

VisionLaw: Inferring Interpretable Intrinsic Dynamics from Visual Observations via Bilevel Optimization

Jiajing Lin¹, Shu Jiang², Qingyuan Zeng^{2,3}, Zhenzhong Wang¹, Min Jiang^{1,*}

¹School of Informatics, Xiamen University, China

²Institute of Artificial Intelligence, Xiamen University, China

³The Hong Kong University of Science and Technology (Guangzhou), China

Presenter Name: Jiajing Lin

Project Page: <https://github.com/JiajingLin/VisionLaw>

The first author: Jiajing Lin, jiajinglin@stu.xmu.edu.cn

The corresponding author*: Min Jiang, minjiang@xmu.edu.cn



Why do objects move so differently?



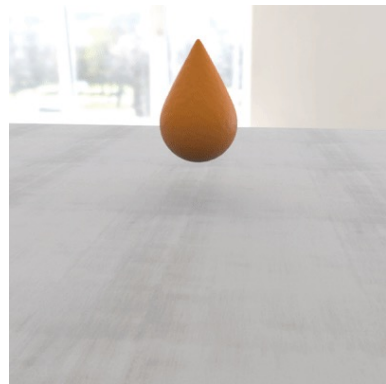
Elasticity



Plasticity



Sand



Newtonian



The **intrinsic dynamics** governs how an object **deforms** and **evolves** under external forces.

It includes:

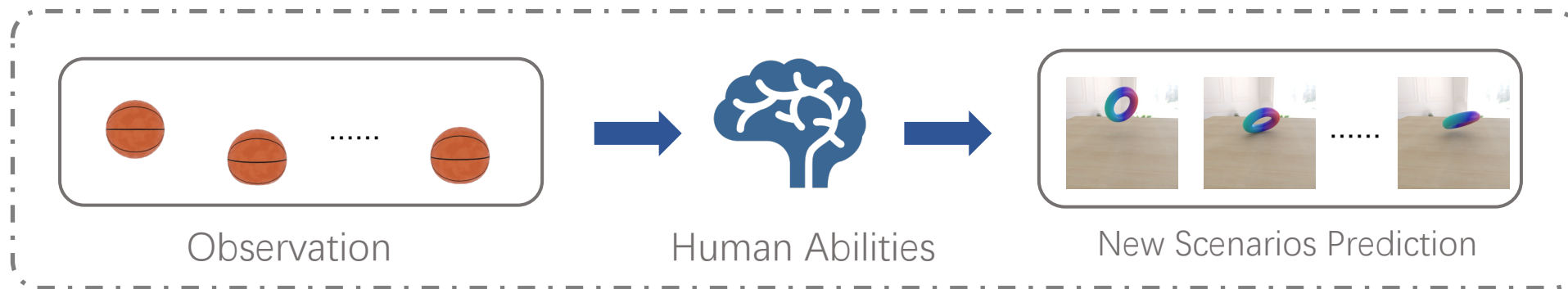
- **material properties** (e.g., stiffness, density)
- **constitutive laws** that describe stress–strain responses

For **physics-driven 4D interaction**, accurately modeling intrinsic dynamics is essential for realistic simulation.

However, current 4D interaction generation relies on **manually specified** intrinsic dynamics.



Human's Intuitive Physics



Humans can intuitively infer the underlying dynamics of an object from its observed motion

