

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

Erik Helmut^{*1}, Luca Dziarski^{*2}, Niklas Funk², Boris Belousov³, Jan Peters^{2,3,4,5}

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 gel 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 the sensor’s raw images. Our model, trained on force distributions inferred from Finite Element Analysis (FEA), demonstrates promising accuracy in predicting normal and shear force distributions. It also shows potential for generalization across sensors of the same type and for enabling real-time application. The codebase, dataset and models are open-sourced and available at <https://feats-ai.github.io>.

I. INTRODUCTION

Tactile sensing plays an important role in advancing the state-of-the-art in robotic manipulation [1], [2], [3], [4], [5], [6], [7], [8]. Successful applications include grip adaptation through slip detection [9], [10], [11], medical procedures [12], [13] and tele-operation [14].

In particular, optical tactile sensors have emerged as a promising technology for capturing contact information due to their high spatial resolution, multimodal sensing capabilities—including shape [15], hardness [16], texture [17], and temperature [18]—and cost-effectiveness [19], [13]. However, many prior works have focused on extracting only low-dimensional tactile information, such as total force [20], [21], [22], limiting operational flexibility. Access to contact force distributions, on the other hand, would enable better handling of multiple contacts and diverse manipulation scenarios [23].

Conventional methods for extracting force distributions require calculating the three-dimensional deformation of the contact medium and utilizing elasticity theory [24], [25], [26], [27]. Yet, accounting for non-linear material behavior, such as with Finite Element Analysis (FEA), is computationally intensive and unsuitable for real-time applications.

^{*}Authors contributed equally.

Corresponding author: Erik Helmut. Email: erik.helmut1@gmail.com.

¹Department of Computational Engineering, Technical University of Darmstadt ²Department of Computer Science, Technical University of Darmstadt ³German Research Center for AI (DFKI) ⁴Centre for Cognitive Science, Technical University of Darmstadt ⁵Hessian Center for Artificial Intelligence (Hessian.AI), Darmstadt

This work received funding from “The Adaptive Mind” grant, the EU’s Horizon Europe project ARISE (Grant no.: 101135959), the AICO grant by the Nexlore/Hochtief Collaboration with TU Darmstadt, and from the Hessisches Ministerium für Wissenschaft & Kunst through the DFKI grant.

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

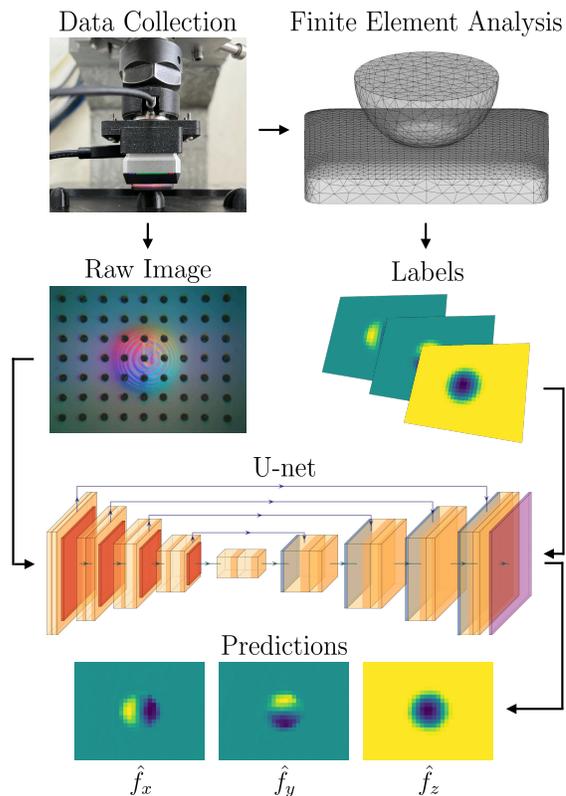


Fig. 1: Complete Method Overview: from data collection to force distribution prediction. After data collection in a precisely calibrated setup with a CNC milling machine, Finite Element Analysis is employed to generate labels (“ground truth” force distributions). Using the labels and raw images captured by the GelSight Mini tactile sensor, we train a U-net for efficiently mapping raw tactile images to the corresponding force distributions.

