

BoxTwin: Learning Elastoplastic Articulated Object Dynamics from Videos

Heng Zhang^{*1}, Gehan Zheng^{*1}, Kaifeng Zhang¹, Jay Song¹, Shivansh Patel¹,
Sonny Hu¹, Yunzhu Li¹, Changxi Zheng¹ and Peter Yichen Chen²

Abstract—Digital twins enable robots to anticipate and adapt to physical interactions, but existing models struggle with elastoplastic articulated objects (EAOs) that exhibit nonlinear elasticity, plastic yielding, and damage accumulation. We present BoxTwin, an interactive digital-twin framework that learns the full dynamics of EAOs from videos. Our pipeline reconstructs the scene, identifies a physics-aware constitutive model for each EAO. Experiments on manual folding and dual-arm manipulation of EAOs show that BoxTwin accurately tracks joint trajectories and reproduces post-contact plastic behavior over long horizons. By integrating video-driven reconstruction with elastoplastic-damage modeling, BoxTwin advances digital twins toward predictive, adaptive control of deformable objects in unstructured environments.

I. INTRODUCTION

Integrating digital twins into robotic platforms heralds a new era of safer, more adaptable human-robot collaboration within complex physical settings[1], [2]. By furnishing physics-based real-time replicas of tangible objects, digital twins allow robots to foresee interactions, refine manipulation tactics, and improve decision making amid unstructured environments[3], abilities that are especially vital when handling deformable items. However, modeling deformable objects persists as a major obstacle in robotic manipulation[4], [5], [6], [7], [8], [9], [10], [11]; current techniques often do not cover the full range of mechanical responses elicited during contact. This issue becomes pronounced for articulated elastoplastic objects, which combine reversible elastic deformations with irreversible plastic changes and feature joints that enable intricate motions (e.g. boxes, doors, hinges, sheet-metal assemblies). Unlike rigid or purely elastic bodies, these objects display highly non-linear dynamics that evolve throughout interaction, posing considerable challenges for precise simulation and control.

The difficulty arises from three tightly coupled physical effects. First, non-linear elasticity produces stress-strain relationships far from Hookean behavior, leading to energy dissipation and interdegree-of-freedom coupling that a linear model cannot capture. Second, plastic deformation introduces permanent shape alterations once critical loads are exceeded, establishing new rest configurations that defy prediction by elastic models alone. Third, damage evolution causes elastoplastic material parameters to drift over repeated loading cycles, demanding time-dependent material

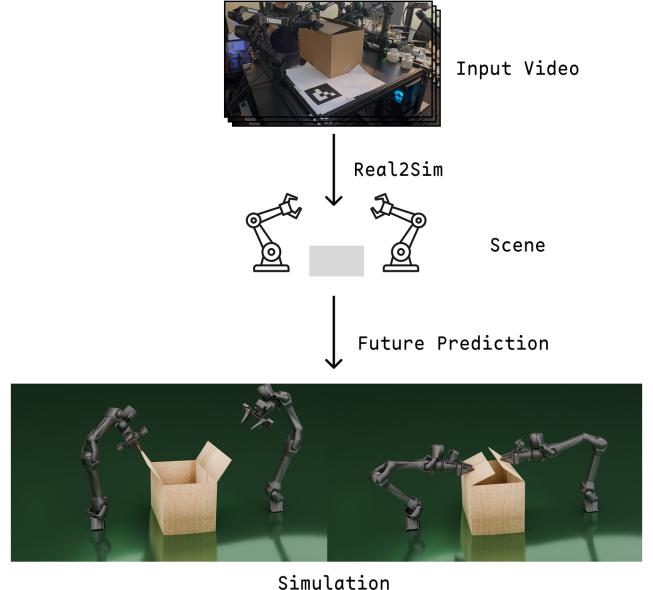


Fig. 1: Our pipeline begins by reconstructing the entire scene—from the input video we recover both the EAOs and the robots. With this reconstructed scene in hand, we can then perform future prediction.

descriptions. Conventional digital twin frameworks, which are typically designed for rigid bodies or linearly elastic materials, do not accommodate this triad of phenomena. Prior research has largely focused on either rigid dynamics[12], [13], [14] or static elastic deformation[15], [16], [17], leaving the intertwined plasticity, nonlinear elasticity, and progressive damage of elastoplastic articulated objects (EAOs) unaddressed. As a result, robots frequently fail in tasks involving grasping, folding, or forecasting post-interaction states, jeopardizing safety and reliability in both domestic and industrial contexts.