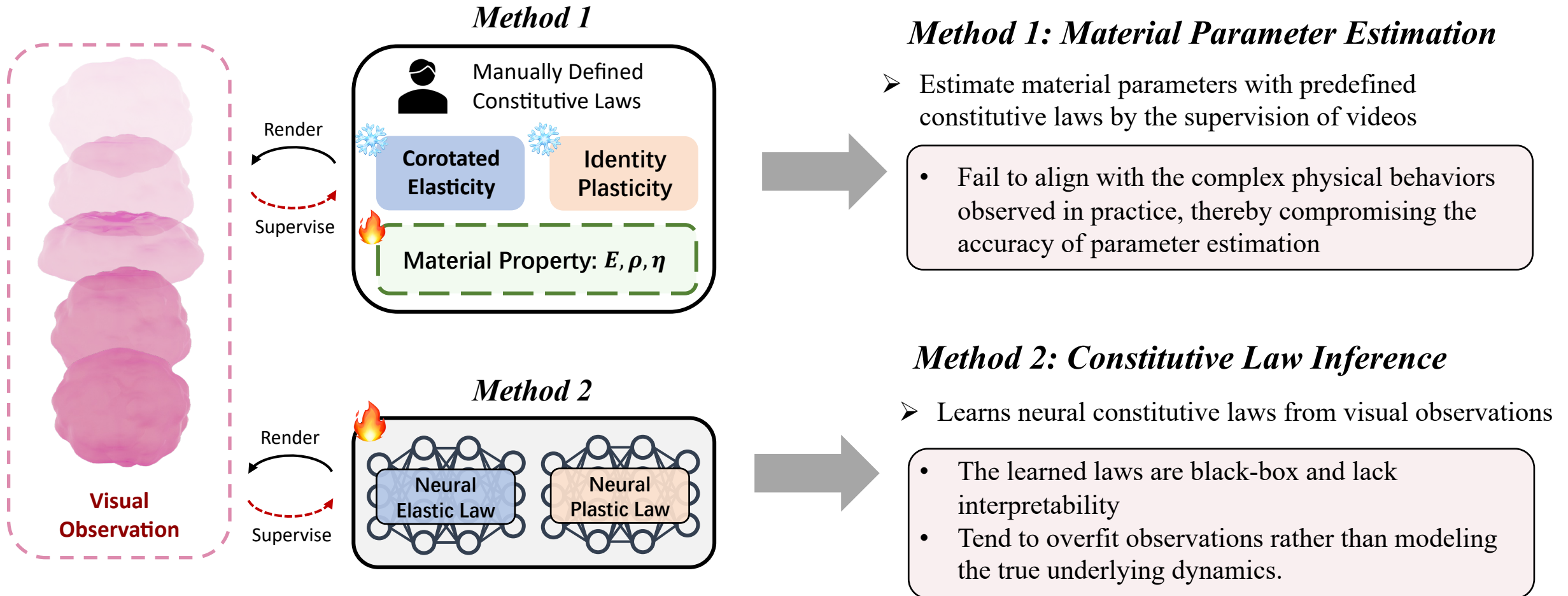


Humans can further generalize this dynamics understanding to new scenarios

can we enable machines to infer the intrinsic dynamics directly from visual observations, as humans do?



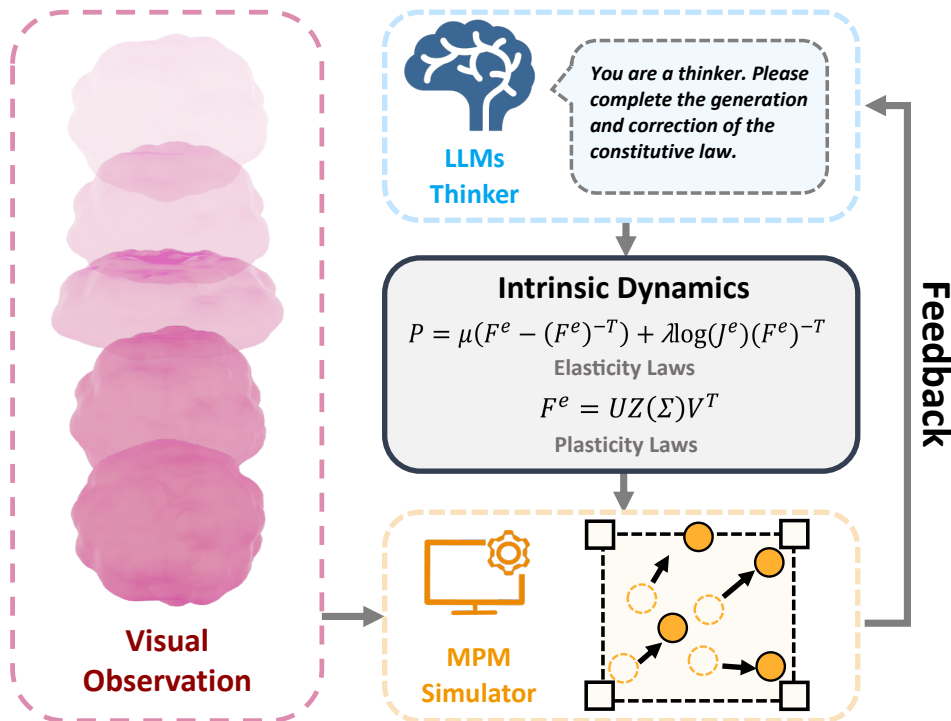
Intrinsic Dynamics Inference from Vision