Recent works leverage Deep Learning to address the challenge of real-time force estimation. In [28], Convolutional Neural Networks (CNNs) were used to predict contact forces from sensor images, while [29] introduced CANFnet for estimating normal force distributions at the pixel level. In [23], [30], [31], FEA-derived data was used to train a model for predicting force distributions, demonstrating the effectiveness of combining simulations with data-driven methods.

In this paper we introduce FEATS (see Fig. 1)—a machine learning approach that directly maps raw tactile images to force distributions, building upon the method by Sferrazza et al. [31]. We utilize FEA to generate labeled data for training, ensuring accurate ground truth across various indenters and force levels. A U-net neural network architecture [32] is employed to estimate force distributions from images captured by the GelSight Mini optical sensor [28], [33].

In contrast to [31], our method is tailored to a widely available commercial sensor GelSight Mini, dropping the requirement of a custom-made gel with immersed particles, thereby drastically extending the applicability of the approach. Furthermore, this sensor allows for a significantly expanded range of measurable forces 0 – 40 N, an 8-fold increase in the maximum measurable force compared to [31]. Finally, we open source our code, dataset and model, aiding reuse and reproducibility.

Experimental results demonstrate that the proposed method accurately predicts high-dimensional contact force distributions from raw tactile images. This capability advances robotic manipulation by accommodating a wider range of contact scenarios and offers a versatile representation applicable to downstream tasks.

II. RELATED WORK

Extracting meaningful contact-related information from the raw RGB images of optical tactile sensors is a major challenge in visual-tactile perception [14], [24], [29], [31], [34], [35]. A number of methods have been proposed for constructing or learning such “tactile representations”.

A. Marker Displacement Methods

Li et al. [34] posit that it is the contact layer deformations that capture the crucial information within tactile images. By analyzing these deformations, various contact features can be extracted, with Marker Displacement Methods (MDMs) being the most common approach [34]. In MDMs, markers are placed on or within the elastomer and appear as features in the sensor’s imagery (Fig. 1). For the GelSight sensor [28], [33], markers were first introduced in [36] to study normal and shear forces, along with slip dynamics. They identified a linear relationship between loads and marker motion, but this applied only in non-slip conditions. Beyond marker motion, optical sensors can capture detailed height maps and contact geometry through careful illumination and photometric stereo [35]. These height maps can be used to estimate contact forces with a third-degree polynomial [14].

In this paper, we use a gel with markers, but their movement is not explicitly tracked. Instead, they serve as implicit features within the sensor image, which is analyzed by a neural network to predict force distributions.

B. Deep Learning-Based Tactile Representations

Advancements in computer vision directly translate to vision-based tactile sensing. Models such as CNNs and LSTMs were adapted to assess object hardness [16] and grip stability [37], whereas SVMs were used for lump detection [38]. More tactile-specific deep learning methods have been developed for overall force prediction [28] and for pixel-wise contact area and normal force estimation [29].

Building on the demonstrated effectiveness of deep neural networks for feature extraction and prediction, we employ a U-net architecture similar to that of [29]. However, in contrast to [29], FEATS estimates both normal and shear forces, thus providing a physically grounded representation in the form of a 3D force distribution acting upon the sensor.

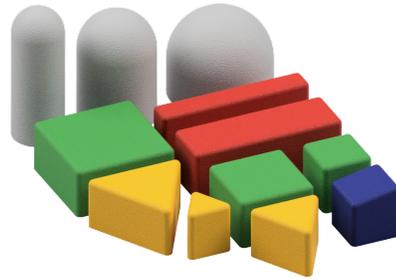


Fig. 2: Models of 3D-printed indenters used for data collection. Different colors represent groups of indenters with similar shapes.

C. Force Distribution Prediction Through Elasticity Theory

Elasticity theory has been effectively applied to create more refined and accurate load distributions acting upon the soft silicone gel of optical tactile sensors. In [24], elasticity theory with MDMs was used to derive force vectors from marker movements assuming a linear elastic, uniform and half-spaced material. This method was later adopted in [25] for the GelSlim sensor [39]. More recently, sensors enabling 3D surface deformation reconstruction have been proposed, such as TacLINK [26] and Tac3D [27]. They compute force distributions from measured 3D marker displacements.

