

Physics-Encoded Graph Neural Networks for Deformation Prediction under Contact

Mahdi Saleh¹, Michael Sommersperger¹, Nassir Navab², Federico Tombari³

Abstract—In robotics, it’s crucial to understand object deformation during tactile interactions. A precise understanding of deformation can elevate robotic simulations and have broad implications across different industries. We introduce a method using Physics-Encoded Graph Neural Networks (GNNs) for such predictions. Similar to robotic grasping and manipulation scenarios, we focus on modeling the dynamics between a rigid mesh contacting a deformable mesh under external forces. Our approach represents both the soft body and the rigid body within graph structures, where nodes hold the physical states of the meshes. We also incorporate cross-attention mechanisms to capture the interplay between the objects. By jointly learning geometry and physics, our model reconstructs consistent and detailed deformations. We’ve made our code and dataset public to advance research in robotic simulation and grasping. †

Index Terms—Contact Modelling; Simulation and Animation; Manipulation and Grasping;

I. INTRODUCTION

Grasping is a cornerstone of robotic manipulation, and accurate deformation prediction is pivotal for a stable and safe grip, especially with delicate or flexible objects. Miscalculations can result in improper handling, dropped items, or even damage. This precision becomes even more crucial in specialized fields like medical surgeries. Specifically, surgical tools must interact with delicate structures during procedures involving soft tissues, requiring high precision to prevent unintentional harm and guarantee patient safety [1].

Traditional techniques like Finite Element Methods (FEM) [2], [3], Boundary Element Methods (BEM) [4], [5], and Dual Mortar Methods [6], [7] have been the widely used to predict deformation for a long time. However, they often struggle with complex, detailed meshes due to their computational intensity. Their optimization process can be tedious and time-consuming, especially for high-resolution models.

Moving beyond classical methods, particle-based methods [8] offer a more flexible approach. Systems like Smoothed Particle Hydrodynamics (SPH) [9] or Mass-Spring Systems [10] approach deformation prediction by breaking down the object into a series of interconnected particles. These methods can handle more complex deformations but still can

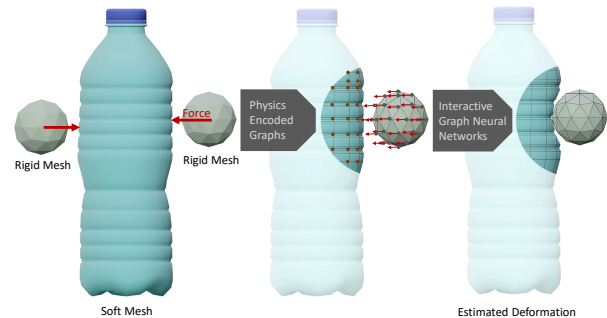


Fig. 1. Given the input meshes of a soft and rigid body and their contact, we embed physical states into mesh graphs. Our network then processes the graphs to understand their interaction and predicts the deformation of the soft mesh.

not ensure the accuracy of the predicted deformation on the original mesh.

Meanwhile, graph-based representations have gained traction in computer vision and graphics, especially in registration and deformable shape matching [11], [12]. These methods can learn and match deformation patterns by viewing 3D shapes as a network of interconnected nodes. Graphs are particularly relevant for our problem formulation since we can encapsulate the physical characteristics of an object into the graph embeddings, ensuring a more detailed and structured representation.

Contrary to subsampled and holistic graph structures, we aim to use the original triangle mesh. By interpreting the connections of the triangle mesh as graph edges, our approach preserves the mesh structure, ensuring accurate deformation predictions without the additional steps typically associated with graph construction. We balance the reliability of classical methods and the flexibility of graph-based representations while reducing the computational costs. As depicted in Fig. 1, we introduce a novel learning pipeline that embeds physics information, such as force features, into the graph nodes of both soft and rigid objects. Using GNNs, we capture the geometric shape features, and by using a cross-attention mechanism, we allow messages to pass between the soft and the rigid graphs. Finally, we reconstruct the post-contact mesh from the conditioned encoded features.

Our methodology can learn intricate surface deformations through a data-driven approach. The proposed network, adapted for our objectives, aims to bridge the gap between precision and computational efficiency. As the fields of robotic grasping and medical surgeries evolve, our integrated

Corresponding author: Mahdi Saleh (m.saleh@tum.de)

¹ M. Saleh, M. Sommersperger, are with Department of Computer Science, Technische Universität München, München 85748 Germany.

² N. Navab is a full professor and head of the Chair for Computer Aided Medical Procedures & Augmented Reality, Technical University of Munich, 85748 Munich, Germany.

³ Federico Tombari is with the Faculty of Computer Science, Technische Universität München, 85748 Garching bei München, Germany, and also with the Google, 8002 Zurich, Switzerland (e-mail: tombari@in.tum.de).