Core Idea:

We propose an **interpretable intrinsic dynamics inference** framework based on bilevel optimization, which can jointly **infer symbolic constitutive laws** and their corresponding **continuous material properties** solely from **visual observations**



Upper-level Constitutive Evolution

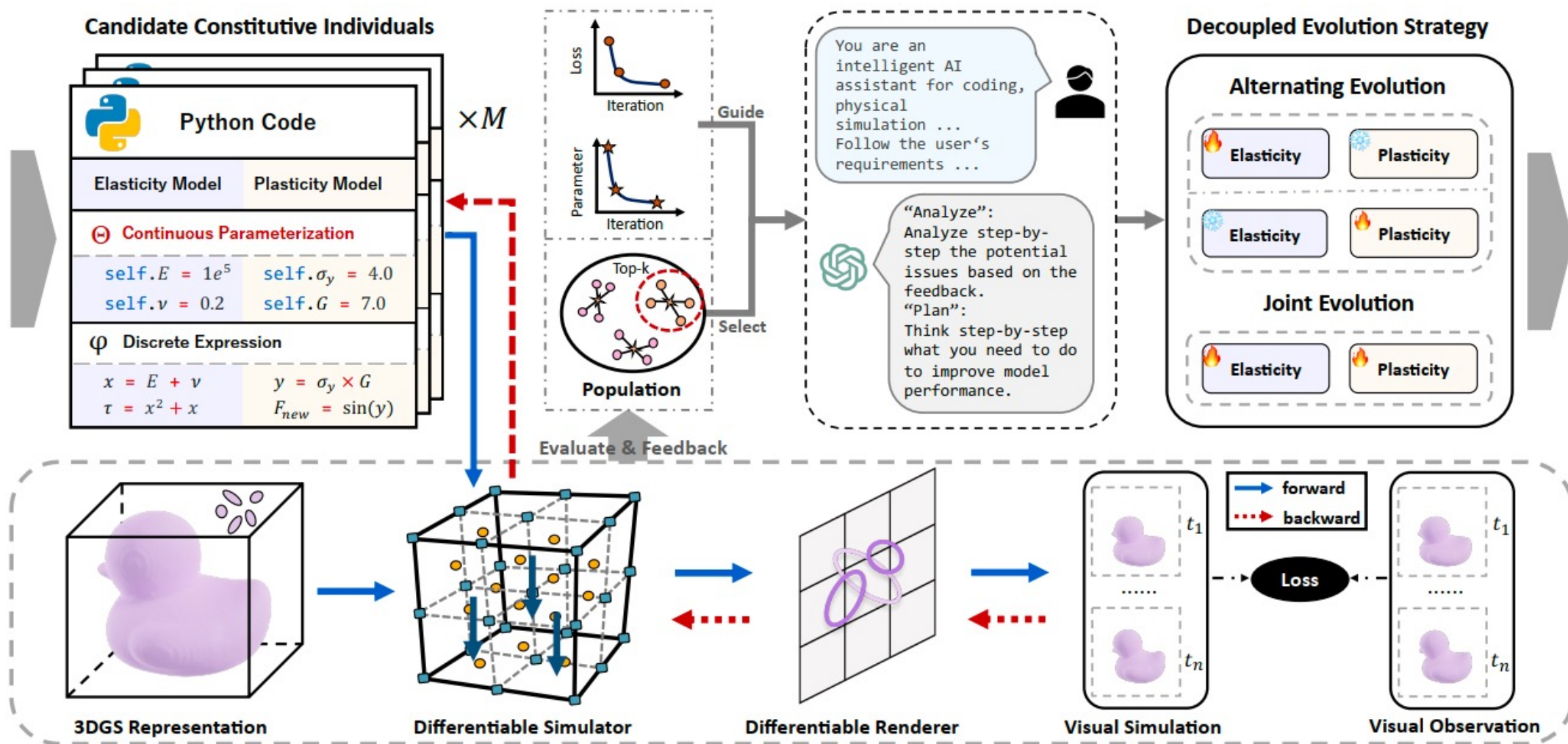
- 1) unleashes the capabilities of LLMs to generate and refine symbolic constitutive hypotheses
- 2) introduces a decoupling mechanism to effectively alleviate the search space explosion caused by jointly evolving elastic and plastic components