However, direct prediction of force distributions from displacements, usually through a linear stiffness matrix, does not account for the nonlinearities of soft elastomers. Sun et al. [40] addressed this limitation by employing ResNet [41], which was trained on sensor images with approximated force distributions. Similarly, Sferrazza et al. [31] utilized a Deep Neural Network (DNN) trained on image features with force distributions obtained from FEA, and later improved this method by incorporating simulated training data [23], [30].

Building on these approaches, we also aim to estimate contact forces using supervised deep learning. However, instead of working on optical flow features or grayscale images [30] applicable to the custom-made sensor with a dense 3D-marker field [23], [30], [31], we use raw RGB images from a widely available GelSight Mini sensor. Crucially, we develop specific procedures for data collection and model training that enable the efficient use of this widely accessible sensor, thereby significantly lowering the entry barrier into the field.

Our key contributions are: i) method for creating force labels from FEA outputs tailored to GelSight Mini + implementation in CalculiX, ii) data collection procedure + dataset, iii) trained model applicable to varying objects and gels.

III. METHOD

Our proposed method—Finite Element Analysis for Tactile Sensing (FEATS)—estimates the force distribution acting upon the gel of the GelSight Mini sensor by approximating the output of a Finite Element Analysis (FEA) computation with a neural network that takes the raw RGB image as input. Querying a neural network is much faster than running FEA, and importantly, for FEA computation, one needs a precise geometrical description of the contact, whereas FEATS only needs a raw input image and no further geometric information. Thus, FEATS enjoys fast inference time and does not

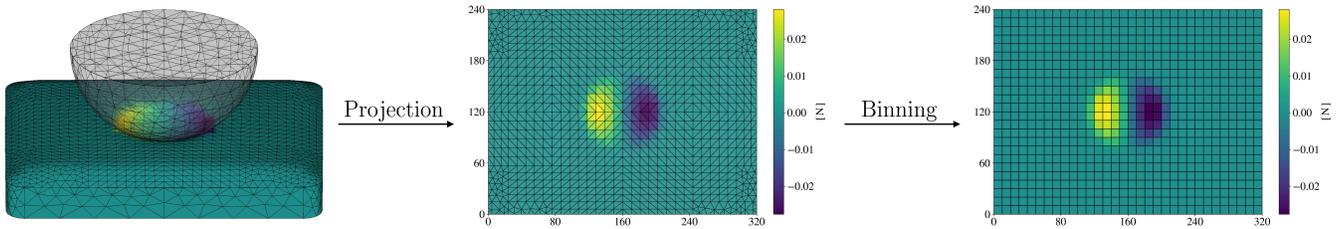


Fig. 3: Label creation process when a spherical indenter presses into the sensor’s soft silicone gel. The visualized contact force distributions correspond to the shear force’s x component. Left: Simulation of the contact configuration and the raw output from running the 3D FEA. Middle: Projecting the result from the 3D FEA into the coordinate system of Gelsight Mini, i.e., into an image plane. Right: Force labels after changing the resolution to 24×32 .

require any additional object tracking equipment at run time. Only at training time FEATS requires a dataset collected in a controlled environment on a set of calibration objects.

A. Data Collection

The data collection process involves a series of precise indentation experiments. We attach a GelSight Mini sensor to the fixed spindle of a Computer Numerical Control (CNC) milling machine, with a positional tolerance of $\pm 0.25 \mu\text{m}$ to automatically create different contact configurations between the sensor and selected indenters (cf. Fig. 1). Following [30], we use 12 indenters of varying shapes and sizes (see Fig. 2). The Gelsight Mini sensor is equipped with the dotted gel to better track the indentation motion. We additionally place a six-axis *RESENSE-HEX-21* Force/Torque (F/T) sensor above the Gelsight Mini to have a complementary external force measurement, which is later used to validate the material model. With this setup, we collected a total of 5173 samples. Each sample contains the respective GelSight Mini RGB image, the CNC motion data, i.e., the current indentation, and the F/T sensor’s readings. The forces in the z direction reach up to 40 N, and in the x and y directions up to ± 5 N.

