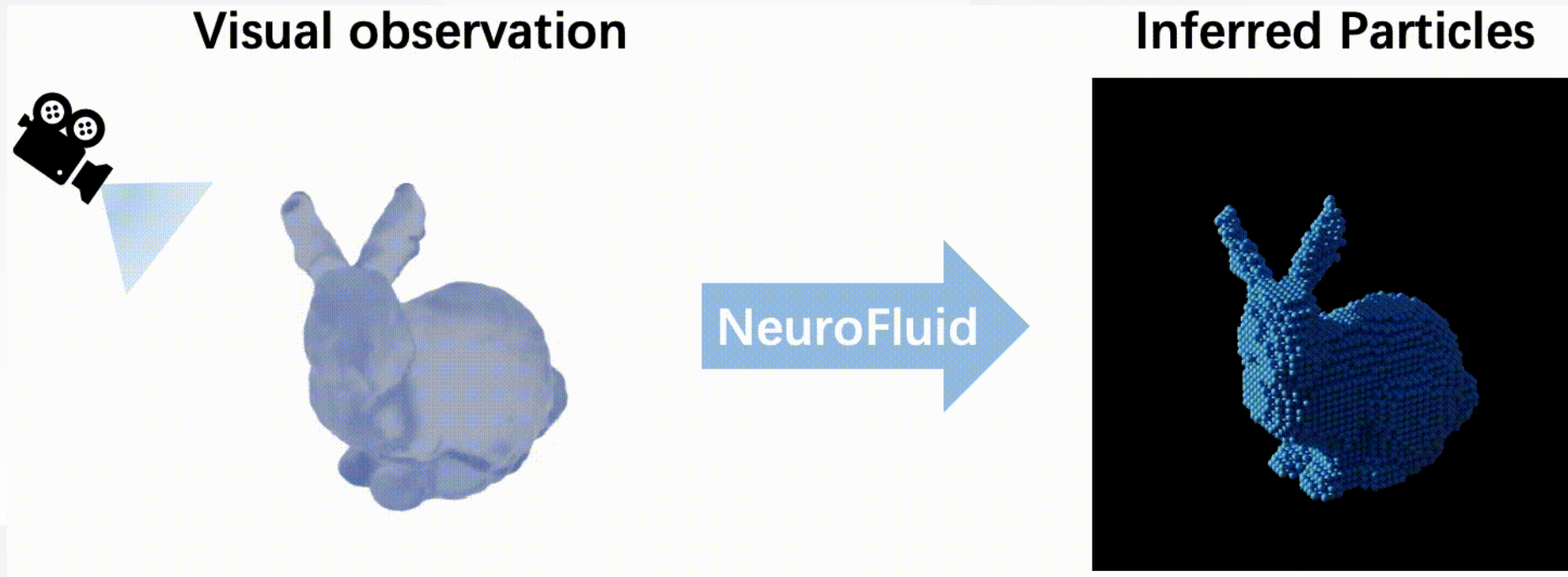
- **非机器学习的计算物理方法**，如SPH（光滑粒子动力学）等，通常采用数值方法求解Navier-Stokes方程
 - 要求物理规律可被形式已知的方程刻画
 - 场景描述困难，求解困难，泛化性较差
- **机器学习方法**，如GNS (DeepMind, 2020)，以流体粒子的后续状态为监督，训练深度预测网络
 - 要求严格采集的粒子数据，即粒子标号前后时刻需对应，**难以拓展到真实场景**
 - 仿真数据本身基于SPH等传统方法，**学到的物理动态规律无法超越传统流体仿真器上限**
 - 确定性的仿真预测网络，预测单一仿真序列，**无法涵盖自然界的随机影响的效应**



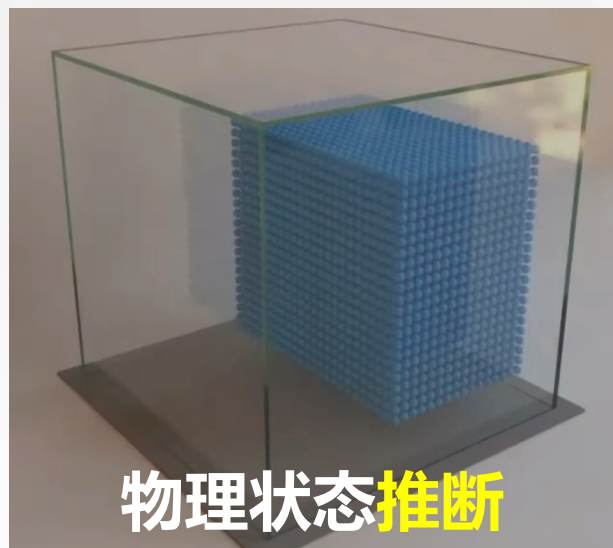
How about REAL scenes?

NeuroFluid:

仅使用视觉空间观测作为优化监督来推断流体力学。（没有粒子状态空间的监督）



核心思想：将物理模拟建模为神经渲染的反问题

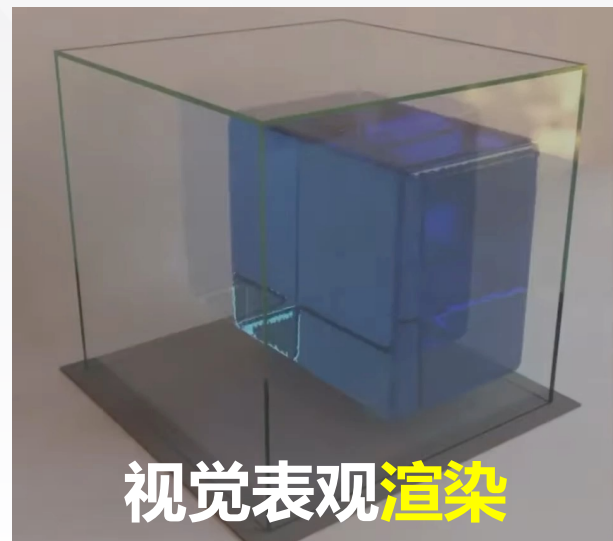


神经渲染

现有神经渲染技术不支持流体

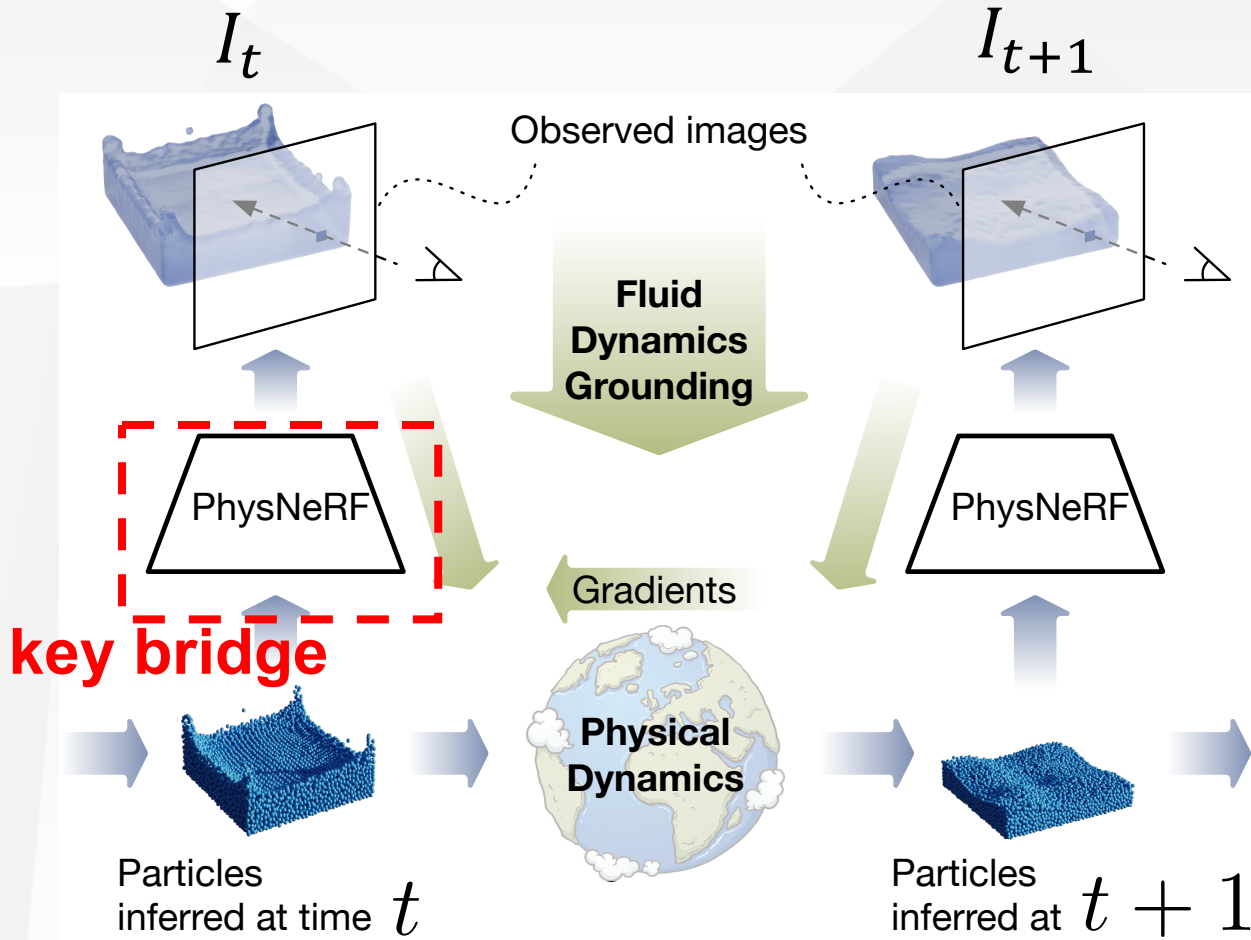
NeuroFluid

物理推断



基本假设：

- 采用流体的拉格朗日描述，通过学习一组有限粒子的运动状态，刻画流体内部的运动规律。拉格朗日描述具有易于建模动力学特性，且易于描述流体的几何外观的优势（仿真：动力学→图像）
- 流体的内部运动规律，可由三维视觉连续图像序列推测（推理：图像→动力学规律）



两个网络模块

- 粒子状态转移模型 T_θ
- 物理粒子驱动的可微渲染器 R_ϕ .
- 拉格朗日空间的粒子状态转移：
$$s_{t+1} \leftarrow T_\theta(s_t)$$
- 粒子状态空间到图像的映射： $\hat{I}_{t+1} \leftarrow R_\phi(s_{t+1}, d)$

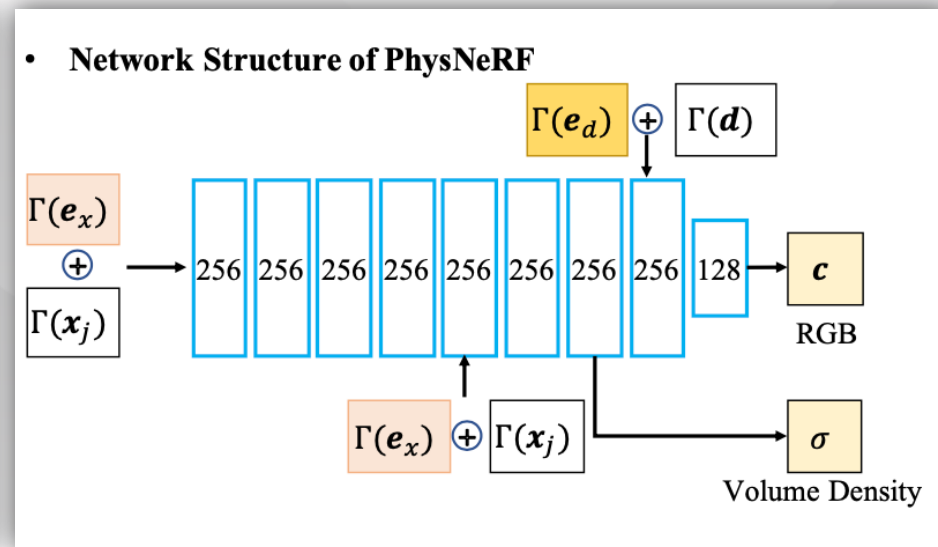
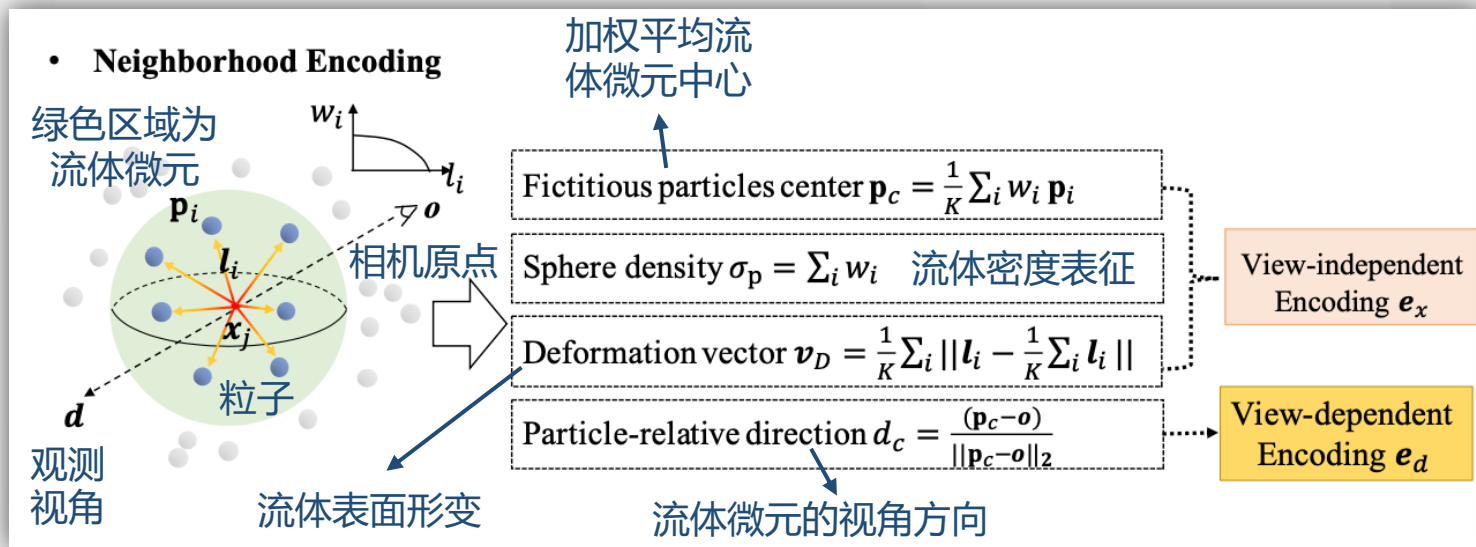
关联光线粒子与流体粒子，从真实观测反推动力学模型

模块1：粒子驱动的可微渲染器 PhysNeRF

神经渲染 ← **结合** → 物理流体粒子规律

神经渲染网络考虑：

1. **与观测视角相关的流体表征**：如视角方向、流体微元表面切线方向
2. **与观测视角无关的流体表征**：如流体微元表面形变、流体微元位置、流体微元中的粒子密度



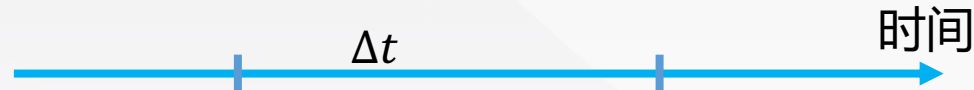


模块2：粒子状态转移网络 DLF (Deep Lagrangian Fluids : Ummenhofer et al. ICLR 2020)

- 对三维空间中，球形局部邻域内的流体粒子做卷积，忽略宏观几何信息，可泛化到复杂场景 → 

粒子状态：(1)粒子位置 x ；(2)粒子速度 v

- DLF：① 与容器边界粒子 P 的相互作用
② 与相邻粒子 X_N 的相互作用
③ 粒子 x_t^* 本身的特征



受重力作用的转移结果
(x_t^*, v_t^*)