†Project page: <https://mahdi-slh.github.io/deform>

approach to learning physics and geometry together will significantly influence their direction.

We highlight our work’s main contributions:

- We curated a dataset of everyday objects with varied physical characteristics, associated metadata of contact, and deformations released together with the paper.
- We innovatively integrate the physical state of both soft and rigid bodies and their interaction into graph representations.
- Our novel GNN framework learns geometry and the underlying physics between objects, able to predict the distortion and deformations upon contact for different applications.

II. RELATED WORK

A. Classical Methods

This section discusses classical and non-learning methods to estimate surface deformations. FEM provides a fundamental approach for estimating surface deformations by discretizing the problem domain into smaller elements. FEM has been widely used for deformation of plastic and other materials under stress or contact. [2], [3], [13]. Boundary Element Methods (BEM), by contrast, focus on the boundary rather than the entire domain, optimizing computational efficiency for certain problems [4], [5]. Mass-spring models are simplistic representations where a system is depicted using masses and springs to estimate deformations [10].

Particle-based methods, such as Smoothed Particle Hydrodynamics (SPH), utilize particles to approximate deformations [8]. Dual mortar methods are mathematical formulations that allow for a more precise integration over contact interfaces [6], [7]. However, they have complexity in implementation for many problems. Classical methods for surface deformation estimation provide a solid foundation and are used in various engineering applications. While classical methods offer robustness, learned techniques can promise efficiency and adaptability.

B. Neural Physics

Recent advances combine data-driven machine learning with physics, creating a paradigm known as Physics-Informed Machine Learning [14]. It enhances model performance by incorporating physical prior knowledge. While this interdisciplinary field has progressed, it remains underexplored in areas like robotic control [15]. The work of [16] showcases a Neural Physics Engine, a neural network architecture that breaks down object dynamics into pairwise interactions, though it’s mainly applied to simple two-dimensional dynamics. In a different approach, [17] uses deep learning models trained on analytical solutions to predict deformations in poroelastic media. Meanwhile, [18] merges subspace simulation techniques with machine learning for efficient physics simulations, mainly focusing on dynamic cloths interacting with external objects. [19] offers a framework for deformable solid dynamics using autoencoder neural networks, emphasizing a condensed representation through implicit integration. Additionally, [20]

employs physics-informed neural networks to model crustal deformations. [21] and [21] leverage GNNs to predict cloth deformation and particle dynamics. Lastly, [22] uses GNNs to understand material mechanics and predict different materials’ structural deformations.

C. Graph Neural Networks

Graph Neural Networks (GNNs) are extensively employed to process 3D geometry, including meshes and point clouds. Unlike point-based methods, which lack geometric intuitions [23], GNNs leverage graphs to facilitate information exchange between neighboring points. Graph Attention Networks (GATs) employ masked self-attentional layers tailored for graph data [24]. The multi-head attention concept can also be combined with point and graph-based methods to attend to local and global features [25], [26]. In parallel, Spectral CNN [27] facilitates 3D Shape Segmentation through weight sharing in a spectral domain through graph Laplacian eigenbases. EdgeConv dynamically refines edges across network layers, improving point cloud classification and segmentation tasks [28]. Graphs have been then used for rigid [11], [12] and non rigid registration [11], [29]. For instance, Graphite [12] extracts keypoints and representations from a fixed spatial graph. Meanwhile, Bending Graphs uses mesh structures, offering a hierarchical approach to deformable shape matching [11]. Despite the capabilities of GNNs, much of the graph-centric and Geometric Deep Learning [30] research is restricted to correspondence estimation.

D. Hand/Robot Object Interaction

Robotic manipulation, once primarily focused on rigid objects, now frequently engages with deformable entities that alter shape upon contact [31]. Such interactions make the underlying physics of deformation intricate and challenging to capture accurately. [32] proposes a subspace method for simulating dynamic deformations, merging subspace methods with learning-based techniques. Subsequent research by the same authors enhances these simulations with more detailed contact-driven deformations [33]. In another study on deformable object simulations, [34] utilizes the hyperbolic space’s unique properties for reconstructing hand and object meshes from images. Addressing deformations in a medical context, [35] offers the PhysGNN framework, using GNNs to approximate soft tissue deformation during neurosurgical procedures.

While existing techniques offer significant insights into deformation prediction, the trade-offs between computational efficiency and precision become complex, especially in real-time scenarios. Recognizing this, our method aims to utilize the robustness of classical techniques, the adaptability of learning-based approaches, and the intricate modeling of GNNs, all integrated into a cohesive framework.