To overcome these limitations, we introduce BoxTwin, an interactive digital-twin system specifically designed for EAOs. Instead of relying on hand-crafted physics models with static parameters, BoxTwin incorporates a physics-aware architecture that captures the full dynamics of such objects, including nonlinear elasticity, plastic yield thresholds, and damage progression, through calibrated constitutive models. The system continuously updates material properties to reflect time-varying damage accumulation, enabling robust simulation and control of EAOs in real-world robotic tasks. BoxTwin lays the groundwork for interactive digital twins that enable safe, adaptive robotic manipulation

^{*} Equal contribution

¹Heng Zhang, Gehan Zheng, Kaifeng Zhang, Jay Song, Shivansh Patel, Sonny Hu, Yunzhu Li and Changxi Zheng are with SceniX, Inc.

²Peter Yichen Chen is with the University of British Columbia.

in real-world environments.

II. METHOD

A. Articulated Body

An EAO is represented as a collection of rigid links that are connected by ideal hinge joints modeling the flexible creases of the structure. The topology is captured by a rooted kinematic tree

$$\mathcal{G} = (\mathcal{V}, \mathcal{E});$$

a vertex $i \in \mathcal{V}$ denotes a link \mathcal{L}_i equipped with a body-fixed frame \mathcal{F}_i , while an edge $e = (p, i) \in \mathcal{E}$ corresponds to a hinge joint that allows a single relative rotation about a known axis \mathbf{a}_e . The joint angle $q_e \in \mathbb{R}$ is the only configuration variable associated with that edge.

The homogeneous transform that maps the world frame to link i is obtained by composing the transforms along the unique path from the root to i :

$$\mathbf{T}_{0 \rightarrow i}(\mathbf{q}) = \mathbf{T}_{0 \rightarrow p} \mathbf{T}_{p \rightarrow i}(q_e),$$

$$\mathbf{T}_{p \rightarrow i}(q_e) = \begin{bmatrix} \mathbf{R}(\mathbf{a}_e, q_e) & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix},$$

where $\mathbf{R}(\mathbf{a}_e, q_e)$ is the rotation generated by the hinge angle. Collecting all joint angles into the vector

$$\mathbf{q} = [q_{e_1}, \dots, q_{e_{|\mathcal{E}|}}]^\top$$

provides a compact description of the configuration. Joint-space dynamics follow the standard Lagrange formulation

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}) = \boldsymbol{\tau},$$

where $\mathbf{M}(\mathbf{q})$ is the inertia matrix, $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})$ combines Coriolis and centrifugal effects, $\mathbf{g}(\mathbf{q})$ denotes the generalized gravity forces, and $\boldsymbol{\tau}$ collects the generalized forces generated by external contacts.

B. Elastic response of the creases

The torque produced by a crease is highly nonlinear with respect to its bending deformation and cannot be faithfully captured by a simple linear spring. Let q_e be the instantaneous hinge angle and \bar{q}_e the associated rest angle; the deformation that drives the elastic response is the scalar gap

$$\Delta_e = q_e - \bar{q}_e.$$

The elastic torque is introduced as a smooth, differentiable mapping

$$\tau_e^{\text{el}} = f_{\theta_e}(\Delta_e),$$

where $f_{\theta_e} : \mathbb{R} \rightarrow \mathbb{R}$ is parameterised by a set of coefficients θ_e that are specific to edge e . This formulation captures the strongly nonlinear hinge behaviour while remaining amenable to analytical differentiation, which is required for subsequent identification and optimization tasks.