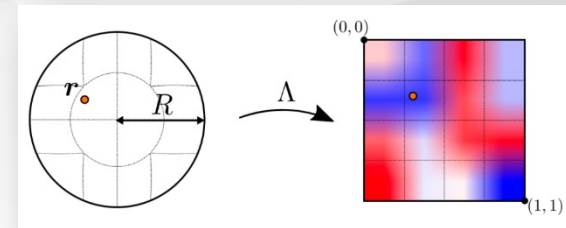
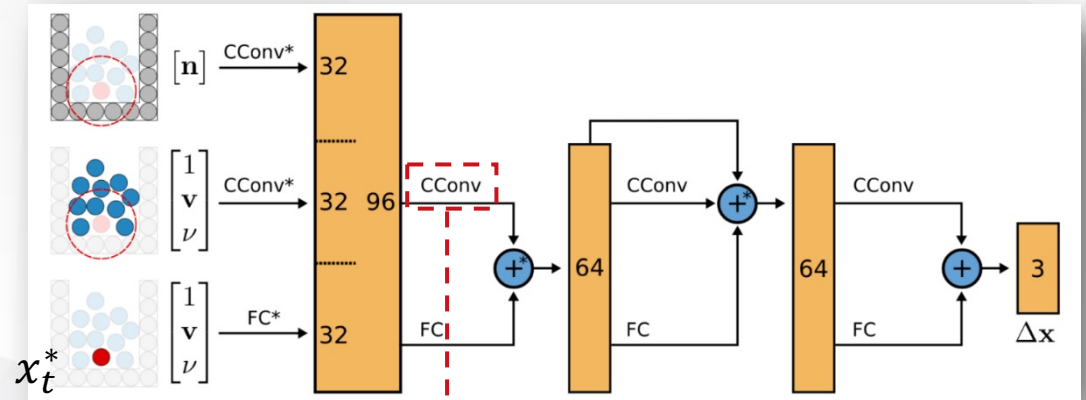
$$\begin{aligned} v_t^* &= v_t + g\Delta t \\ x_t^* &= x_t + v_t\Delta t \end{aligned}$$

$$\begin{aligned} x_{t+1} &= x_t^* + \Delta x \\ v_{t+1} &= \frac{x_{t+1} - x_t}{\Delta t} \end{aligned}$$

修正 x_t^* 并
计算 v_{t+1}

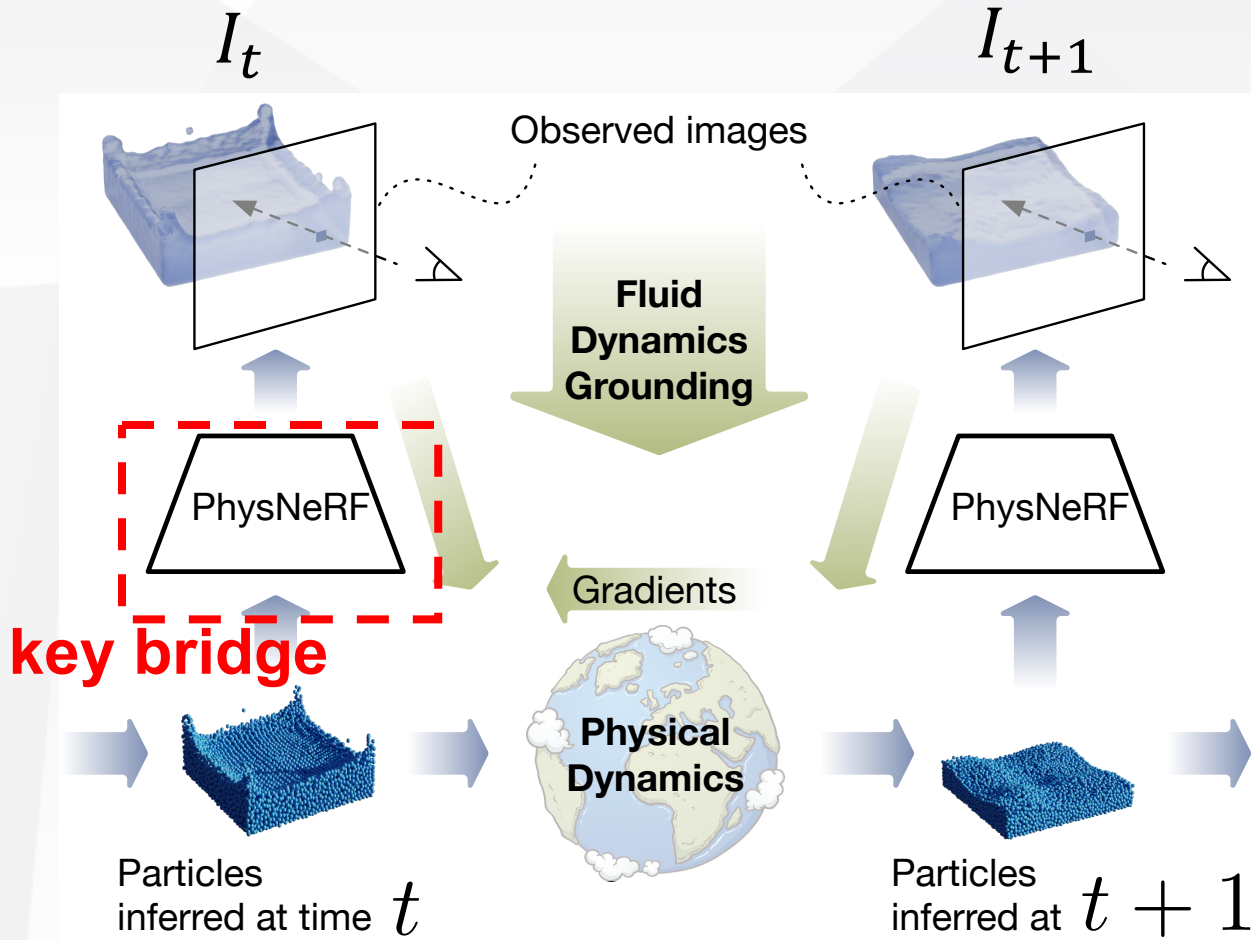
$$\Delta x = \text{DLF}(P, X_N, [x_t^*, v_t^*])$$

考虑与容器边界粒子 P 以及相邻粒子 X_N 的影响



球面空间点
↓
三维规则网格点

CConv通过映射，得到球形空间任意一点的卷积核参数



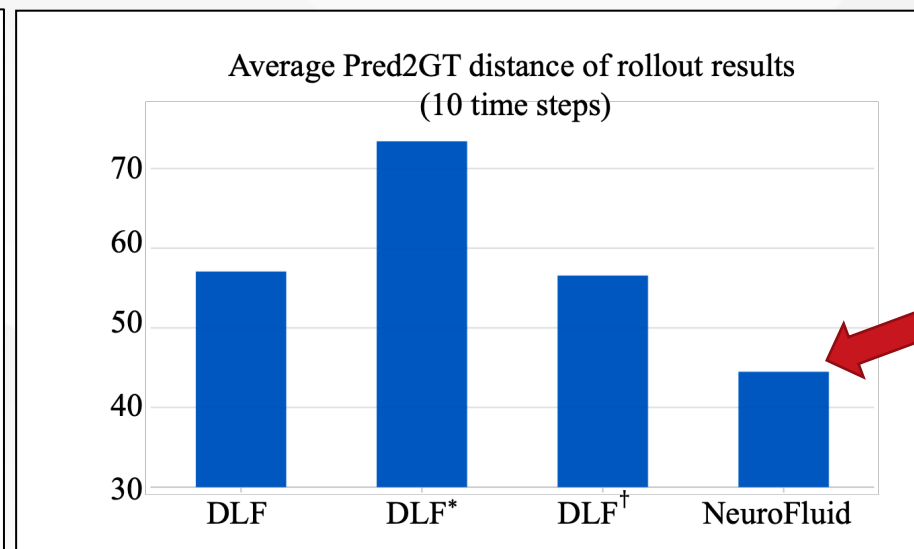
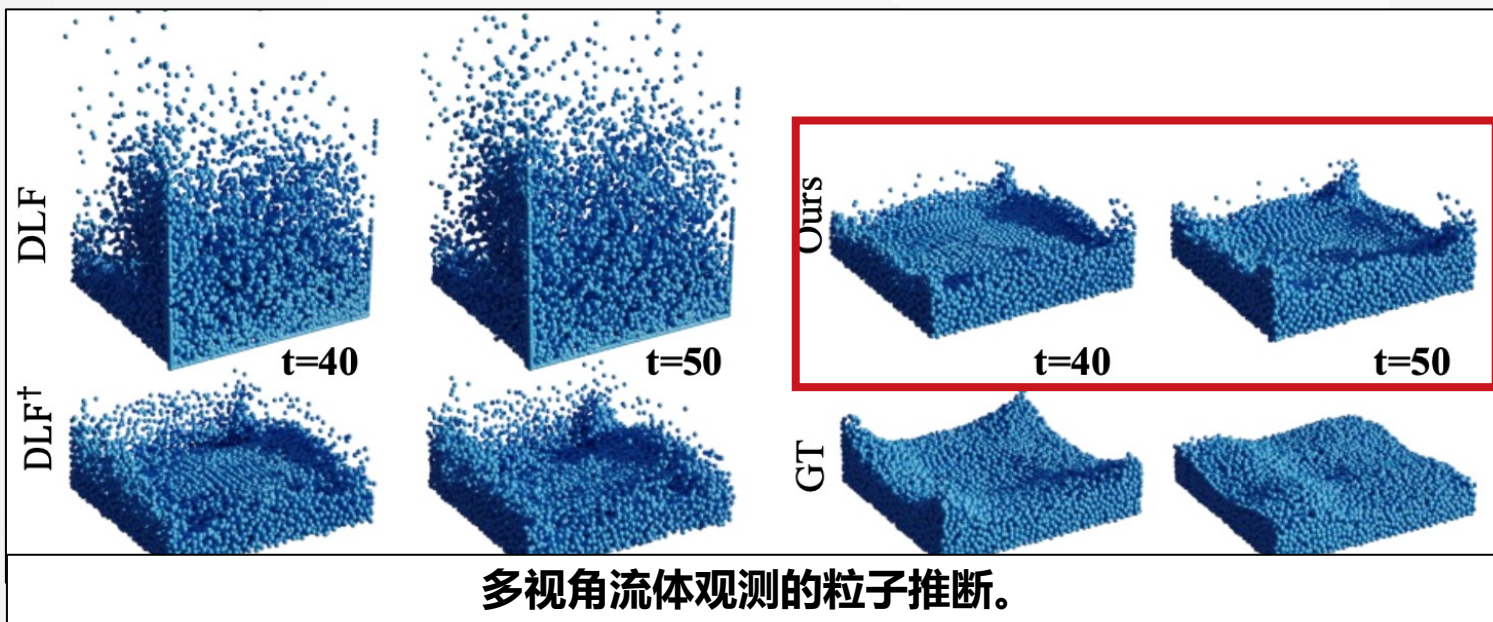
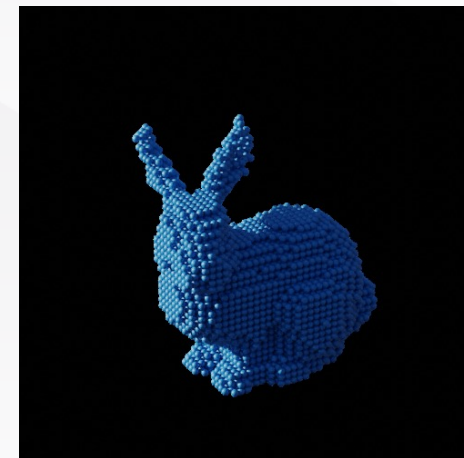
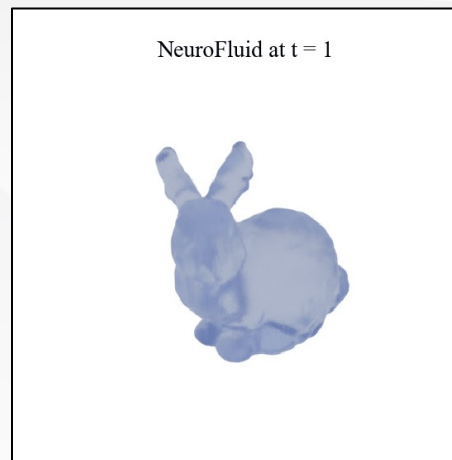
端到端的训练两个网络模块