III. METHODOLOGY

We present a compact pipeline to predict post-contact deformations by integrating physics encoding into GNNs. Our pipeline is also illustrated in Fig 2.

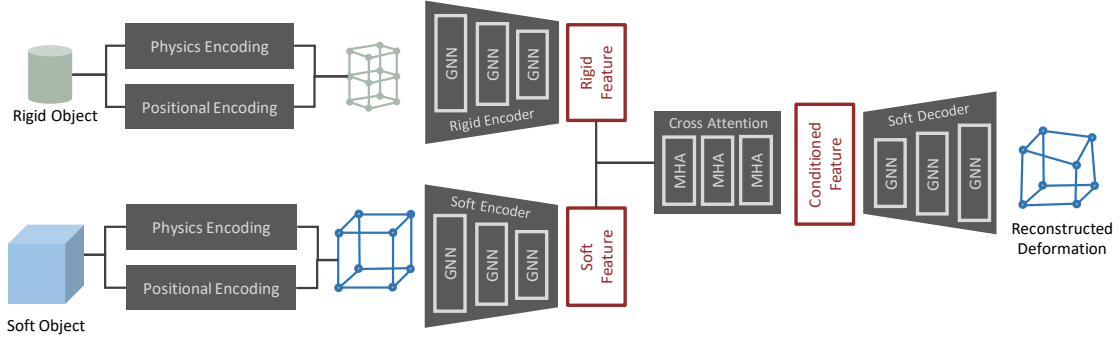


Fig. 2. Our pipeline unfolds in three stages: First, we integrate the physics and positional encoding for both soft and rigid objects. Second, we engage in the encoder phase to extract features for each body. Finally, we employ multi-head attention to facilitate interactions between the features of soft and rigid objects, subsequently leveraging the conditioned feature to decode and reconstruct the deformed mesh or graph.

A. Physics Encoding

This subsection details the process of graph generation and representation. Given a triangle mesh, each vertex is interpreted as a node in the graph. While existing KNN graphs might work effectively for smaller datasets, querying nearest neighbors becomes computationally expensive as the number of vertices increases. Therefore, we produce three edges from each triangular mesh face, connecting each vertex. The information about neighboring vertices and how they relate to each other is preserved by referring to triangle meshes and can be easily manipulated or queried.

Let $P \in \mathbb{R}^3$ denote the 3D positions of these vertices. These 3D coordinates, similar to the encoding in [11], are transformed into log-frequency representations to create more discriminative features. Moreover, we incorporate physics-related attributes for each node, such as force vectors, denoted by $F \in \mathbb{R}^4$. These initial attributes represent the normalized force vector and its magnitude.

The complete node feature for a given vertex then becomes a stack of its position information P and the associated force information F . They are passed through multilayer perceptron (MLP) layers to encode these stacked features further. The resultant encoded node features serve as input for our graph neural network in the subsequent subsection.

B. Network Architecture

Our network is designed to predict how points (or nodes) in a relaxed setting (the "soft" graph) change their positions when they come into contact with another entity, the so-called "rigid" graph. The architecture consists of layers of GNN for both the soft and rigid graphs, an interaction mechanism using Multi-Head Attention (MHA) [36], and finally, a decoder to reconstruct the object following the interaction.

GNN Encoders: We take x_S and x_R as the initial state of our soft and rigid graphs, respectively. These features are processed and refined through GNN layers applied iteratively. Our GNN layers are based on TAGConv (Topology Adaptive Graph Convolution) [37]. TAGConv operates with varying

topologies using adaptive, fixed-size filters for graph convolutions, ensuring more accurate geometric representations.

Feature Interactions: To establish interactions between the soft and rigid graphs, we leverage a cross-attention mechanism with multi-head attention [36]. In this context, soft features from the resting graph and rigid features from the collider graph are input to the attention mechanism to capture intricate dependencies and relations between them. These interactions, illustrated in Fig. 2, allow us to condition the soft body features using the information from the rigid body, resulting in a richer, more representative feature set given the physics.

Soft Decoder: Following this attention-based interaction, we obtain a conditioned feature representation that captures the mutual influence of the soft and rigid bodies on each other. These conditioned features represented as x_C are then processed by a graph-based decoder. This decoder refines and translates the conditioned features into the deformed mesh representation.

The output from the decoder, denoted as $P_D \in \mathbb{R}^3$, signifies the predicted deformed positions of nodes in the resting graph. The predicted graph output is directly used as the reconstructed mesh.