Lower-level Constitutive Evaluation

- 1) optimizes the continuous material parameters of a given constitutive law under visual supervision
- 2) provides evaluation and feedback on how well the candidate matches the ground-truth intrinsic dynamics



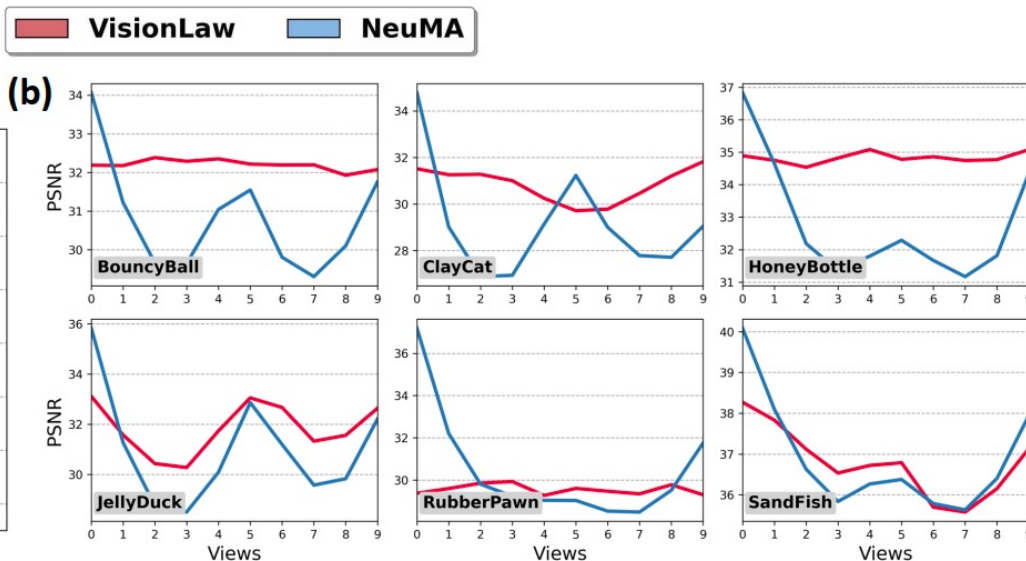
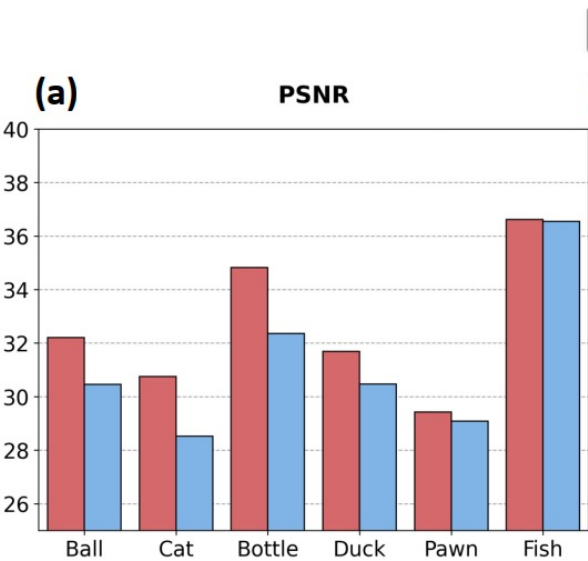
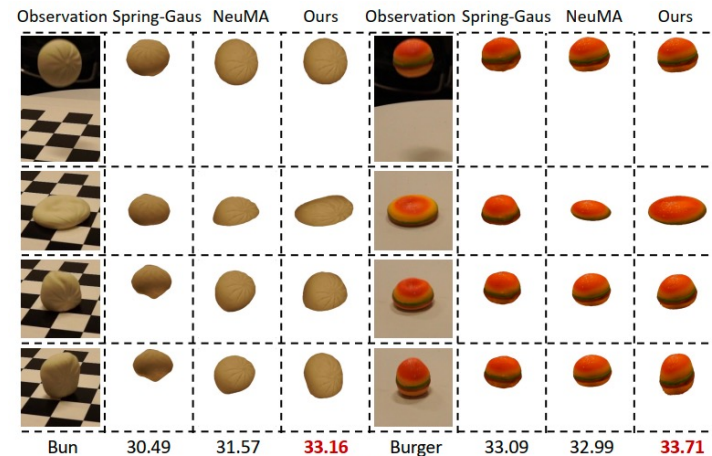
Framework of VisionLaw