B. Finite Element Analysis

For simulating the indentation experiments, we employ FEA using the open-source solver CalculiX [42]. The solver’s capability for nonlinear computations makes it an appropriate tool for the stress analysis of soft elastomers. Consequently, it can calculate the resulting force distributions corresponding to the previously described real-world indentation experiments – an essential component for label generation as described in the next Section III-C. We generate tetrahedral volume meshes for both the gel and indenters, with denser meshing at the contact surface to ensure higher accuracy of the simulated contact forces (see Fig. 3). The gel mesh comprises 2504 contact surface elements out of a total of 7591 elements. The elements are implemented as ten-node tetrahedral (C3D10 [42]) elements. The FEA is conducted as a static analysis, assuming hard contact and applying tied contact constraints. This approach permits the execution of simulations without requiring a friction coefficient, and is substantiated by the fact that the experiments are conducted in a manner that precluded slippage. Assuming no deformation of the indenters, they are characterized as a hard material with a Young’s modulus of 210 GPa and a Poisson ratio of 0.3. In [28], the gel elastomer of the

GelSight sensor is characterized as a material similar to a Neo-Hookean solid with a shear modulus, μ , of 0.145. When employing a hyperelastic Neo-Hookean model in the CalculiX solver [42], the strain energy potential is expressed as $U = C_{10}(\bar{I}_1 - 3) + 1/D_1(J - 1)^2$ where \bar{I}_1 denotes the first invariant of the right Cauchy-Green deformation tensor, while J represents the determinant of the deformation gradient tensor. C_{10} and D_1 are the material constants to be set. In agreement with [28], C_{10} is chosen to be 0.0725. Due to CalculiX’s lack of support for perfectly incompressible materials, the solver assigns a default value to D_1 prior to the simulation. This model is later validated in the results section (cf. Sec. IV-A) by optimizing the material parameters through load-depth indentation data [43], comparing the normal forces from the FEA with F/T sensor measurements.

C. Creating Labels

To obtain the “ground truth” force distribution labels, each real-world indentation experiment (cf. Sec. III-A) is repeated in simulation in order to calculate the contact forces using FEA (cf. Sec. III-B & Fig. 3). 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.

a) *Mesh Projection*: To align the results from the three-dimensional FEA 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 through $\mathbf{x}^{\text{GS}} = \mathbf{P} \cdot \mathbf{X}^{\text{FEA}}$, where \mathbf{X}^{FEA} are the 3D coordinates of the mesh nodes, \mathbf{x}^{GS} the image projections and \mathbf{P} the projection matrix. The projection matrix \mathbf{P} can be determined using the least squares method. Four point correspondences are needed for a minimal solution. To establish point correspondences, we gently press the GelSight Mini against an object, such as a cuboid of known size, using the CNC milling machine. Once the object’s shape is clearly visible in the raw sensor image, we identify distinctive object points, such as its corners, in the image. We subsequently repeat the same indentation experiment in simulation and also extract the corner points in the coordinates of the simulation. This leaves us with the point correspondences needed for calculating \mathbf{P} .

b) *Discretization Process*: Once all nodes have been projected onto the image plane, the force distributions for shear and normal forces are binned within the image boundaries. The number of bins decides the resolution of the force distribution. In most experiments, it is 24×32 , however, the

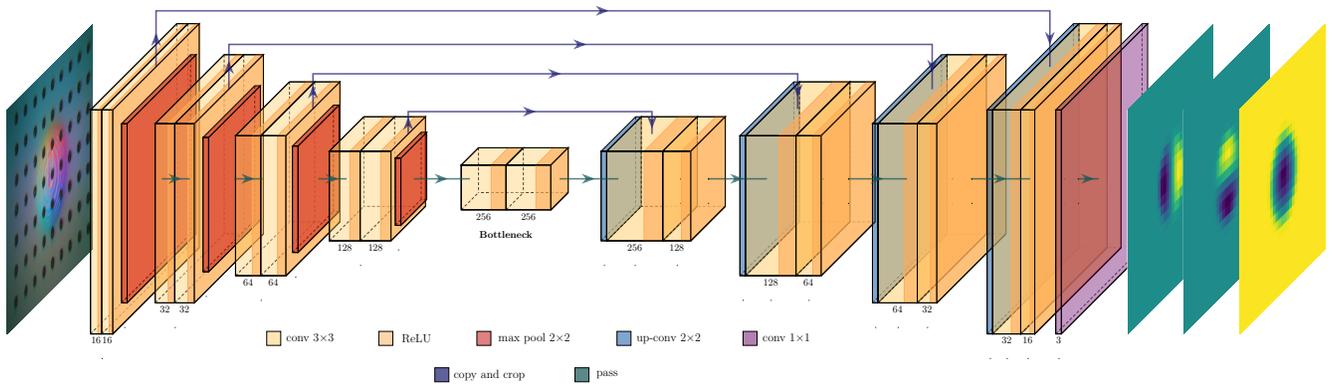


Fig. 4: U-net model which maps raw images from GelSight Mini sensor to shear and normal force distributions. The architecture comprises 4 down-sampling (encoder) and 4 up-sampling (decoder) blocks, connected by skip connections. The number of feature channels at each stage is labeled at the bottom of the corresponding block. Different colors of the boxes and arrows indicate specific operations and activation functions. This image is generated using PlotNeuralNet [44].

