Learning Force Distribution Estimation for the GelSight Mini Optical Tactile Sensor Based on Finite Element Analysis

Erik Helmut[*1], **Luca Dziarski**[*1], **Niklas Funk**[1], **Boris Belousov**[3], **Jan Peters**[1,2,3]

[1]Technical University of Darmstadt [2]German Research Center for AI (DFKI)
[3] Hessian Center for Artificial Intelligence (Hessian.AI), Darmstadt

Abstract

Contact-rich manipulation remains a major challenge in robotics. Optical tactile sensors like GelSight Mini offer a low-cost solution for contact sensing by capturing soft-body deformations of the silicone gel. However, accurately inferring shear and normal force distributions from these deformations has yet to be fully addressed. In this work, we propose a machine learning approach using a U-net architecture to predict force distributions directly from raw images. Our model, trained on force distributions inferred from Finite Element Analysis (FEA), demonstrates promising accuracy in predicting these distributions. It also shows potential for generalizing to a different GelSight sensor than the one used for data collection and enables real-time application. This approach holds the promise of improving tactile perception in contact-rich robotic manipulation tasks.

1 Introduction

The integration of tactile sensing into robotics is crucial for enhancing functionality across various applications, from object manipulation to medical tasks and tele-operation [1, 4]. Optical tactile sensors like the GelSight Mini [9] capture detailed contact information through image processing. These sensors offer high spatial resolution and cost-effectiveness, making them attractive for bringing tactile sensing to robotics.

However, most research has focused on low-dimensional force measurements rather than comprehensive contact force distributions, which include both total force and contact area [10]. Estimating accurate force distributions is challenging due to the non-linear behavior of contact mediums. While traditional methods like FEA offer detailed solutions, they are computationally demanding. Although inverse FEA has been applied in real-time scenarios [6], it remains resource-intensive and is generally not well-suited for most real-time applications.

Recent advancements leverage deep learning to address the challenge of real-time force estimation. For example, Yuan et al. [9] used Convolutional Neural Network (CNN)'s to predict contact forces from sensor images, while Funk et al. [3] introduced CANFnet for estimating normal force distributions at the pixel level. Sferrazza and D'Andrea [8] employed FEA-derived data to train deep learning models for predicting force distributions, demonstrating the effectiveness of combining simulations with data-driven methods.

This study builds on these advancements, particularly on the approach in [8], by employing a U-net architecture [7] to predict force distributions from raw images, using labels generated through FEA. By incorporating a diverse set of indenters and applied forces, we aim to enhance the model's

[*]Equal contribution. Email: `erik.helmut1@gmail.com`. Website: `https://feats-ai.github.io`

NeurIPS 2024 Workshop on Touch Processing: a new Sensing Modality for AI .

generalization across various contact scenarios. This approach not only provides an interpretable metric for the GelSight Mini but also delivers a real-time solution applicable to downstream tasks.

2 Method

Recovering the force distributions acting onto the GelSight Mini sensor provides an interpretable representation of the RGB tactile images. While FEA provides physically grounded force distributions, it is computationally expensive and requires precise contact information which is often unavailable in real-world scenarios. To overcome these drawbacks, we propose using a supervised machine learning approach. By training a U-net architecture on a dataset of raw images and corresponding FEA-derived force distributions obtained in a calibrated setup, the model learns to predict force distributions directly from images. The following subsections detail the dataset generation and the neural network-based method.

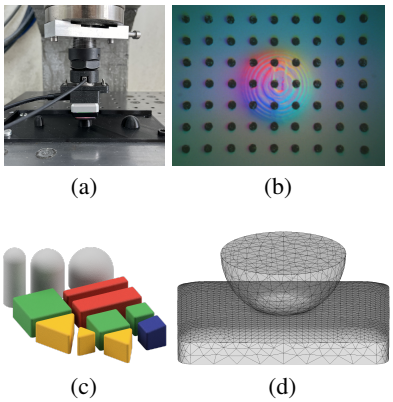

Figure 1: Overview of dataset generation for ground truth force distributions: (a) Computer Numerical Control (CNC) milling machine pressing the GelSight Mini against a 3D-printed sphere indenter; (b) Raw image from the GelSight Mini; (c) Various indenters used in training; (d) Simulation model from FEA.