C. Plastic deformation

Permanent set of a crease is modelled by allowing the rest angle \bar{q}_e to evolve. An internal scalar variable α_e records the accumulated plastic rotation, so that

$$\bar{q}_e = q_e^0 + \alpha_e,$$

with q_e^0 denoting the nominal angle. Plastic flow is activated once the elastic gap exceeds a material-specific yield threshold Δ_e^y . The yield condition

$$\Phi_e(\Delta_e) = |\Delta_e| - \Delta_e^y \leq 0$$

defines an elastic regime ($\Phi_e \leq 0$). When $\Phi_e > 0$ the internal variable evolves according to a smooth flow rule

$$\dot{\alpha}_e = g_{\psi_e}(\Delta_e),$$

where $g_{\psi_e} : \mathbb{R} \rightarrow \mathbb{R}$ is a function parameterised by coefficients ψ_e .

D. Damage evolution

Repeated folding induces micro-fracture, fibre delamination and other degradation mechanisms, which manifest as a gradual loss of stiffness and a reduced capacity to accommodate permanent set[18], [19]. To capture this effect a scalar damage variable $d_e \in [0, 1]$ is introduced for each hinge; $d_e = 0$ corresponds to an undamaged crease and $d_e = 1$ to a mechanically inactive joint. Damage attenuates both the elastic torque and the plastic flow:

$$\tau_e^{\text{el}} = (1 - d_e) f_{\theta_e}(q_e - \bar{q}_e), \quad \dot{\alpha}_e = (1 - d_e) g_{\psi_e}(q_e - \bar{q}_e).$$

The multiplicative form preserves the sign and monotonicity of the underlying maps while ensuring a physically meaningful reduction of the joint's load-bearing capability. The evolution of d_e is driven by the cumulative amount of plastic slip that has occurred at the hinge. Define the plastic-slip accumulator

$$\mathcal{P}_e(t) = \int_0^t |\dot{\alpha}_e(\tau)| \, d\tau,$$

which records the total absolute plastic rotation experienced up to time t . The damage-rate law is taken as a smooth, monotone function of this scalar measure,

$$\dot{d}_e = h_{\xi_e}(\mathcal{P}_e),$$

where $h_{\xi_e} : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ is parameterised by a modest set of coefficients ξ_e .

III. EVALUATION

To assess the effectiveness of the proposed model, we carried out two experiments. In the first, we manually deformed an EAO composed of two panels and a single joint, keeping one panel anchored to the table. In the second experiment, we employed Aloha robotic arms to manipulate the EAOS and reproduced the entire sequence within a simulation environment. Both experiments were implemented using the MuJoCo simulator[20].

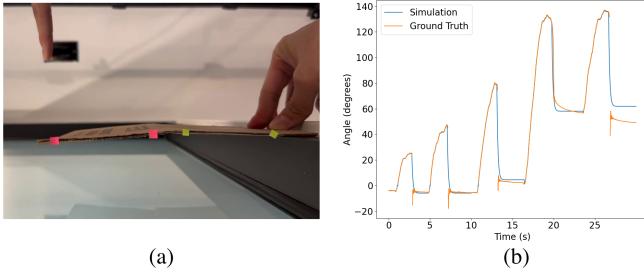


Fig. 2: Figure (a) depicts the experimental setup. Figure (b) showcases the expressive power of our proposed model in capturing the long-horizon dynamics of the EAOs.

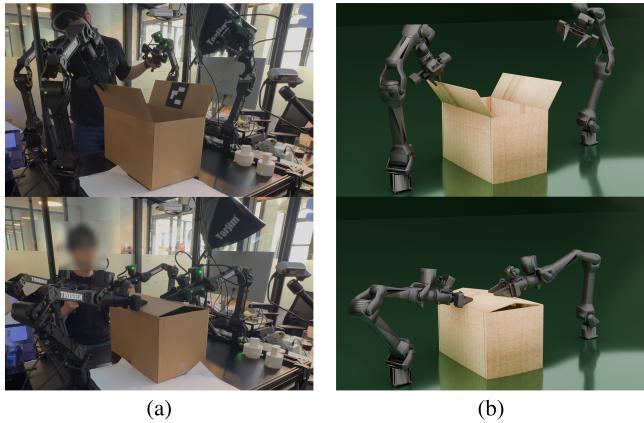


Fig. 3: Figure (a) shows the frames captured from the real-world video, and Figure (b) presents the corresponding simulated views.