C. Loss Functions

Mean Squared Error (MSE): The MSE Loss measures the mean squared difference between the predicted positions (P_D) of nodes in the graph and the ground truth deformations of nodes (P_D^*). The MSE Loss is given by:

$$L_{mse} = \frac{1}{N_{nodes}} \sum_{i=1}^{N_{nodes}} |P_D - P_D^*|_2^2 \quad (1)$$

Graph Consistency Regularization: The Graph Consistency Regularization as a loss measures the consistency of gradient directions between the predicted graph (D) and the resting pose graph (R). Inspired by the graph Laplacian, which captures the intrinsic geometric structure of data, this term ensures that predicted deformations respect the underlying spatial relationships, mirroring physical constraints. The loss is computed as follows:

$$L_G = \frac{1}{N_{\text{edges}}} \sum_{i=1}^{N_{\text{edges}}} |\nabla(R) - \nabla(D)|_2 \quad (2)$$

Where:

N_{edges} is the total number of edges in the graph. $\nabla(R)$ represents the gradients of the resting graph (R) positions. $\nabla(D)$ represents the gradients of the predicted graph (D) positions. Graph Consistency Regularization is essential in preserving the object’s structural integrity during deformation, ensuring that the model predictions adhere to realistic physical transformations.

Finally, the total loss, L_T , is computed as the weighted sum of the loss terms:

$$L_T = L_{mse} + \lambda_G \cdot L_G \quad (3)$$

The inclusion of λ_G allows for the fine-tuning of the regularization effect, providing a mechanism to balance the influence of accurate position prediction with the preservation of structural consistency.

By tuning λ_G , we calibrate the model’s focus between precision in spatial deformation prediction and adherence to realistic deformation behaviors, facilitating control over the model’s generalization capabilities.

IV. SIMULATION AND DATASETS

In this section, we detail our experimental setup and results. We start with an overview of our Simulation environment, designed for mesh integration and soft body simulation. Finally, we provide insights into our implementation details.

A. Simulation System

We select a subset of 8 objects from the Objaverse [38] dataset and employ the Unity Engine [39] for simulation. The soft body is positioned in the center of world coordinates, and Rigid objects are initialized with random translations and positions in space. Subsequently, we set random velocities and forces directed towards the soft body and observe the resulting collisions. For physics and deformation estimation, we utilize a unified particle physics [40] engine¹.

B. Everyday Deform Dataset

Leveraging our simulation system, we generate a dataset comprising eight distinct objects, each with unique deformation and stiffness attributes, as shown in Fig. 4. These objects exhibit varied mesh densities. While some are sourced through 3D scanning, others are modelled using 3D CAD software, leading to diverse mesh topologies. The dataset includes the initial resting pose of each object and its deformation across 1.6k distinct contact scenarios. Each object has a fixed, manually determined stiffness and dampness. For every contact, we store the poses of the interacting entities, force vectors, velocities, and coordinates of contact points. In total, our dataset features around 11k meshes. We divided the dataset into an 80% training split and a 20% test split.

¹<http://obi.virtualmethodstudio.com/>

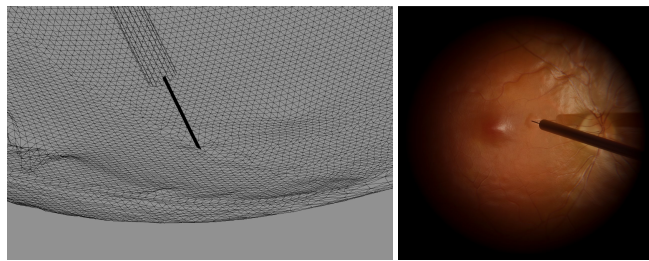


Fig. 3. Images from the SynthesEyes simulation framework: On the left, the 3D mesh depicts the retina’s deformation, while the right presents a 2D camera projection showcasing the intricate deformation patterns of the retina. We utilize and simulate these deformations for our training.

C. Retina Dataset

Diving deeper into specialized applications, we create a dataset on the intricate task of reconstructing the retina surface mesh. Here, our primary objective is to model the deformation pattern of the retina’s surface mesh upon its interaction with a surgical needle. Like the Everyday Deform dataset, we initiated our process by randomly positioning a needle around the retina surface. Here, we use custom software for vitreoretinal surgery simulation². We observed and recorded the subsequent deformations by applying varying degrees of force. This allowed us to simulate the complex physics involved, considering the known characteristics of the retina’s delicate structure. This dataset consists of 1.5k meshes in different deformations upon contact. Fig. 3 shows a view of the surgical tool interaction and rendering in the simulation software.