2.1 Data Collection

The data collection process involves a series of indentation experiments conducted with the GelSight Mini, equipped with a dotted gel to better track the indentation motion (see Fig. 1b). Inspired by Sferrazza and D'Andrea [8], we use 12 indenters of varying shapes and sizes (see Fig. 1c). A CNC milling machine, with a positional tolerance of $\pm 0.25\,\mu\mathrm{m}$, is utilized to automate the data collection process and to ensure precise indentation movements. Furthermore, a six-axis *RESENSE-HEX-21* Force/Torque (F/T) sensor is positioned above the Gelsight Mini to record contact forces during indentation (see Fig. 1a). A total of 5173 samples were collected, combining GelSight Mini images, F/T readings, and CNC motion data. The force in the z direction reach up to $40\,\mathrm{N}$, and in the x and y directions up to $\pm 5\,\mathrm{N}$.

2.2 Finite Element Analysis

For simulating indentation experiments, we employed FEA using the open-source solver CalculiX [2], which is capable of nonlinear analysis. Tetrahedral volume meshes for both the gel and indenters were generated, with denser meshing at the contact surface to ensure higher accuracy of the simulated contact forces (see Fig. 1d). The simulation is set up as a static analysis, assuming hard contact and no friction, as slippage is avoided in the experiments. Indenters were modeled as steel, with a Young's modulus of $210\,\mathrm{GPa}$ and a Poisson coefficient of 0.3. A hyperelastic Neo-Hookean model, as detailed in [9], is applied for the gel material. The shear modulus μ is set to 0.145, corresponding to a material constant C_{10} of 0.0725. Incompressibility is neglected by assigning a default value to D_1 due to solver limitations. This model is validated by optimizing the material parameters through load-depth indentation data, comparing the normal forces from the FEA with F/T sensor measurements (see Sec. 3.1). The results closely matched the literature values reported in [9].

2.3 Creating Labels

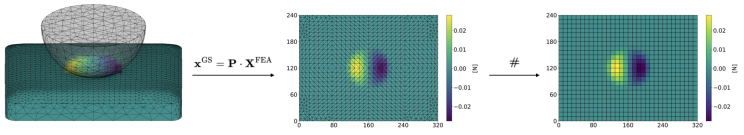

Figure 2: Label creation process shown as an example for the shear force distribution in x direction with a sphere indenter.

To generate the ground truth force distributions, each indentation performed by the CNC milling machine is simulated using FEA, which calculates the contact forces across the surface of the GelSight Mini's gel. For each element on the surface of the gel, three force components are computed: shear forces in the x and y directions, and normal forces in the z direction. To align these

FEA results with the sensor's raw images, the 3D coordinates of the gel mesh nodes are initially projected onto the 2D image plane of the GelSight Mini using a precomputed projection matrix (see Fig. 2). Finally, the force distributions are discretized into grid cells within the image boundaries, summing the force contributions of all elements intersecting one cell.

2.4 U-Net for Force Distribution Estimation

To predict shear and normal force distributions from the raw GelSight images, we employ a U-net architecture [7], which is well-suited for spatially-detailed tasks. The network's encoder-decoder structure allows for efficiently mapping raw sensor images to force distributions, as demonstrated by Sferrazza and D'Andrea [8]. The model takes 240×320 RGB images as input and outputs force maps for both shear and normal forces at a reduced resolution of 24×32 (see Fig. 3).

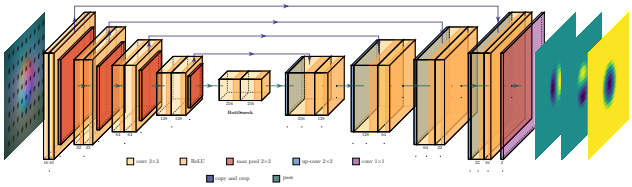

Figure 3: U-net model mapping raw GelSight images to shear and normal force distributions. The architecture comprises 4 down- and 4 up-sampling blocks, connected by skip connections. Image generated using PlotNeuralNet [5].