- 粒子状态转移模型 T_θ
- 物理粒子驱动的可微渲染器 R_ϕ .
- 拉格朗日空间的粒子状态转移：
 $s_{t+1} \leftarrow T_\theta(s_t)$
- 粒子状态空间到图像的映射： $\hat{I}_{t+1} \leftarrow R_\phi(s_{t+1}, d)$

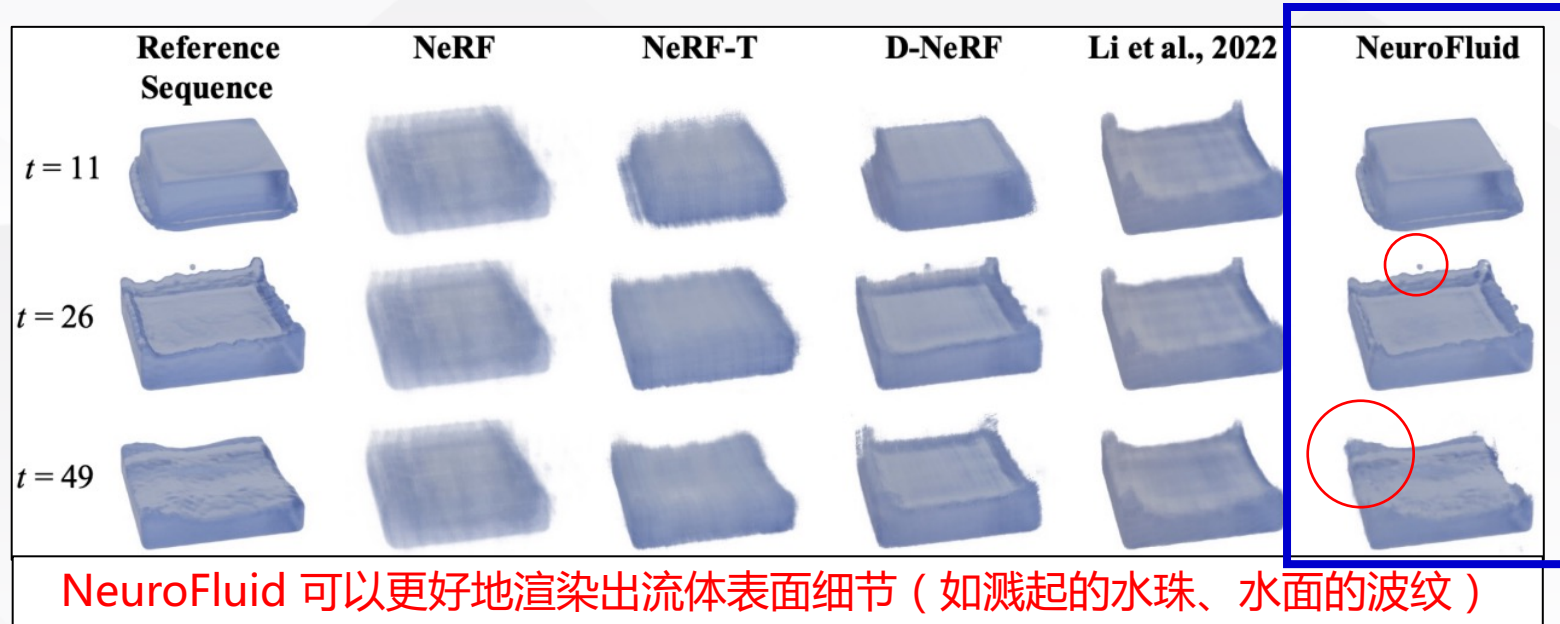
关联光线粒子与流体粒子，从真实观测反推动力学模型

实验结果：当前视觉观测下流体粒子动态反演 + 未来状态预测

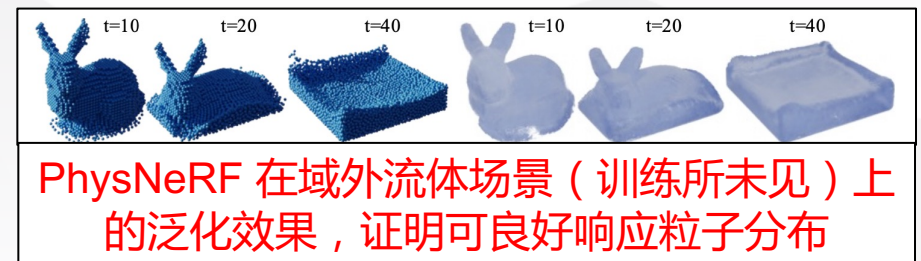
- DLF 在大规模**粒子状态**数据上进行有监督训练；
DLF[†]表示在测试场景的**粒子状态**上进行模型微调
- NeuroFluid 反演结果的粒子运动更加自然，更贴近真实情况，能更好地匹配真实流体动态



未来状态预测误差更小。DLF*表示在与测试场景物理属性相近的粒子数据上进行模型微调



Model	PSNR	SSIM	LPIPS	$d_{t < 50}^{avg}$	$d_{t \geq 50}^{avg}$
Unknown initial positions	29.21	0.94	0.12	35.69	30.26
Low-reso train, 200 × 200 px	28.67	0.94	0.16	29.90	31.67
NeuroFluid (Final)	30.06	0.95	0.10	29.80	31.49
w/o Fictitious particles center	27.42	0.92	0.18	30.79	43.66
w/o Sphere density	26.65	0.92	0.19	34.64	59.60
w/o Deformation vector	27.07	0.92	0.18	33.93	66.29
w/o Particle-relative direction	31.48	0.94	0.17	52.97	84.83



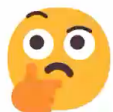
目 录

1. 什么是视觉直觉学习？
2. 流体力学视觉推断
3. **流体力学视觉概率式
知识迁移**





“物理场景理解需要的**不仅仅是定位和识别物体**，它还需要了解物体**可能具有不同的潜在特性**（例如质量或弹性），以及这些特性会**影响物理动态的结果**。” [1]



How can you tell honey from oil?