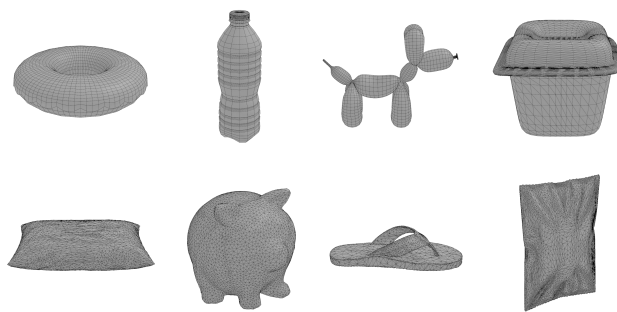


Fig. 4. The ‘Everyday Deform’ dataset features eight commonly deformable objects, each with distinct mesh topologies, densities, and deformation characteristics. Using a simulator, each object undergoes random collisions with rigid bodies. The resultant mesh and associated physics data are recorded and made available to the community.

D. Implementation Details

We’ve constructed our mesh and deformation simulations within the Unity Engine [39] framework. Post simulation, for the processing and learning phases, we integrated the

²<http://syntheseyes.de>

PyTorch library for deep learning operations. For operations specific to graphs and mesh processing, we used Open3D and PyTorch Geometric libraries.

Hardware-wise, our system is powered by a Nvidia Titan RTX GPU and an Intel(R) Core(TM) i7-7820X CPU. For the training phase, we set our batch size at 4. The meshes we processed ranged from 2-6k vertices, dynamic with the object in the pipeline. Notably, we process the full meshes, avoiding sampling techniques. For the retina dataset, our graphs incorporated a detailed 64-node structure to reflect the granularity of the deformation we aimed to capture. Our input features to GNNs are of size 21 and 25 for soft and rigid graphs, and the intermediate encoded feature size is 256.

V. RESULTS AND DISCUSSIONS

To assess the accuracy and efficacy of our proposed method, we adopt specific evaluation metrics and conduct a series of experiments. Our primary evaluation metric is MSE error. MSE calculates the average squared Euclidean distance between the nodes of the predicted deformation and the corresponding nodes in the ground truth. Moreover, we calculate the Mean Absolute Error (MAE) as the average absolute difference between the predicted deformation and the ground truth. Finally, Graph Consistency evaluates the consistency and smoothness of the predicted deformation relative to the ground truth.

A. Quantitative Evaluation

Experiment 1: Object Contact. In our initial experiment, we assess the efficacy of our methodology on the Everyday Deform dataset. Our evaluation strategy is twofold. First, as detailed in Table I, we train a single network per object for 40 epochs. These objects, characterized by diverse densities and physical characteristics, are effectively learned by our model to estimate their full-body deformation. The Dog demonstrates optimal shape consistency while achieving the lowest MSE error. On the other hand, despite the Bottle’s high deformability, it performs the best in terms of MAE error.

Object	Size	Consistency	MAE	MSE
Bottle	3927	0.0015	3.02e-05	3.74e-05
Cat	5002	0.0023	0.0001	0.0002
Dog	4612	0.0010	2.92e-05	1.87e-05
Donut	2560	0.0012	7.38e-06	1.89e-05
Doritos	5002	0.0013	4.27e-05	2.94e-05
Flipflop	4993	0.0009	2.63e-05	1.35e-05
Pillow	4002	0.0019	3.87e-05	6.02e-05
Average	-	0.0013	3.79e-05	2.57e-05

TABLE I

COMPARISON OF DEFORMATION PREDICTION FOR EVERYDAY OBJECTS. EACH NETWORK IS TRAINED SEPARATELY ON THE FULL MESH.

In another experiment, we evaluated our model’s performance, trained using the complete training set on the closest mesh patch comprising 1024 vertices. The results of this evaluation are outlined in Table II. Even though

this dataset encompasses a spectrum of physical attributes and topologies, our rich representation facilitates consistent performance on unseen deformation instances.

Object	Consistency	MAE	MSE
Bottle	0.0045	6.11e-05	0.000143
Box	0.0034	5.75e-05	0.000105
Cat	0.0050	0.000111	0.000312
Dog	0.0013	7.02e-05	2.99e-05
Donut	0.0030	1.89e-05	7.44e-05
Doritos	0.0027	8.71e-05	7.75e-05
Flipflop	0.0019	6.31e-05	6.44e-05
Pillow	0.0037	7.74e-05	0.000158
Average	0.0032	6.69e-05	0.000119

TABLE II

COMPARISON OF DEFORMATION PREDICTION ERROR FOR RANDOM LOCAL OBJECTS. A SINGLE NETWORK IS TRAINED FOR ALL PATCHES.

Experiment 2: Retina Tool Interaction. Our second experiment focuses on the challenging task of retina surface mesh reconstruction. Specifically, we aim to predict the deformation of the retina surface mesh when it comes into contact with a surgical needle.

	MAE (μm)	MSE (μm)	Variance (μm)
Retina	57.326	10.88	0.00010348

TABLE III

ERROR METRICS FOR RETINA TOOL DEFORMATION ESTIMATION, IN MICROMETERS.