resolution can be seamlessly adapted. Each bin’s force value is calculated by summing the contact force contributions of all elements intersecting that bin (see Fig. 3), where the element position refers to its state before deformation. The contribution of each element to a bin is proportional to the fraction of the element that falls within the bin.

D. U-net for Learning Force Distribution Estimation

To predict shear and normal force distributions from the raw GelSight images, we employ a U-net architecture [32], 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 [30]. The model takes 240×320 RGB images as input and outputs three force maps for both shear force components and the normal forces (see Fig. 4).

a) Architecture: The U-net follows an encoder-decoder structure, consisting of a contracting path (encoder) and an expansive path (decoder). High-level image features are extracted by the encoder, by iteratively applying two 3×3 convolutions with zero padding, followed by a ReLU activation function, and a 2×2 max-pooling with a stride of 2. This leads to down-sampling while progressively doubling the feature channels at each step. The decoder up-samples the feature map at each step and then applies a 2×2 convolution, which halves the number of feature channels but doubles the resolution. The up-sampled feature map is then concatenated with the corresponding cropped feature map from the encoder. Subsequently, two 3×3 convolutions are applied, both of which are followed by a ReLU activation. In the final decoder layer, the last feature map is transformed into the specified number of classes via 1×1 convolution.

b) Training: The model is trained by minimizing the Mean Squared Error (MSE) loss

$$\text{MSE}(\mathbf{f}, \hat{\mathbf{f}}) = \frac{1}{3WH} \sum_{i=0}^W \sum_{j=0}^H \|\mathbf{f}^{(i,j)} - \hat{\mathbf{f}}^{(i,j)}\|_2^2 \quad (1)$$

between predicted ($\hat{\mathbf{f}}$) and ground truth (\mathbf{f}) FEA force distribution components. The MSE is averaged across all entries of the force distribution grid.

The dataset is split into three parts: 85% are used for training, 5% for validation, and 10% for testing. Prior to training, both input images and the ground truth force distribution labels are normalized with respect to the training dataset to ensure a consistent baseline across the data. Shear forces are normalized to a range of -1 to $+1$, while normal forces range from 0 to $+1$. Data augmentation techniques, such as adding Gaussian noise and adjusting image brightness, contrast, saturation, and hue, are used to enhance the model’s generalization capabilities. Training is carried out using the Adam optimizer [45], with an initial learning rate of 0.001 and a batch size of 8. The learning rate is adjusted adaptively based on validation performance. As validation loss, the Mean Absolute Error (MAE) on the sum of total forces in the unnormalized space is used. We train for 600 epochs and select the model with the lowest validation loss.

IV. RESULTS

This section evaluates the performance of our proposed FEATS method. The employed FEA model, which is the basis for our label generation procedure, has one free parameter C_{10} that characterizes the response of the material to shear stress and needs to be set correctly. Therefore, we start by validating a value for C_{10} previously reported for GelSight [28] for applicability to GelSight Mini (Sec. IV-A). Subsequently, we evaluate the U-net accuracy in predicting shear and normal force distributions, along with its ability to reconstruct total forces (Sec. IV-B). Particular attention is given to how different force distribution resolutions impact total force reconstruction accuracy. We end with assessing the U-net’s inference speed (Sec. IV-C).