Comparison on Synthetic and Real-World Datasets

Method	BouncyBall	ClayCat	HoneyBottle	JellyDuck	RubberPawn	SandFish	Average
PAC-NeRF (Li et al. 2023)	516.30	15.38	2.21	137.73	15.47	1.71	114.80
NeuMA (Cao et al. 2024)	1.78	1.24	1.09	10.96	1.01	1.07	2.86
VisionLaw (Ours)	1.08	0.77	0.79	5.19	0.94	1.10	1.65



```

class PlasticityModel(nn.Module):
    def __init__(self, yield_stress: float = 0.30):
        super().__init__()
        self.yield_stress = nn.Parameter(torch.tensor(yield_stress))
    def forward(self, F: torch.Tensor) -> torch.Tensor:
        U, sigma, Vh = torch.linalg.svd(F)
        sigma_clamped = torch.clamp_min(sigma, 1e-6)
        epsilon = torch.log(sigma_clamped)
        epsilon_mean = epsilon.mean(dim=1, keepdim=True)
        epsilon_dev = epsilon - epsilon_mean
        epsilon_dev_norm = torch.norm(epsilon_dev, dim=1, keepdim=True).clamp_min(1e-12)
        delta_gamma = epsilon_dev_norm - self.yield_stress
        delta_gamma_clamped = torch.clamp_min(delta_gamma, 0.0)
        shrink_factor = 1.0 - delta_gamma_clamped / epsilon_dev_norm
        epsilon_dev_corrected = epsilon_dev * shrink_factor
        epsilon_corrected = epsilon_mean + epsilon_dev_corrected