Table III summarizes this experiment’s results, demonstrating our methodology’s capability to accurately reconstruct the retina’s local surface under varying force conditions from surgical tools.

B. Qualitative Evaluation

Visualization results in Figures 5 and 6 present an intuitive perspective into the model’s predictions. Our model can clearly predict fine details, be it in terms of surface deformations or interactions with surrounding environments. The comparative visualizations, especially between predicted and ground truth deformations, highlight the minimal discrepancy, reaffirming the quantitative findings.

C. Ablation Studies

To better understand the intricacies of our model and the influence of distinct encodings, we executed ablation studies, the results of which are presented in Table IV. By selectively activating the positional and physics encodings — individually and in combination — we could assess their individual and combined effects on prediction accuracy. The results after ten epochs indicate that the highest accuracy in our node’s initial state is achieved when both geometric and physics information are simultaneously encoded.

D. Runtime

Practical applications require not just accuracy but also efficiency. Our runtime evaluations, presented in Table V,

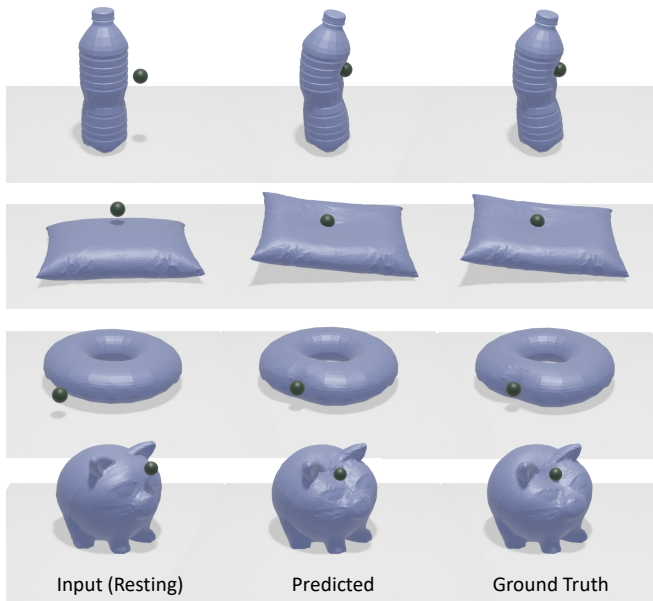


Fig. 5. The figure presents our quantitative results calculated from the Everyday Deform dataset. The resting pose is visualized on the right, with the rigid object in its initial position. At the center, we illustrate the deformation prediction from our network. The left images depict the ground truth deformation present in the dataset. Our method captures fine surface deformation rigid motion and even reflects interaction with the ground floor.

	Positional	Physics	MAE	MSE
Only Physics		✓	1.05e-05	3.119e-04
Only Positional	✓		1.14e-05	3.389e-04
Both	✓	✓	0.90e-05	2.695e-04

TABLE IV

ABLATION STUDY ON DIFFERENT ENCODINGS USING MAE, MSE, WITH ACTIVE ENCODINGS INDICATED BY A CHECKMARK. THE STUDY SHOWS BOTH ENCODINGS CONTRIBUTE TO THE FINAL PREDICTIONS.

suggest that our method is computationally efficient, rendering it suitable for real-time applications. The table shows latency for graph and data preparation and network inference pass. This is particularly crucial for surgical systems and simulations where timely and quick predictions are necessary.

Dataset	Size	Data (ms)	Network (ms)
Retina	64	0.992	2.978
Everyday Deform	1024	1.661	4.261

TABLE V

RUNTIME EVALUATION PER BATCH (OF 4 SAMPLES) ON DIFFERENT DATASETS, SPLIT INTO DATA LOADING AND GRAPH GENERATION, AND NETWORK PASS AND ESTIMATION.

E. Future Works

We plan to tackle a few key areas to improve our method in our upcoming work. One idea is to use time series analysis. Our method currently looks at deformation at a single point in time, but with time series, we could track how objects change over a more extended period. This could

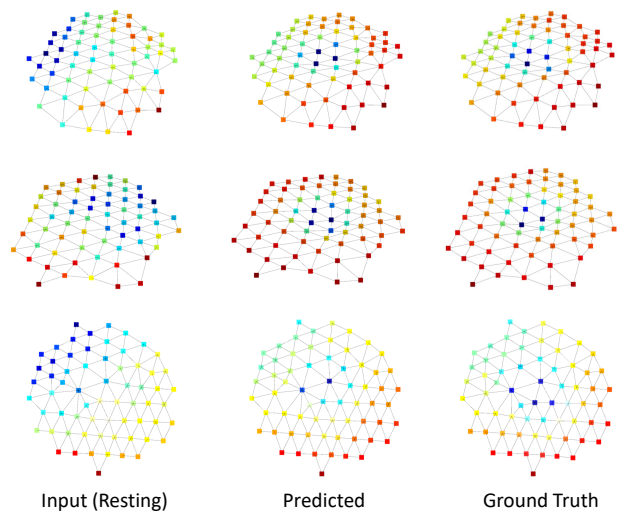


Fig. 6. The figure provides a visualization of the retina mesh graph: the input is displayed on the left, the predicted deformations in the middle, and the ground truth on the right. Predicted vertex normal vectors are depicted, using color mappings to highlight the curvature.