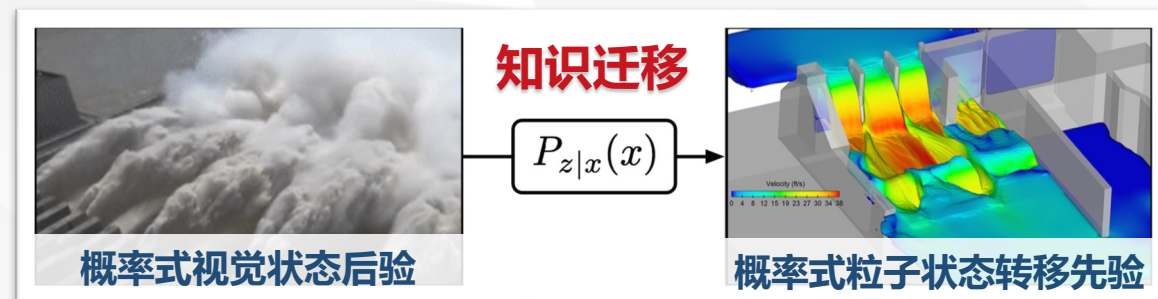


NeuroFluid或其他的动力学模型依赖于**当前观测场景的物理属性**，推断**当前观测场景或相同属性的物理动态**。

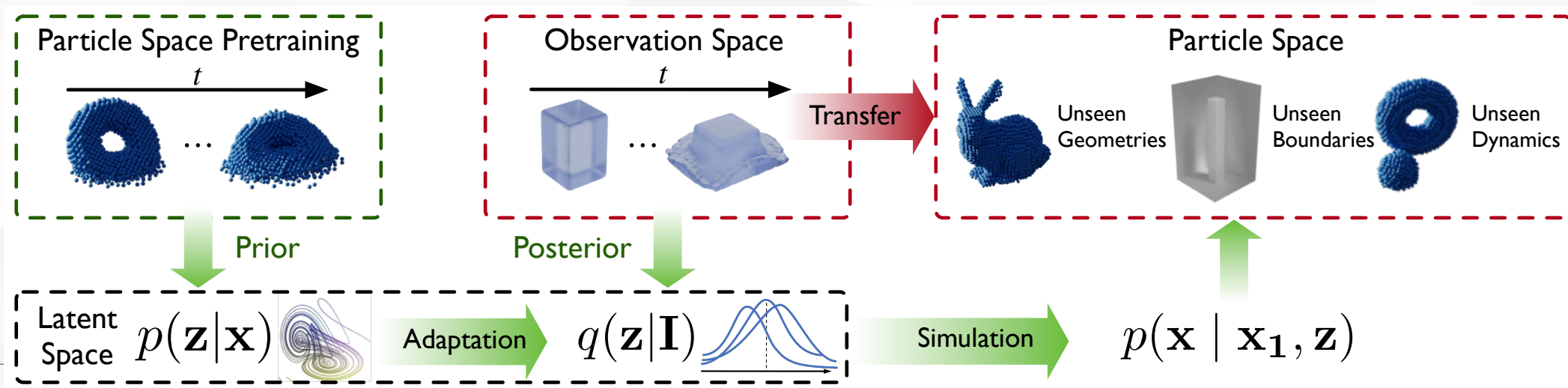
核心问题：真实世界中数据稀少，如何避免模型**过拟合**至当前视觉观测？如何有效利用仿真先验？

最终目标：仿真**视觉可观测的、属性未知的、初始状态与边界条件可变的**流体场景

思考：我们能在对物理属性了解有限的情况下预测物理系统吗？是否有可能将容易获取的视觉观测中隐藏的物理信息转移到基于神经网络的物理模拟器中？

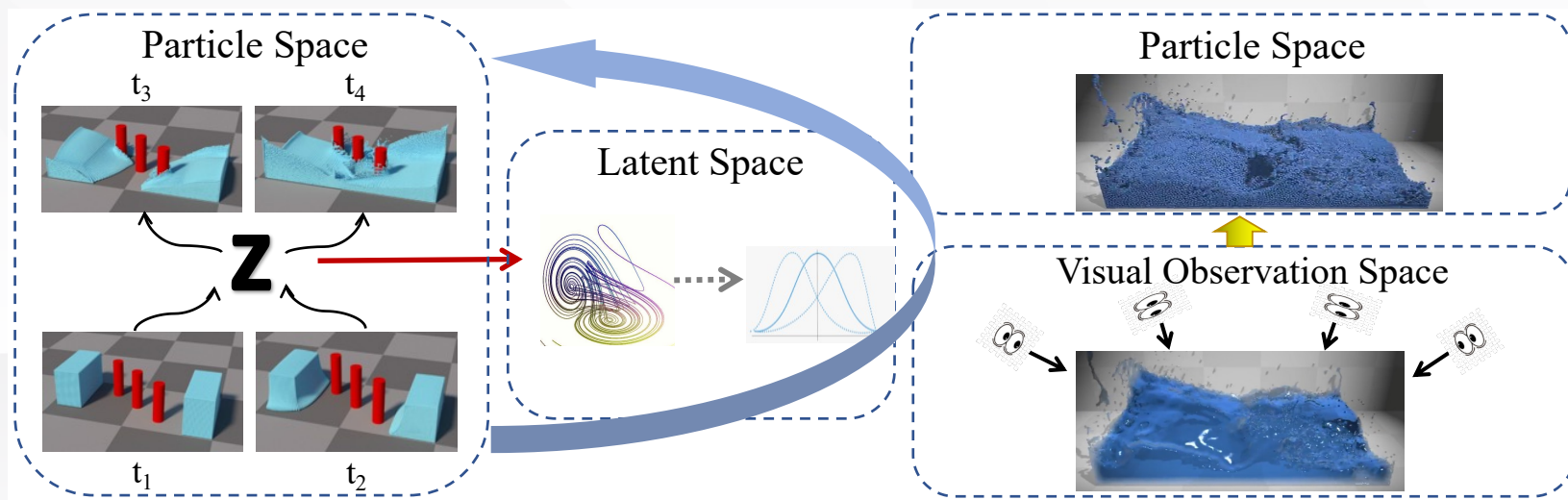


引入**概率式隐变量动力学系统**，将大规模仿真数据中的预训练知识与真实流体场景快速适配



概率式流体力学模型的优势：

- 将不同物理属性分布理解成**隐变量概率式分布**，它允许我们在不知道真实物理参数的情况下预测流体力学。
- 概率式流体力学模型依赖于从**连续输入**中推断潜在特征。
- 当提供**视觉观测**时，我们可以通过潜在隐变量空间将概率式流体力学仿真集成到视觉场景的**变分推理**方法中。



推断流体的**隐藏属性**，作为仿真域与现实域之间的桥梁

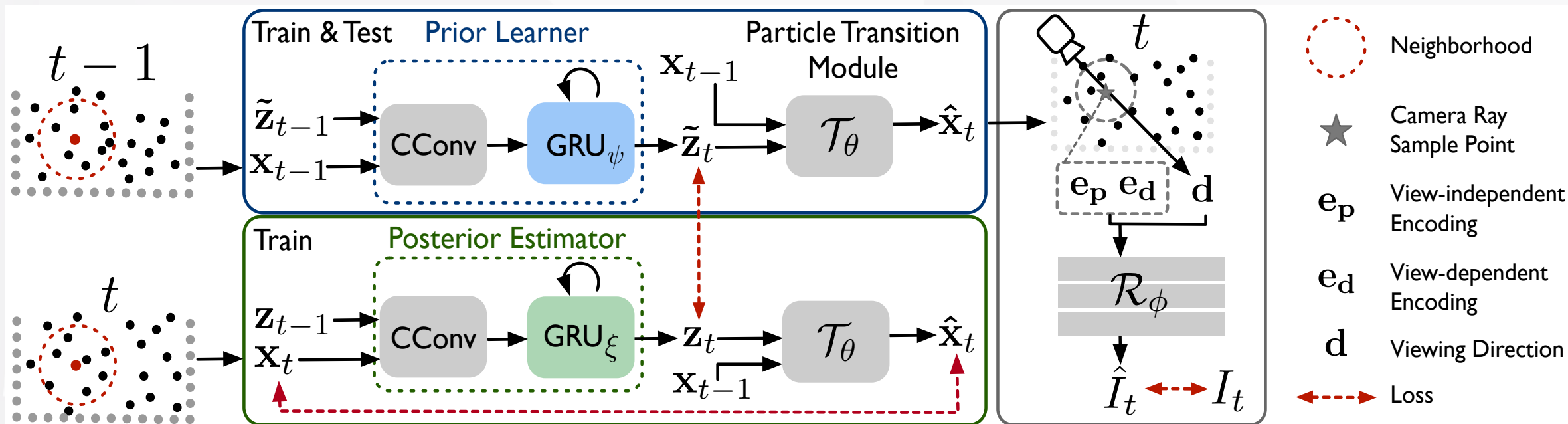
Latent Intuitive Physics: Learning to Transfer Hidden Physics from A 3D Video