A. Material Characterization

To validate the Neo-Hookean material model parameter C_{10} , load-depth indentation data can be utilized (cf. Sec. III-B). The load-depth measurements are obtained, using a sphere indenter with a 15 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 $\tilde{f}_z^{(i)}$ are compared with

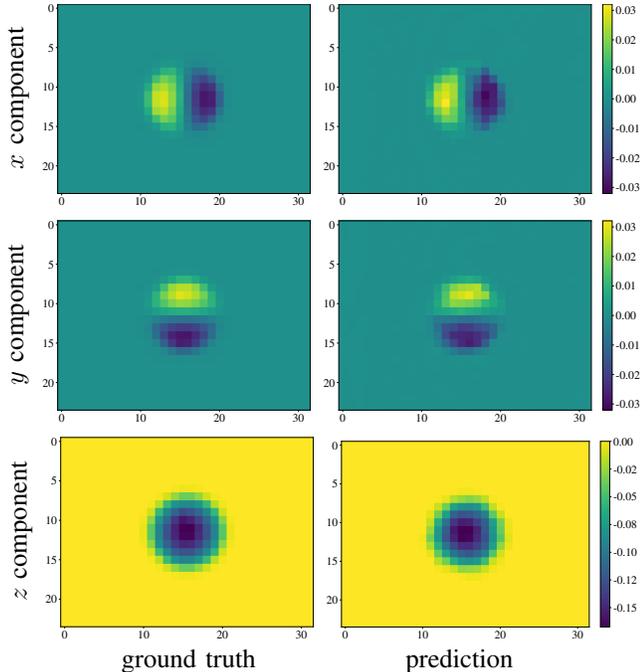


Fig. 5: Ground truth labels (left column) and predictions of our FEATS model (right column) of the force distributions, in Newtons. The sphere indenter (cf. Fig. 3) penetrates the gel for 1.2mm, exerting a significant normal force. The resulting gel deformation causes shear forces, in this case roughly cancelling each other due to the absence of horizontal movement of the indenter.

the forces $f_z^{(i)}(C_{10})$ calculated by the FEA via the MSE loss

$$J(C_{10}) = \frac{1}{N} \sum_{i=1}^N \left(f_z^{(i)}(C_{10}) - \tilde{f}_z^{(i)} \right)^2. \quad (2)$$

When running Bayesian optimization to find the best fit for these $N = 4$ measurements, we found that $\hat{C}_{10} = 0.0792$ provided the best fit with a MAE of 0.5166 N between FEA and F/T sensor measurements. Thus, we confirm that our estimate is within the range of $C_{10} = 0.0725$ previously reported in [28] for the GelSight sensor. In the experiments, we use that value as it was stated in the original GelSight paper and our estimate closely matches it.

B. Evaluation of U-net Force Distribution Estimation

To assess the U-net’s capacity to accurately estimate shear and normal force distributions from the raw RGB images across a wide range of forces, after training, we use a test dataset consisting of 492 samples. The test data is recorded using the same sensor and approach as for the training data, described in Sec. III-A. The value ranges of the ground truth force distribution in the test dataset for the Total Force (TF) estimation task are shown in Fig. 6. Given the nature of the indentation experiments, the GelSight Mini sensor primarily

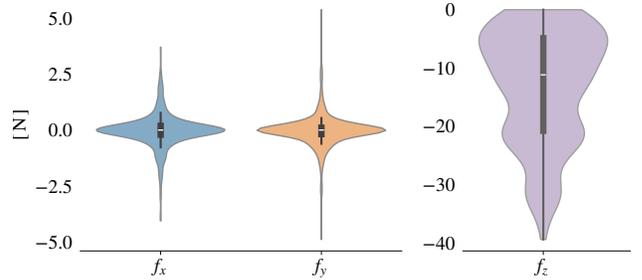


Fig. 6: Test data ranges for the Total Force (TF) estimation task. The range of the tested normal forces is considerably larger than for the shear forces.

encounters normal forces, while the range of the encountered shear forces is comparatively smaller.

a) Evaluation of U-net Predictions on the Test Dataset:

Our main U-net model outputs the force distribution with dimensions $24 \times 32 \times 3$, striking a good balance between maintaining a sufficiently high spatial resolution and at the same time low error on the test data (cf. Tables I and II). An example of the model prediction is shown in Fig. 5, resulting from the contact with a spherical object as in Fig. 3. The U-net demonstrates a noteworthy capability to predict shear and normal force distributions which closely align with the ground truth data (cf. left vs. right column in Fig. 5). This visual comparison highlights the U-net’s effectiveness in capturing the fundamental structures of the underlying force distribution patterns. Additional examples that underscore this finding are provided on the accompanying website.