be useful in tasks like robotic planning. Another thing we're excited about is using Neural Ordinary Differential Equations (Neural ODEs). These tools are great at showing continuous changes, which could give a smoother picture of how things like retina tools or soft objects interact.

Additionally, our current dataset doesn't account for different deformable parts of an object. One can pay more attention to this in the future so the simulations are closer to true-to-life objects, especially the ones with mixed materials. And finally, instead of the focus on soft-to-rigid interaction, one can build on top of our pipeline to learn how soft objects interact with each other. These updates will make our contribution even better and more in tune with real-world situations.

VI. CONCLUSION

This work bridges the gap between computational geometry and Physics-Informed Deep Learning, utilizing a data-driven approach to model object interactions. We process mesh graphs for soft and rigid shapes, integrating physical and geometric data. By leveraging cross-attention mechanisms, we capture fine-grained physical dynamics. Subsequently, we reconstruct post-contact deformations under various conditions. Supplemented by a new dataset featuring diverse everyday objects, our approach presents a practical and efficient solution for simulated robotics, grasping, surgical tool interactions, and other contact modeling and robotic manipulation applications.

VII. ACKNOWLEDGMENT

The authors wish to thank SynthesEyes for providing the excellent simulation setup for the data generation.

REFERENCES

- [1] S. Dehghani, M. Sommersperger, P. Zhang, A. Martin-Gomez, B. Busam, P. Gehlbach, N. Navab, M. A. Nasser, and I. Iordachita, "Robotic navigation autonomy for subretinal injection via intelligent real-time virtual ioc volume slicing," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 4724–4731.
- [2] R. W. Clough and R. J. Woodward III, "Analysis of embankment stresses and deformations," *Journal of the Soil Mechanics and Foundations Division*, vol. 93, no. 4, pp. 529–549, 1967.
- [3] A. M. Group, J. Merwin, and K. Johnson, "An analysis of plastic deformation in rolling contact," *Proceedings of the Institution of Mechanical Engineers*, vol. 177, no. 1, pp. 676–690, 1963.
- [4] W. S. Hall and W. Hall, *Boundary element method*. Springer, 1994.
- [5] B. Xue, X. Du, J. Wang, and X. Yu, "A scaled boundary finite-element method with b-differentiable equations for 3d frictional contact problems," *Fractal and Fractional*, vol. 6, no. 3, p. 133, 2022.
- [6] A. Popp, M. Gitterle, M. W. Gee, and W. A. Wall, "A dual mortar approach for 3d finite deformation contact with consistent linearization," *International Journal for Numerical Methods in Engineering*, vol. 83, no. 11, pp. 1428–1465, 2010.
- [7] A. Popp, B. I. Wohlmuth, M. W. Gee, and W. A. Wall, "Dual quadratic mortar finite element methods for 3d finite deformation contact," *SIAM Journal on Scientific Computing*, vol. 34, no. 4, pp. B421–B446, 2012.
- [8] R. A. Gingold and J. J. Monaghan, "Smoothed particle hydrodynamics: theory and application to non-spherical stars," *Monthly notices of the royal astronomical society*, vol. 181, no. 3, pp. 375–389, 1977.
- [9] M. Liu and G. Liu, "Smoothed particle hydrodynamics (sph): an overview and recent developments," *Archives of computational methods in engineering*, vol. 17, pp. 25–76, 2010.
- [10] X. Provot *et al.*, "Deformation constraints in a mass-spring model to describe rigid cloth behaviour," in *Graphics interface*. Canadian Information Processing Society, 1995, pp. 147–147.
- [11] M. Saleh, S.-C. Wu, L. Cosmo, N. Navab, B. Busam, and F. Tombari, "Bending graphs: Hierarchical shape matching using gated optimal transport," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11 757–11 767.
- [12] M. Saleh, S. Dehghani, B. Busam, N. Navab, and F. Tombari, "Graphite: Graph-induced feature extraction for point cloud registration," in *2020 International Conference on 3D Vision (3DV)*. IEEE, 2020, pp. 241–251.
- [13] A. Amor-Martin and L. E. Garcia-Castillo, "Adaptive semi-structured mesh refinement techniques for the finite element method," *Applied Sciences*, vol. 11, no. 8, p. 3683, 2021.
- [14] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.