A. Folding Test

To capture the joint angle of the EAO, color markers are affixed to the side of the panel; their positions are continuously monitored, allowing the joint angle to be derived as illustrated in Figure 2(a). Subsequently, the same actuation applied to the physical object is transferred in simulation, and the resulting joint angles are recorded throughout the motion. Figure 2(b) presents the simulated joint angle trajectory together with the corresponding measurements from the real-world EAO.

Our proposed model captures the actual behavior of the object across the long sequence, faithfully reproducing both the instantaneous kinematic responses and the gradual evolution of material properties; it accurately tracks the joint-angle trajectories despite the accumulation of hysteresis and wear, preserves the nuanced interplay between elastic, plastic, and damage mechanisms throughout repeated actuation cycles, and thereby provides a robust, high-fidelity digital twin that remains reliable even as the system experiences prolonged exposure to complex loading patterns.

B. Prediction Test

In this experiment we command the Trossen Aloha dual-arm platform to interact with a variety of EAOs. The system operates in a leader–follower configuration: the leader

arms receive the high-level motion commands, while the follower arms execute the physical actions on the objects. Both simple EAOs that possess a single revolute joint and more complex assemblies that contain several coupled joints are handled by the same control framework. The follower arms are responsible for the complete manipulation sequence. They approach the object, apply the necessary forces to translate, rotate, or fold its individual links, and release it once the desired configuration is achieved.

After each trial we logged the full follower-arm trajectory, including joint positions, velocities, torques and end-effector pose at every step. Using these data we reconstructed an identical virtual scene in a physics-based simulator: the Trossen Aloha dual-arms were instantiated with the same kinematic and dynamic parameters (link lengths, masses, joint limits, friction), and the EAOs were modeled with the experimentally identified parameters. The recorded control signals were then replayed on the simulated follower arms, preserving the exact timing and magnitude of every command. By comparing the simulated joint angles and overall object configurations with the measurements obtained on the physical hardware, we observed a close correspondence across all test cases, from single-joint modules to multi-joint assemblies. Figure 3 shows the replay results for a box with multiple joints. This validation demonstrates that the simulated environment can be reliably used for further algorithm development, predictive analysis, and transfer-learning studies without the need for repetitive physical trials.

IV. CONCLUSIONS

We introduce the first interactive digital-twin framework that captures the full dynamics of elastoplastic articulated objects (EAOs) during real-world robotic manipulation. By integrating a video-driven reconstruction pipeline that simultaneously handles nonlinear elasticity, plastic deformation, and damage evolution, our physics-aware twins predict post-contact plastic behavior—something prior EAO models could not achieve. BoxTwin thus moves beyond static simulations, enabling adaptive, predictive control of deformable objects in unstructured environments where rigid-body assumptions fail.

REFERENCES

- [1] H. Jiang, H.-Y. Hsu, K. Zhang, H.-N. Yu, S. Wang, and Y. Li, “Physitwin: Physics-informed reconstruction and simulation of deformable objects from videos,” *arXiv preprint arXiv:2503.17973*, 2025.
- [2] A. A. Malik and A. Brem, “Digital twins for collaborative robots: A case study in human-robot interaction,” *Robotics and Computer-Integrated Manufacturing*, vol. 68, p. 102092, 2021.
- [3] Y. Mu, T. Chen, Z. Chen, S. Peng, Z. Lan, Z. Gao, Z. Liang, Q. Yu, Y. Zou, M. Xu *et al.*, “Robotwin: Dual-arm robot benchmark with generative digital twins,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 27 649–27 660.
- [4] Y. Feng, Y. Shang, X. Li, T. Shao, C. Jiang, and Y. Yang, “Pie-nerf: Physics-based interactive elastodynamics with nerf,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 4450–4461.
- [5] Y. Jiang, C. Yu, T. Xie, X. Li, Y. Feng, H. Wang, M. Li, H. Lau, F. Gao, Y. Yang *et al.*, “Vr-gs: A physical dynamics-aware interactive gaussian splatting system in virtual reality,” in *ACM SIGGRAPH 2024 Conference Papers*, 2024, pp. 1–1.