The model is trained by minimizing the Mean Squared Error (MSE) loss between predicted and ground truth FEA forces. The dataset is split into three parts: 85% for training, 5% for validation, and 10% for testing, with normalization applied to both input images and labels. Shear forces are normalized to a range of -1 to $+1$, while normal forces range from 0 to $+1$. Data augmentation techniques like Gaussian noise and brightness, contrast, saturation, and hue adjustments are used to enhance generalization. Training is carried out using the Adam optimizer, with an initial learning rate of 0.001 and a batch size of 8. The learning rate is adjusted adaptively based on validation performance. We train for 600 epochs and select the model with the lowest validation loss.

3 Experimental Results

This section evaluates the U-net's performance in predicting force distributions. First, the material constants for the hyperelastic model used in label generation are validated. Then, the U-net's accuracy in predicting shear and normal force distributions, along with its ability to reconstruct total forces, is analyzed. Particular attention is given to how different force distribution resolutions impact total force reconstruction accuracy. Finally, the inference speed of the U-net is assessed.

3.1 Material Characterization

To validate the Neo-Hookean material model parameter C_{10}, a load-depth curve must be defined through indentation. The load-depth curve is obtained, using a sphere indenter with a $15\,\mathrm{mm}$ diameter, by sampling data points at different indentation depths ranging from 0.5 mm to 2.0 mm, increasing in 0.5 mm intervals. The measured normal forces from the F/T sensor are compared with those from the FEA. Bayesian optimization revealed that $C_{10} = 0.0792$ provided the best fit, with a mean absolute error (MAE) of $0.5121\,\mathrm{N}$ between FEA and F/T sensor measurements. Although this is close to the literature value of 0.0725 from [9], the literature value was chosen for ongoing simulations due to potential measurement uncertainties.

3.2 Force Distribution Estimation

To evaluate the U-net's ability to predict force distributions, we tested it on 492 samples recorded using the same sensor as in training dataset. The test data captures a wide range of normal forces and smaller shear forces, due to the nature of the indentation experiments.

Baseline U-net Model Performance. The baseline U-net model predicts force distributions of size $24 \times 32 \times 3$, outputting two shear and one normal force distributions simultaneously. Figure 4 illustrates an example prediction, which closely aligns with the ground truth, highlighting the model's effectiveness (also see Fig. A.I). Quantitative evaluation using MAE shows accurate predictions, with MAE values under $1\,\mathrm{N}$ (see Table 1).

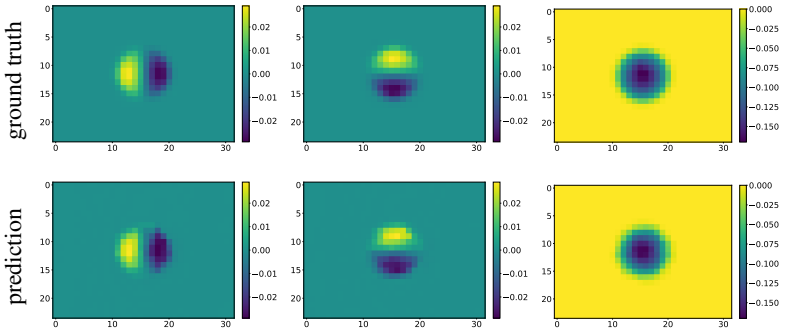

Figure 4: Prediction of baseline U-net and ground truth labels of the force distributions for Fig. 1b. The first row is showing the ground truth force distributions in x-, y- and z-direction. In the second row the corresponding predictions of the U-net are presented. The entities within the force distributions are in N.

The U-net performs better with shear forces than with normal forces, likely due to the smaller force range for shear and the richer feature set provided by marker displacement in shear force estimation. In contrast, normal force estimation relies more on depth information, which is harder to extract.