In order to quantitatively evaluate the performance of predicting force distributions, the MAE metric was used. With MAE, the deviations between the predicted and ground truth force values can be measured, both for per-Grid Unit Forces (GUF) and for the Total Force (TF) values. We are showing the GUF errors since they directly reflect the error in the model outputs and in addition capture the contact area coverage, whereas TF is obtained as the sum of the GUFs and therefore is missing the local force information. The results of the main U-net evaluation on the test dataset are shown in Table I. The U-net is capable of accurately predicting the total shear and normal forces, with MAE values below 1 N. It has been shown that the U-net performs better when dealing with shear forces in the x - and y -directions than with normal forces in the z -direction. This difference in performance may be attributed to two factors. First, the range of force values encountered by the U-net for shear forces are generally smaller than for normal forces. Second, the markers in the GelSight mini sensor’s gel inherently encompass a richer set of features for extracting and representing shear behavior. Shear forces cause a greater displacement of the markers compared to normal forces, and might therefore trigger a clearer signal.

b) Impact of Label Resolution on Prediction Accuracy:

We compare our main U-net model with variants of the U-net architecture and against a ResNet to investigate whether changing the output resolution and network structure affects the prediction quality. The results on the Total Force (TF) estimation task in Table II show that the ResNet, which solely

TABLE I

U-NET MEAN ABSOLUTE ERROR (MAE) ON THE TEST SET

	MAE _{GUF} [N]	MAE _{TF} [N]
f_x	0.0006 ± 0.0006	0.2242 ± 0.4007
f_y	0.0005 ± 0.0003	0.0934 ± 0.1356
f_z	0.0015 ± 0.0010	0.3720 ± 0.4727

TABLE II
MODEL ABLATION ON THE TOTAL FORCE ESTIMATION TASK

Method	MAE _{TF} [N]		
	f_x	f_y	f_z
ResNet ^{1×3}	0.085 ± 0.115	0.069 ± 0.085	1.593 ± 1.131
U-net ^{12×16×3}	0.102 ± 0.216	0.089 ± 0.123	0.447 ± 0.539
3×U-net ^{24×32×1}	0.438 ± 0.585	0.189 ± 0.225	0.448 ± 0.523
U-net ^{24×32×3} (ours)	0.224 ± 0.401	0.093 ± 0.136	0.372 ± 0.473
U-net ^{48×64×3}	0.318 ± 0.436	0.119 ± 0.184	0.459 ± 0.516

regresses to the total force, outperforms all considered U-net architectures in predicting the shear forces but is significantly less accurate in predicting the normal forces. Importantly, the ResNet only outputs a single 3D vector of force for the whole sensor, and therefore significantly lacks in resolution.

The top performing model according to the MAE on the total force is the U-net with output dimension of $12 \times 16 \times 3$. Within the variance of the results, it is however very close to our main architecture with output $24 \times 32 \times 3$, outperforming it on x -shear force but being worse on the z -normal force. That is why we recommend the network with a higher-resolution output, as it strikes the best balance between resolution and achieving small error on the test dataset.

We also compare against training three separate U-net models, where each model is designed to solely predict a single force component, i.e., each individual model outputs a force distribution of shape $24 \times 32 \times 1$. Although these models collectively possess three times the number of parameters, their predictive accuracy is generally inferior to that of the main U-net. This suggests that predicting all three force directions simultaneously can capture correlations between normal and shear forces, yielding more accurate predictions.

Lastly, when training a U-net model with shape $48 \times 64 \times 3$, its predictive accuracy remains relatively strong but exhibits a slight decline compared to our main U-net.

c) Generalization to Different Sensors: To explore the generalization capability of our FEATS method, it is tested on images captured with a different GelSight Mini sensor of the same type (cf. Fig. 7). The predictions remain largely accurate, although there is a notable downward and slight rightward shift in the localization of the contact area. This shift is likely due to differences in the alignment of the camera relative to the gel, thus causing the markers in the image of the test sensor to appear lower than in the image of the GelSight Mini sensor used during training. Furthermore, sensor-related factors such as image brightness, contrast, saturation and hue may also contribute to the shift in the localization of the contact area.