Xiangming Zhu*, Huayu Deng*, Haochen Yuan*, Yunbo Wang†, Xiaokang Yang

ICLR 2024

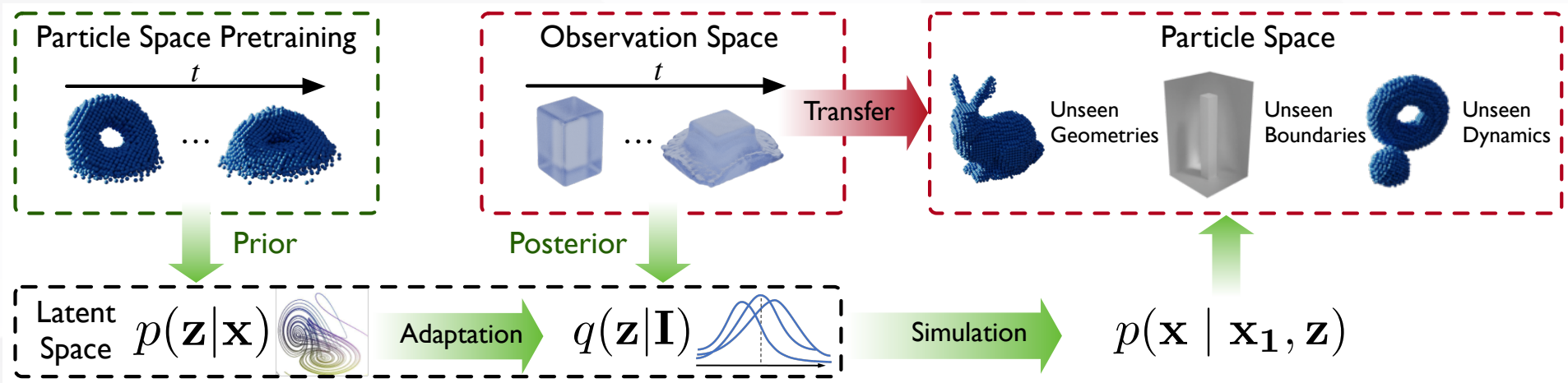
Latent Intuitive Physics模型由四个网络组件组成：

- 概率粒子转移模块
- 物理先验学习器
- 基于粒子的后验估计器
- 粒子驱动的神神经渲染器



训练流程：

- 阶段一：利用数值模拟，预训练**概率式的粒子状态转移模块**
- 阶段二：基于视觉表观，利用粒子驱动的可微渲染器，优化**视觉后验隐变量**
- 阶段三：将**视觉后验**迁移到仿真器的**物理先验网络**中。可仿真属性未知、**初始状态/边界条件可变**的流体场景



物理先验推理

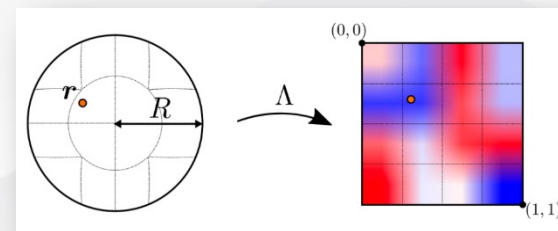
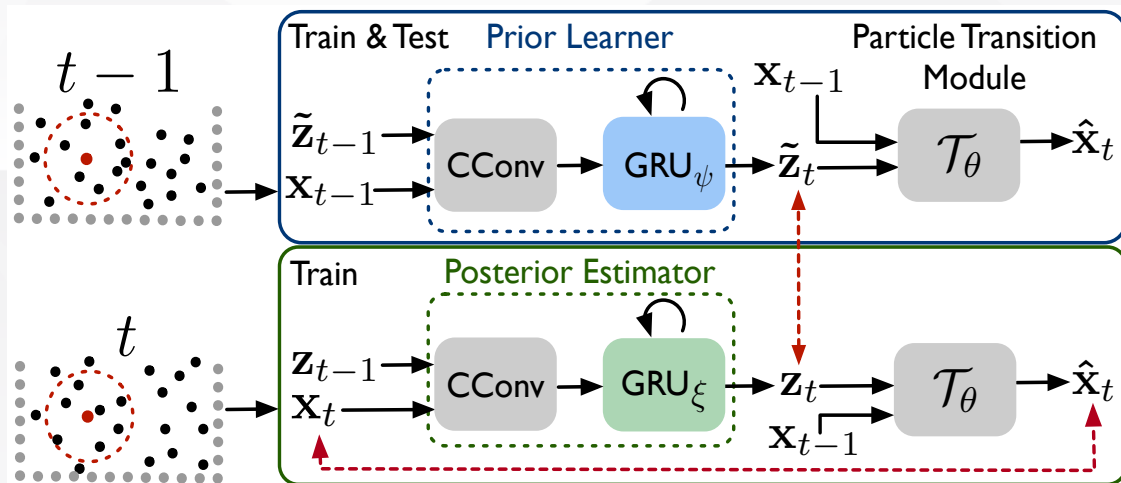
域适应

视觉后验推断

知识迁移

新场景模拟

● 阶段一：利用数值模拟，预训练概率式的粒子状态转移模块



球面空间点
↓
三维规则网格点

连续卷积 (Cconv) 作为特征编码器，以获取每个粒子的特征表示。

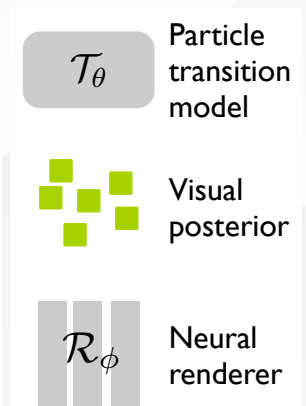
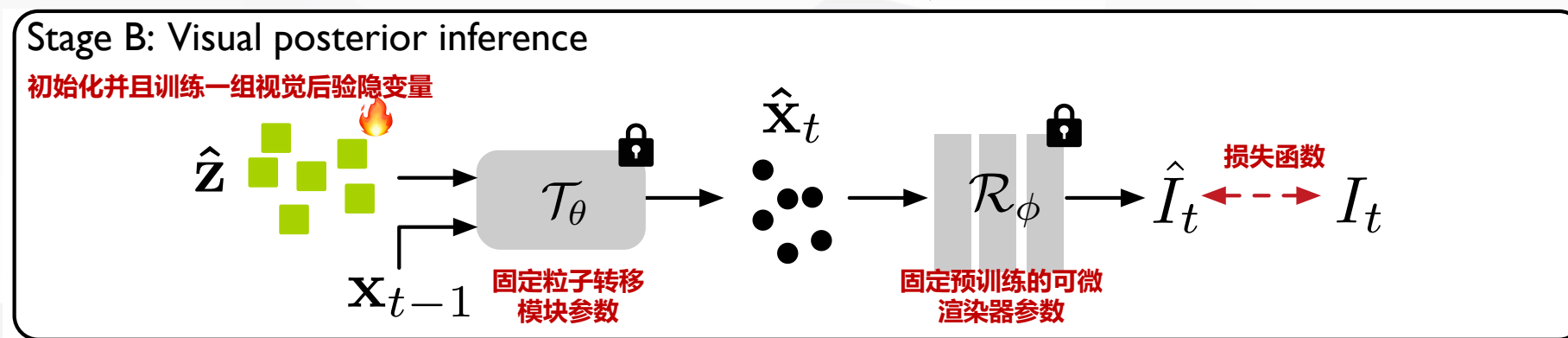
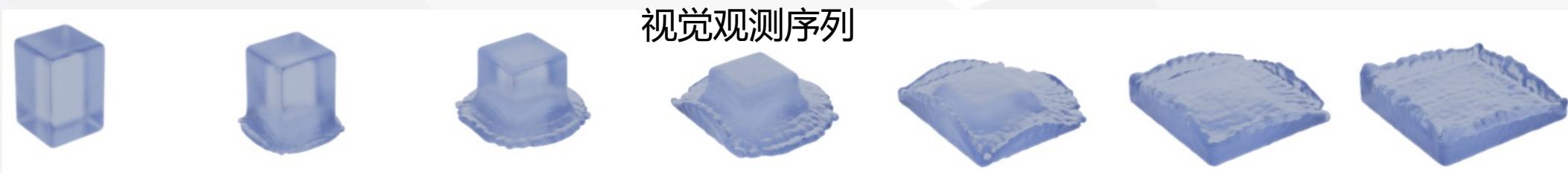
- 预训练数据集中包括不同粘滞系数和密度系数仿真的流体序列。
- 由于不可见的物理属性无法从单个状态推断出来，使用GRU来收集历史信息并推断先验潜在分布。**概率式先验学习器**与一个单独的**后验估计器**一起训练（在测试时不使用）。

$$\text{Prior: } \tilde{\mathbf{z}}_t \sim \text{GRU}_\psi(\text{CConv}(\tilde{\mathbf{z}}_{t-1}, \mathbf{x}_{t-1})); \quad \text{Posterior: } \mathbf{z}_t \sim \text{GRU}_\xi(\text{CConv}(\mathbf{z}_{t-1}, \mathbf{x}_t))$$