Table 1: Comparison of U-net Variants on Total Force Estimation

Method	MAE_{TF} [N]		
	f_x	f_y	f_z
ResNet$^{1\times3}$	0.085 ± 0.115	0.069 ± 0.085	1.593 ± 1.131
U-net$^{12\times16\times3}$	0.102 ± 0.216	0.089 ± 0.123	0.447 ± 0.539
$3\times$U-net$^{24\times32\times1}$	0.438 ± 0.585	0.189 ± 0.225	0.448 ± 0.523
U-net$^{24\times32\times3}$	$\mathbf{0.224 \pm 0.401}$	$\mathbf{0.093 \pm 0.136}$	$\mathbf{0.372 \pm 0.473}$
U-net$^{48\times64\times3}$	0.318 ± 0.436	0.119 ± 0.184	0.459 ± 0.516

Impact of Label Resolution. To assess the impact of label resolution on prediction accuracy, the baseline U-net is compared with other models, including different U-net configurations and a ResNet (see Table 1). The ResNet, which predicts total forces exclusively, excels in shear force prediction but is less accurate with normal forces. A U-net trained on lower-resolution distributions ($12 \times 16 \times 3$) improves shear prediction but slightly reduces normal force accuracy. Training three separate U-nets for each force direction ($24 \times 32 \times 1$) led to worse results, indicating that predicting all forces simultaneously captures inter-force correlations better. A higher-resolution U-net ($48 \times 64 \times 3$), despite its higher dimensional output only yields slightly bigger errors.

Generalization to Unseen Indenters and Sensors. To evaluate the U-net's generalization capability, it is tested on unseen indenters and images from a different GelSight Mini sensor. When tested on images from another sensor, predictions remain relatively accurate, though with a noticeable shift in contact localization, likely due to differences in camera positioning and variations in the illumination (see Fig. A.II). When considering predictions for unknown indenters, the quality of the predictions decreases (see Fig. A.III). However, since other works (e.g. [3]) successfully showed this kind of generalization, we conclude that a more extensive and versatile dataset is required to improve these results. This is subject to future work.

Inference Speed. On average, the U-net achieves an inference time of $4.2 \pm 1.3\,\text{ms}$ on a *NVIDIA Quadro RTX 5000* GPU with Max-Q Design and a *Intel Core i7-10875H 8-Core* $2.3\,\text{GHz}$ CPU, which is sufficient for real-time applications, as the GelSight Mini runs at $25\,\text{Hz}$.

4 Conclusion

This study introduces a machine learning approach for estimating force distributions using the GelSight Mini tactile sensor. By training a U-net model on FEA-derived data, we achieved accurate predictions of both shear and normal forces from raw sensor images, particularly excelling in shear force estimation. The model shows potential for real-time applications and generalization to different sensors, despite shifts in contact localization. This method offers an efficient alternative to traditional approaches like inverse FEA, with potential for improving tactile perception in robotic manipulation tasks. Future work will focus on enhancing the model's robustness to different tactile scenarios and refining its generalization capabilities across a broader range of objects and sensors.

Acknowledgment

We thank LAB^3, Kai Ruf & Uwe Faltermeier for their great support and access to the CNC milling machine. This research was supported by Research Clusters "The Adaptive Mind" and "Third Wave of AI", funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and the Arts. We acknowledge the grant "Einrichtung eines Labors des Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) an der Technischen Universität Darmstadt" of the Hessisches Ministerium für Wissenschaft und Kunst. The authors gratefully acknowledge the computing time provided to them on the high-performance computer Lichtenberg II at TU Darmstadt, funded by the German Federal Ministry of Education and Research (BMBF) and the State of Hesse. This work received funding from the EU's Horizon Europe project ARISE (Grant no.: 101135959) and through the AICO grant by the Nexplore/Hochtief Collaboration with TU Darmstadt.

References

[1] A. Billard and D. Kragic. Trends and challenges in robot manipulation. *Science*, 364(6446):eaat8414, 2019.

[2] G. Dhondt. Calculix crunchix user's manual version 2.21. `https://www.dhondt.de/ccx_2.21.pdf`, 2023. Accessed: 2024-03-05.