        sigma_corrected = torch.exp(epsilon_corrected)
        F_corrected = U @ torch.diag_embed(sigma_corrected) @ Vh
        return F_corrected

class ElasticityModel(nn.Module):
    def __init__(self, youngs_modulus_log: float = 9.82, poissons_ratio_sigmoid: float = 4.07):
        super().__init__()
        self.youngs_modulus_log = nn.Parameter(torch.tensor(youngs_modulus_log))
        self.poissons_ratio_sigmoid = nn.Parameter(torch.tensor(poissons_ratio_sigmoid))
    def forward(self, F: torch.Tensor) -> torch.Tensor:
        B = F.size(0)
        E = self.youngs_modulus_log.exp()
        nu = self.poissons_ratio_sigmoid.sigmoid() * 0.49
        mu = E / (2 * (1 + nu))
        lam = E * nu / ((1 + nu) * (1 - 2 * nu))
        I = torch.eye(3, dtype=F.dtype, device=F.device).unsqueeze(0).expand(B, -1, -1)
        J = torch.linalg.det(F).clamp_min(1e-12).view(-1, 1)
        logJ = torch.log(J)
        F_inv = torch.inverse(F)
        F_inv_T = F_inv.transpose(1, 2)
        P = mu * (F - F_inv_T) + lam * logJ * F_inv_T
        Ft = F.transpose(1, 2)
        kirchhoff_stress = torch.matmul(P, Ft)
        return kirchhoff_stress

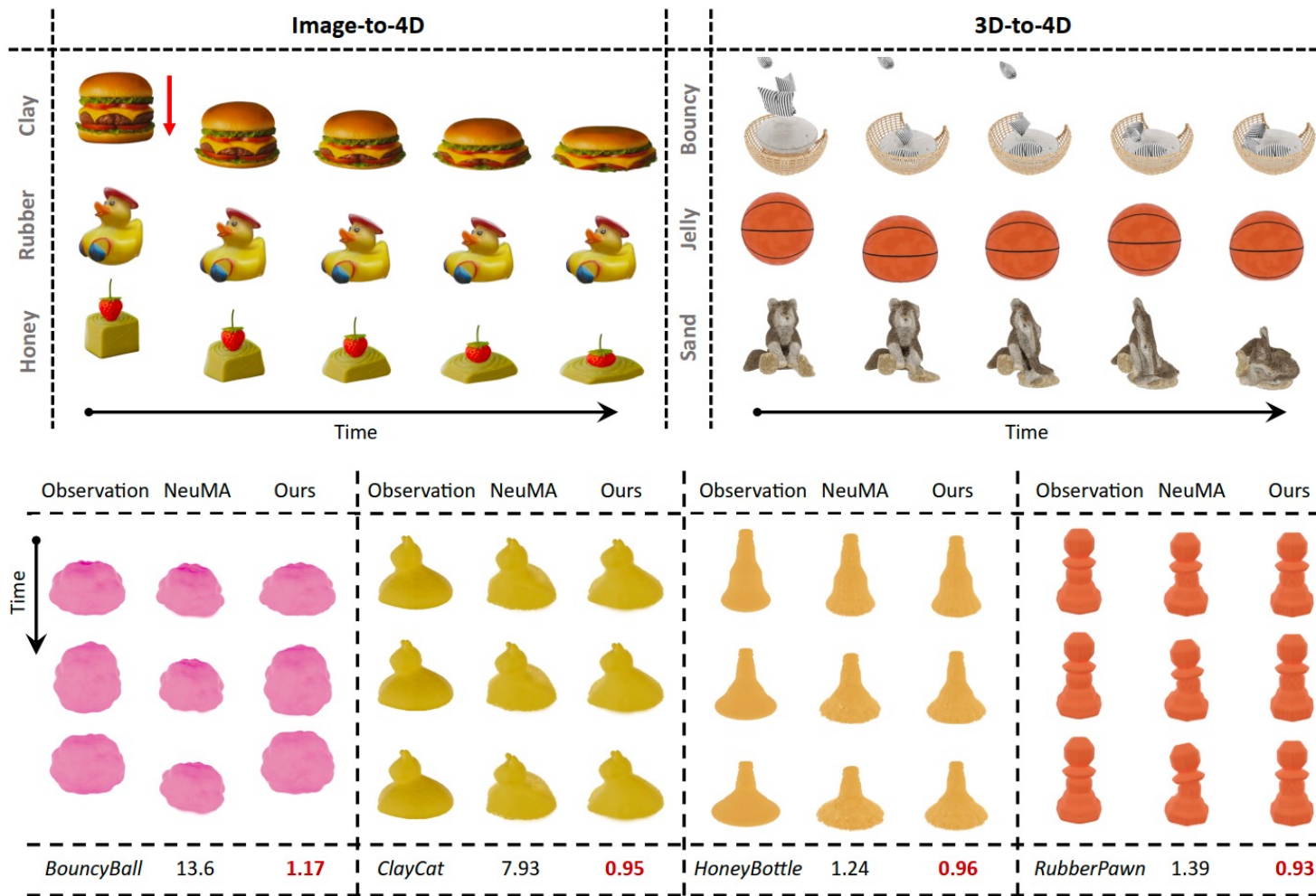
```