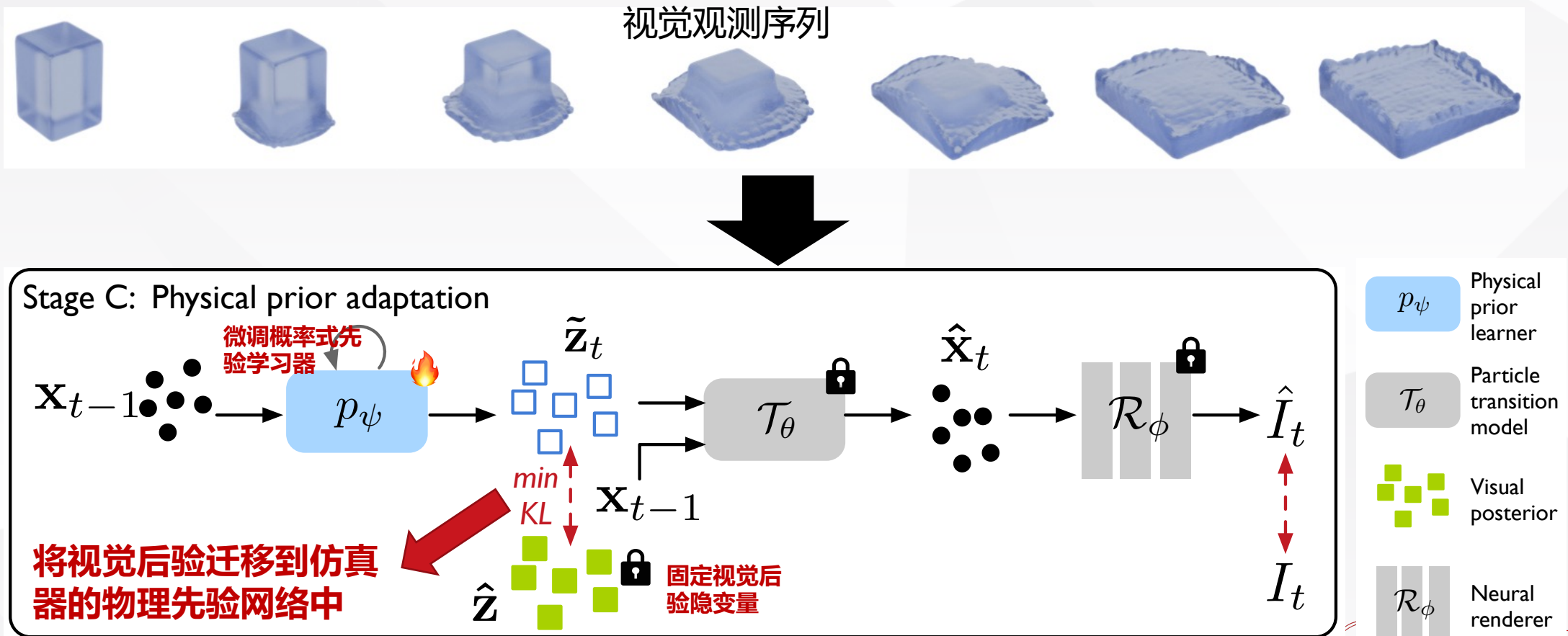
- 粒子转移模块的输入为后验粒子特征，前一个时刻的粒子状态
- 优化损失函数包括两个部分：

$$\mathcal{L}_{\theta, \psi, \xi} = \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N w_i \|\hat{p}_t^i - p_t^i\|_2^\gamma + \beta \mathcal{D}_{KL}(q_\xi(\mathbf{z}_t | \mathbf{x}_{1:t}, \mathbf{z}_{t-1}) \| p_\psi(\tilde{\mathbf{z}}_t | \mathbf{x}_{1:t-1}, \tilde{\mathbf{z}}_{t-1})) \right]$$