[3] N. W. Funk, P.-O. Müller, B. Belousov, A. Savchenko, R. Findeisen, and J. Peters. CANFnet: High-Resolution Pixelwise Contact Area and Normal Force Estimation for Visuotactile Sensors Using Neural Networks. 2023.

[4] B. Gates. A Robot in Every Home. *Scientific American*, 296(1):58–65, 2007. ISSN 0036-8733.

[5] H. Iqbal. Harisiqbal88/plotneuralnet v1.0.0, Dec. 2018. URL `https://doi.org/10.5281/zenodo.2526396`.

[6] D. Ma, E. Donlon, S. Dong, and A. Rodriguez. Dense Tactile Force Estimation using GelSlim and inverse FEM. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 5418–5424, May 2019. doi: 10.1109/ICRA.2019.8794113. ISSN: 2577-087X.

[7] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation.

[8] C. Sferrazza and R. D'Andrea. Sim-to-real for high-resolution optical tactile sensing: From images to three-dimensional contact force distributions. *Soft Robotics*, 9(5):926–937, 2022. doi: 10.1089/soro.2020.0213. URL `https://doi.org/10.1089/soro.2020.0213`. PMID: 34842455.

[9] W. Yuan, S. Dong, and E. Adelson. GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force. *Sensors*, 17(12):2762, Nov. 2017. ISSN 1424-8220. doi: 10.3390/s17122762.

[10] C. Zhang, S. Cui, Y. Cai, J. Hu, R. Wang, and S. Wang. Learning-based six-axis force/torque estimation using gelstereo fingertip visuotactile sensing. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3651–3658, 2022. doi: 10.1109/IROS47612.2022.9981100.