- [15] Z. Hao, S. Liu, Y. Zhang, C. Ying, Y. Feng, H. Su, and J. Zhu, "Physics-informed machine learning: A survey on problems, methods and applications," *arXiv preprint arXiv:2211.08064*, 2022.
- [16] M. B. Chang, T. Ullman, A. Torralba, and J. B. Tenenbaum, "A compositional object-based approach to learning physical dynamics," *arXiv preprint arXiv:1612.00341*, 2016.
- [17] Y. W. Bekele, "Physics-informed deep learning for flow and deformation in poroelastic media," *arXiv preprint arXiv:2010.15426*, 2020.
- [18] D. Holden, B. C. Duong, S. Datta, and D. Nowrouzezahrai, "Subspace neural physics: Fast data-driven interactive simulation," in *Proceedings of the 18th annual ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 2019, pp. 1–12.
- [19] L. Fulton, V. Modi, D. Duvenaud, D. I. Levin, and A. Jacobson, "Latent-space dynamics for reduced deformable simulation," in *Computer graphics forum*, vol. 38, no. 2. Wiley Online Library, 2019, pp. 379–391.
- [20] T. Okazaki, T. Ito, K. Hirahara, and N. Ueda, "Physics-informed deep learning approach for modeling crustal deformation," *Nature Communications*, vol. 13, no. 1, p. 7092, 2022.
- [21] T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. W. Battaglia, "Learning mesh-based simulation with graph networks," *arXiv preprint arXiv:2010.03409*, 2020.
- [22] M. Maurizi, C. Gao, and F. Berto, "Predicting stress, strain and deformation fields in materials and structures with graph neural networks," *Scientific Reports*, vol. 12, no. 1, p. 21834, 2022.
- [23] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.
- [24] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.
- [25] H. Zhao, L. Jiang, J. Jia, P. H. Torr, and V. Koltun, "Point transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 16 259–16 268.
- [26] M. Saleh, Y. Wang, N. Navab, B. Busam, and F. Tombari, "Cloudattention: Efficient multi-scale attention scheme for 3d point cloud learning," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 1986–1992.
- [27] L. Yi, H. Su, X. Guo, and L. J. Guibas, "Syncspecnn: Synchronized spectral cnn for 3d shape segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2282–2290.
- [28] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *ACM Transactions on Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.
- [29] H. Yu, Z. Qin, J. Hou, M. Saleh, D. Li, B. Busam, and S. Ilic, "Rotation-invariant transformer for point cloud matching," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5384–5393.
- [30] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.
- [31] V. E. Arriola-Rios, P. Guler, F. Ficuciello, D. Kragic, B. Siciliano, and J. L. Wyatt, "Modeling of deformable objects for robotic manipulation: A tutorial and review," *Frontiers in Robotics and AI*, vol. 7, p. 82, 2020.
- [32] C. Romero, D. Casas, J. Perez, and M. Otaduy, "Learning contact corrections for handle-based subspace dynamics," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, pp. 1–12, 2021.
- [33] C. Romero, D. Casas, M. M. Chiaramonte, and M. A. Otaduy, "Contact-centric deformation learning," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, pp. 1–11, 2022.
- [34] Z. Leng, S.-C. Wu, M. Saleh, A. Montanaro, H. Yu, Y. Wang, N. Navab, X. Liang, and F. Tombari, "Dynamic hyperbolic attention network for fine hand-object reconstruction," 2023.
- [35] Y. Salehi and D. Giannacopoulos, "Physgcn: A physics-driven graph neural network based model for predicting soft tissue deformation in image-guided neurosurgery," *Advances in Neural Information Processing Systems*, vol. 35, pp. 37 282–37 296, 2022.
- [36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [37] J. Du, S. Zhang, G. Wu, J. M. Moura, and S. Kar, "Topology adaptive graph convolutional networks," *arXiv preprint arXiv:1710.10370*, 2017.
- [38] M. Deitke, D. Schwenk, J. Salvador, L. Weihs, O. Michel, E. Vanderbilt, L. Schmidt, K. Ehsani, A. Kembhavi, and A. Farhadi, "Objaverse: A universe of annotated 3d objects," *arXiv preprint arXiv:2212.08051*, 2022.
- [39] U. Technologies, "Unity, version 2023.1," <https://unity.com/>, 2023, accessed: 2023-09.
- [40] M. Müller, B. Heidelberger, M. Hennix, and J. Ratcliff, "Position based dynamics," *Journal of Visual Communication and Image Representation*, vol. 18, no. 2, pp. 109–118, 2007.