- 阶段二：基于视觉表观，粒子转移模块与可微分的神经渲染器结合，优化**视觉后验隐变量**

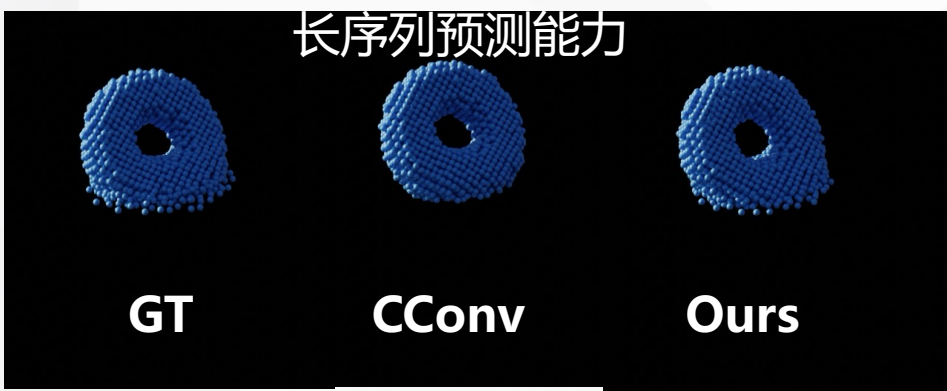


- 阶段三：将**视觉后验**迁移到仿真器的**物理先验网络**中。可仿真属性未知、初始状态/边界条件可变的流体场景

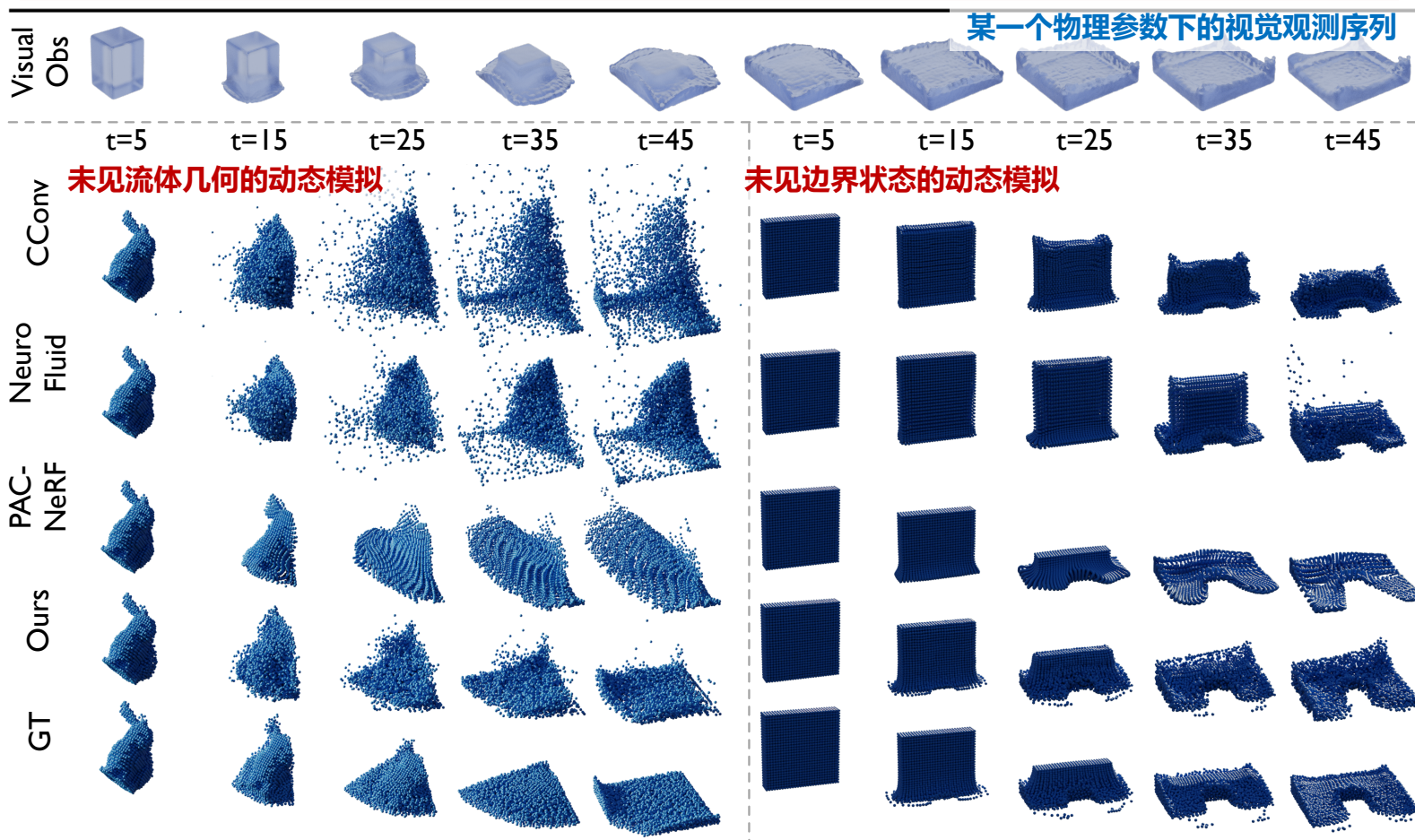


实验结果：

- 概率式粒子状态转移网络长序列预测效果强于基线方法
- 迁移了视觉后验知识的概率式粒子状态转移网络在未见流体的动态模拟上更贴近真实状况，符合人类直觉
- 更好地处理未见边界状态的流体动力学



METHODS	d_{t+1}	d_{t+2}	\bar{d}
DPI-NET (LI ET AL., 2019)	0.95	2.99	90.21
CConv (UMMENHOFER ET AL., 2020)	0.34	1.03	44.79
DMCF (PRANTL ET AL., 2022)	0.54	1.23	39.70
TIE (SHAO ET AL., 2022)	0.52	1.36	41.82
Ours	0.31±0.003	0.94±0.011	38.37±0.860



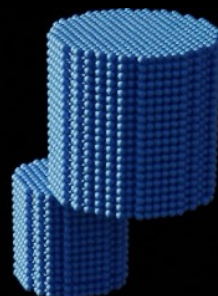


实验结果：在不同物理属性的流体交互上具有更好的动力学模拟能力

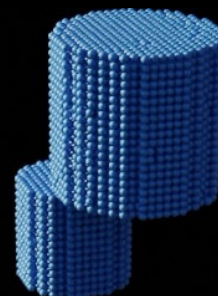
不同属性流体的动态交互能力



Source Scene



Global Latent



Ours



Novel Scene

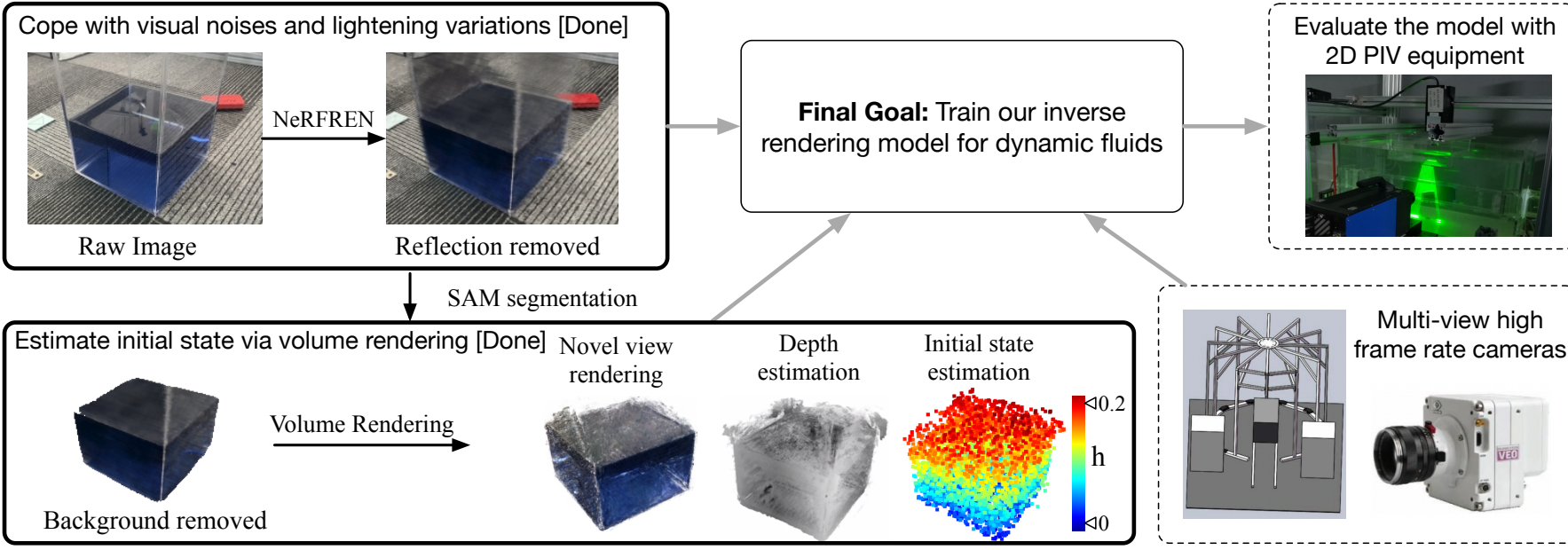


Global Latent



Ours

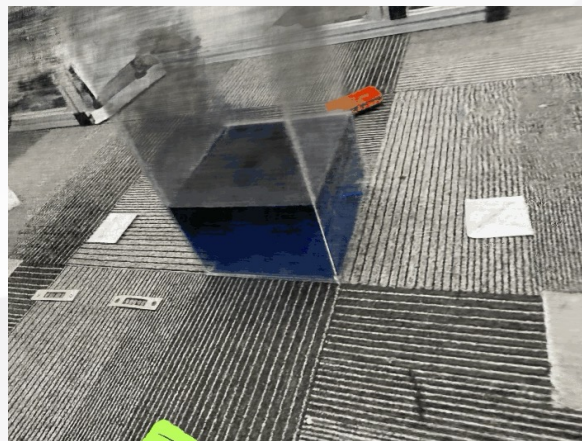




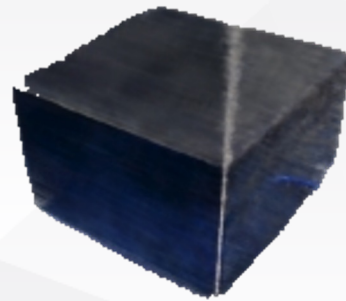
整体流程：

- 拍摄静态场景多视角图像，去除反射光对流体的影响
- 估计初始粒子状态
- 未来将引入高速摄像机对高速流体动态进行拍摄捕捉

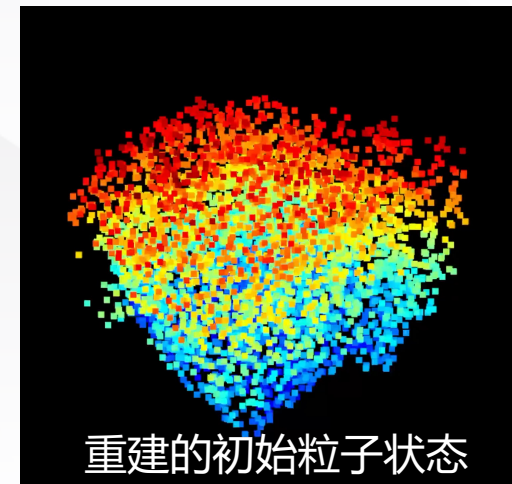
现实场景实验结果



去除反射光



去除背景





上海交通大學

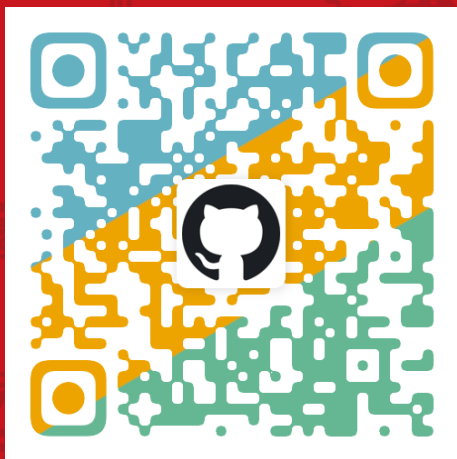
SHANGHAI JIAO TONG UNIVERSITY



人工智能研究院

Artificial Intelligence Institute

谢谢



NeuroFluid



Latent Intuitive Physics

视觉直觉学习项目网站

<https://syguan96.github.io/NeuroFluid/>

<https://sites.google.com/view/latent-intuitive-physics/>

爱国荣校