Appendix

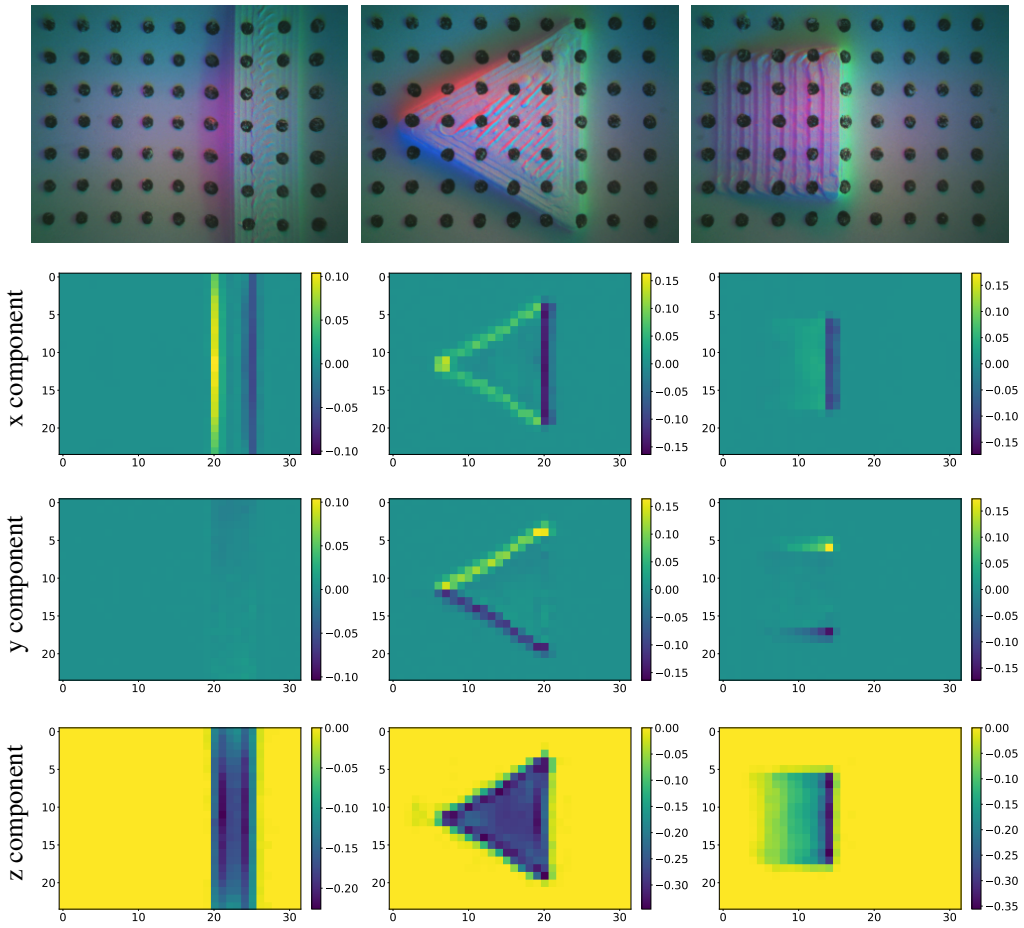

Figure A.I: Example predictions of the baseline U-net on test data using indenters that are included in the training dataset. The first row displays the raw images captured by the GelSight Mini sensor, using the same sensor employed to generate the training dataset. These images are taken as input for the U-net. In the following rows, the corresponding predictions of the U-net for both the shear forces in x and y direction and the normal force in z direction are presented. The entities within the force distributions are in N.

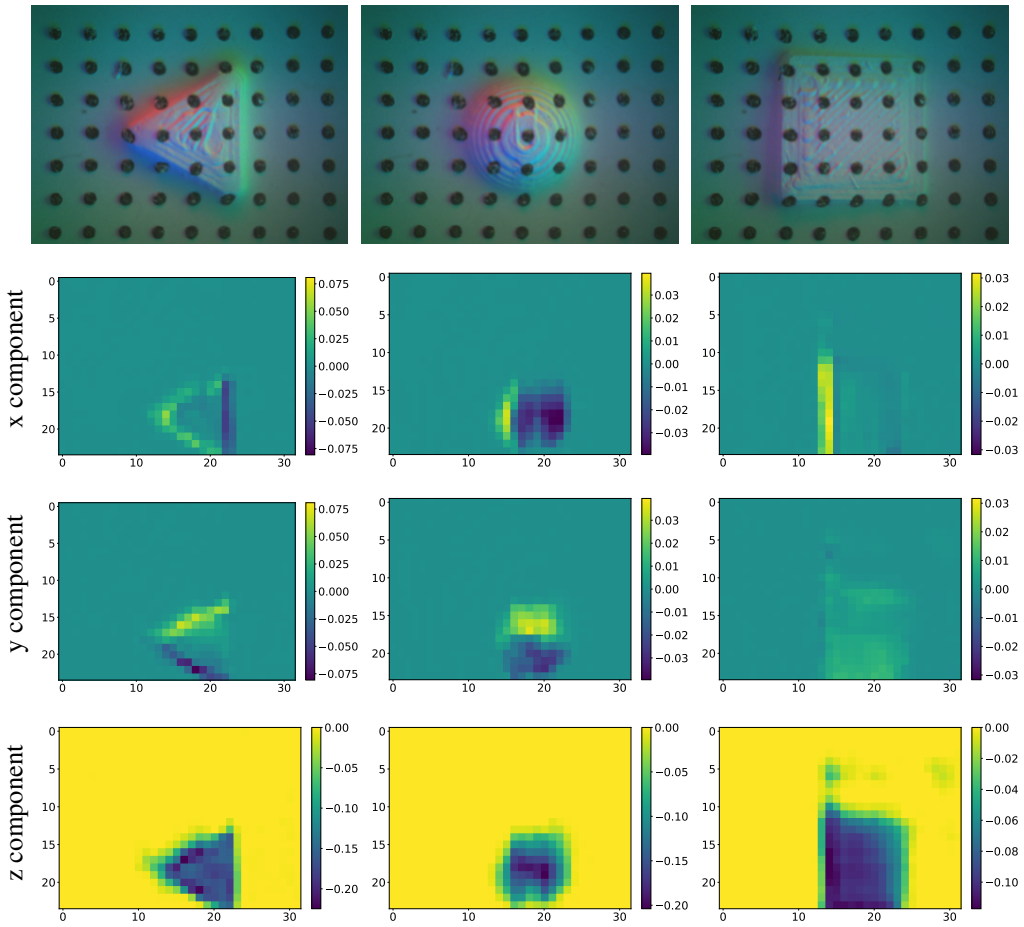

Figure A.II: Example predictions of the baseline U-net on test data using indenters that are included in the training dataset. The first row displays the raw images captured by the GelSight Mini sensor, using a different sensor that is not employed to generate the training dataset. These images are taken as input for the U-net. In the following rows, the corresponding predictions of the U-net for both the shear forces in x and y direction and the normal force in z direction are presented. The entities within the force distributions are in N.