Experimental Conclusions:

These results show that our method can more accurately recover the intrinsic dynamics from visual observations while maintaining interpretability



Generalization Analysis



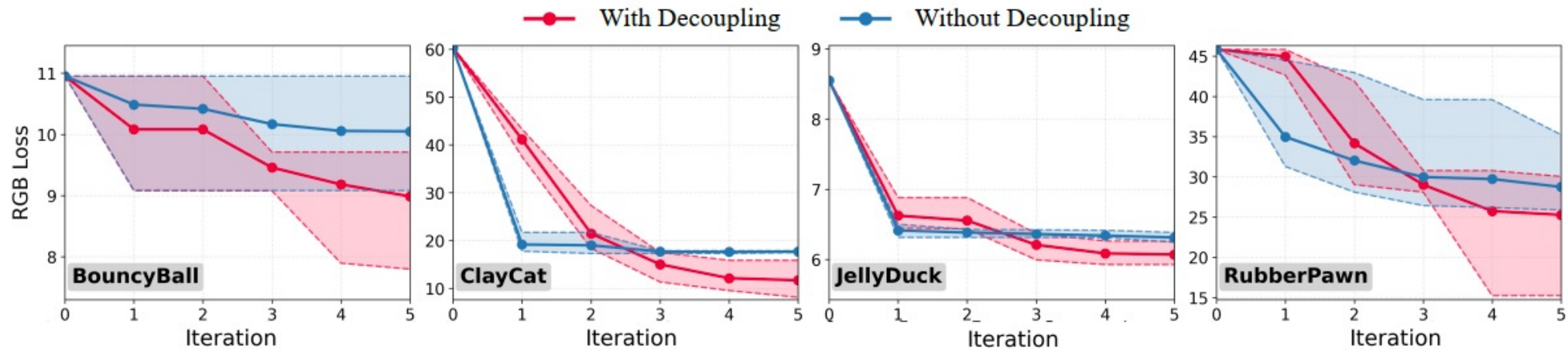
Experimental Conclusions:

- These results show that the intrinsic dynamics inferred by VisionLaw can be transferred to unseen observations and novel scenarios.
- These results further demonstrates that our method learns transferable and physically meaningful dynamics, rather than merely fitting the observed motion.



Ablation on Decoupled Evolution Strategy

Methods	Metric	BouncyBall	ClayCat	HoneyBottle	JellyDuck	RubberPawn	SandFish	Average
w/o decoupling	Accuracy	9.08	17.31	3.64	6.24	25.88	2.24	10.73
	Time	11790	10997	10891	9333	10605	14439	11343
w/ decoupling	Accuracy	7.80	8.13	3.54	5.93	15.25	2.17	7.14
	Time	9319	10636	10699	11249	10839	15689	11405



Conclusion:

These results show that the proposed decoupled evolution strategy significantly improves the quality and stability of constitutive law discovery.



Conclusion

- We propose a bilevel optimization framework that unifies constitutive evolution and visionguided constitutive evaluation, achieving the inference of symbolic constitutive laws and material parameters from visual observations.
- We introduce physical inductive biases through LLMs to guide the evolution of constitutive laws. In addition, a decoupled evolution strategy is proposed to markedly improve both search efficiency and solution quality.
- Extensive experiments on both synthetic and real-world datasets demonstrate that our method effectively captures the interpretable intrinsic dynamics underlying visual observations and generalizes them to novel scenarios for 4D interaction.

Thank you for your attention !

Presenter Name: Jiajing Lin

Project Page: <https://github.com/JiajingLin/VisionLaw>