[6] T. Xie, Z. Zong, Y. Qiu, X. Li, Y. Feng, Y. Yang, and C. Jiang, “Phys-gaussian: Physics-integrated 3d gaussians for generative dynamics,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 4389–4398.

[7] H.-y. Chen, E. Treitschke, T. Stuyck, P. Kadlecak, L. Kavan, E. Vouga, and C. Lassner, “Virtual elastic objects,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 15 827–15 837.

[8] X. Li, Y.-L. Qiao, P. Y. Chen, K. M. Jatavallabhula, M. Lin, C. Jiang, and C. Gan, “Pac-nerf: Physics augmented continuum neural radiance fields for geometry-agnostic system identification,” *arXiv preprint arXiv:2303.05512*, 2023.

[9] Y.-L. Qiao, A. Gao, and M. Lin, “Neuphysics: Editable neural geometry and physics from monocular videos,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 12 841–12 854, 2022.

[10] T. Zhang, H.-X. Yu, R. Wu, B. Y. Feng, C. Zheng, N. Snavely, J. Wu, and W. T. Freeman, “Physdreamer: Physics-based interaction with 3d objects via video generation,” in *European Conference on Computer Vision*. Springer, 2024, pp. 388–406.

[11] L. Zhong, H.-X. Yu, J. Wu, and Y. Li, “Reconstruction and simulation of elastic objects with spring-mass 3d gaussians,” in *European Conference on Computer Vision*. Springer, 2024, pp. 407–423.

[12] M. Anitescu, F. A. Potra, and D. E. Stewart, “Time-stepping for three-dimensional rigid body dynamics,” *Computer methods in applied mechanics and engineering*, vol. 177, no. 3-4, pp. 183–197, 1999.

[13] L. Lan, D. M. Kaufman, M. Li, C. Jiang, and Y. Yang, “Affine body dynamics: Fast, stable & intersection-free simulation of stiff materials,” *arXiv preprint arXiv:2201.10022*, 2022.

[14] S. Stramigioli, *Modeling and IPC control of interactive mechanical systems—A coordinate-free approach*. Springer, 2001.

[15] V. Modi, N. Sharp, O. Perel, S. Sueda, and D. I. Levin, “Simplicits: Mesh-free, geometry-agnostic elastic simulation,” *ACM Transactions on Graphics (TOG)*, vol. 43, no. 4, pp. 1–11, 2024.

[16] Z. Xie, Z. Liu, Z. Peng, W. Wu, and B. Zhou, “Vid2sim: Realistic and interactive simulation from video for urban navigation,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 1581–1591.

[17] M. Li, Z. Ferguson, T. Schneider, T. R. Langlois, D. Zorin, D. Panozzo, C. Jiang, and D. M. Kaufman, “Incremental potential contact: intersection-and inversion-free, large-deformation dynamics.” *ACM Trans. Graph.*, vol. 39, no. 4, p. 49, 2020.

[18] E. Linvill, M. Wallmeier, and S. Östlund, “A constitutive model for paperboard including wrinkle prediction and post-wrinkle behavior applied to deep drawing,” *International Journal of Solids and Structures*, vol. 117, pp. 143–158, 2017.

[19] D. W. Coffin and M. Nygård, “Creasing and folding,” in *Transactions of the 16th fundamental research symposium*, 2017, pp. 69–136.

[20] E. Todorov, T. Erez, and Y. Tassa, “Mujoco: A physics engine for model-based control,” in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 5026–5033.