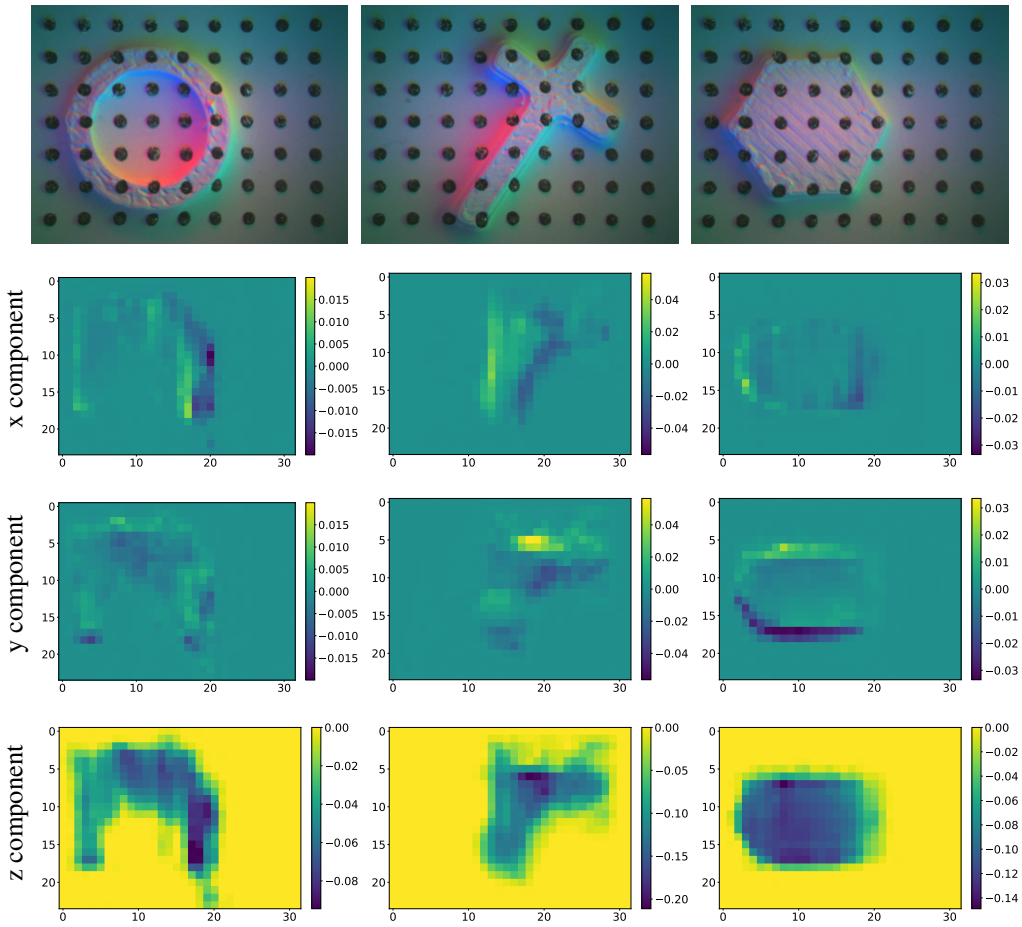

Figure A.III: Example predictions of the baseline U-net on test data using indenters that are not included in the training dataset. The first row displays the raw images captured by the GelSight Mini sensor, using the same sensor employed to generate the training dataset. These images are taken as input for the U-net. In the following rows, the corresponding predictions of the U-net for both the shear forces in x and y direction and the normal force in z direction are presented. The entities within the force distributions are in N.