C. Inference Speed

To determine the speed of the U-net, we measure the time it takes to process a single input on a *NVIDIA Quadro RTX 5000 GPU* and a *Intel Core i7-10875H 8-Core CPU* with a frequency of 2.3 GHz. The time required for transferring data between the CPU and GPU, as well as asynchronous execution and warm-up of the GPU is taken into consideration. On average, the U-net achieves an inference time of 4.1746 ± 1.339 ms over 300 runs, which is sufficient

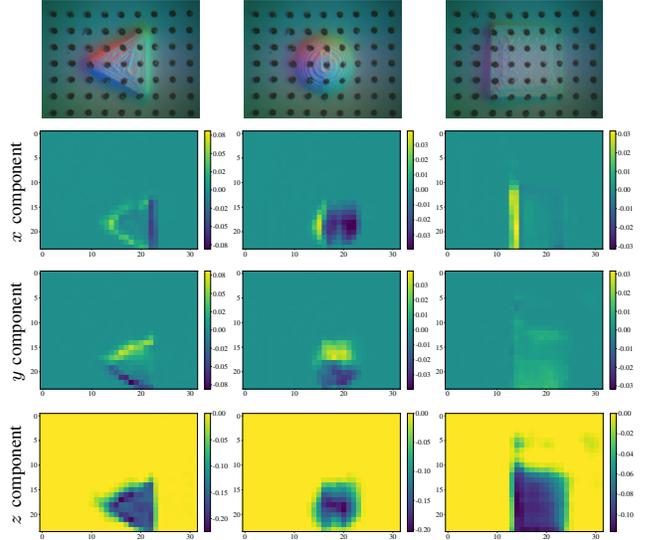


Fig. 7: Generalization to a different sensor. We evaluate our FEATS U-net on indenters from the test set using a different GelSight Mini sensor to validate the model generalization capability. The first row displays the raw images captured by the sensor, and the subsequent rows show the components of the force distribution, in Newtons. While the general outline of the force distribution is captured correctly, there is a noticeable downward and slight rightward shift in the localization of the predicted forces, likely due to different alignments of the gels in different sensors.

for obtaining the force distributions in real-time since the GelSight Mini sensor runs at 25 Hz. This is significantly faster compared to running the FEA simulations, which may take 10 – 120 minutes depending on the contact geometry.

V. CONCLUSION

We introduced FEATS—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, with Mean Absolute Error (MAE) under 0.4 N on average in the range 0 – 40 N for the total normal force and -5 N to 5 N for total shear force in each direction. The model shows potential for real-time applications and generalization to different sensors, although additional improvements are needed to account for the domain gap when a different sensor is used, which could be addressed, e.g., via the recently proposed domain adaptation for optical tactile sensors [46]. Our FEATS method offers an efficient online approach that amortizes the cost of running inverse FEA into a forward pass of a neural network, resulting in physically-grounded, interpretable representations for optical tactile sensors. Future work will focus on enhancing the model’s robustness to different tactile scenarios, refining its generalization capabilities across a broader range of objects and sensor types, and work towards demonstrating the representation’s effectiveness in robotic manipulation tasks.

ACKNOWLEDGEMENT

We thank LAB³, Kai Ruf & Uwe Faltermeier for their great support and access to the CNC milling machine, and we acknowledge the computing time provided on the Lichtenberg cluster at TU Darmstadt.

REFERENCES

- [1] L. D. Harmon, "Automated tactile sensing," *The International Journal of Robotics Research*, vol. 1, no. 2, pp. 3–32, 1982.
- [2] —, "Tactile sensing for robots," *Robotics and Artificial Intelligence*, pp. 109–157, 1984.
- [3] R. D. Howe and M. R. Cutkosky, "Integrating tactile sensing with control for dextrous manipulation," in *IEEE International Workshop on Intelligent Motion Control*, no. 1, 1990, pp. 369–374.
- [4] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019.
- [5] B. Gates, "A Robot in Every Home," *Scientific American*, vol. 296, no. 1, pp. 58–65, 2007.
- [6] A. C. Abad and A. Ranasinghe, "Visuotactile Sensors With Emphasis on GelSight Sensor: A Review," *IEEE Sensors Journal*, vol. 20, no. 14, pp. 7628–7638, Jul. 2020.
- [7] G. Westling and R. S. Johansson, "Factors influencing the force control during precision grip," *Experimental Brain Research*, vol. 53, no. 2, pp. 277–284, Jan. 1984.
- [8] L. Lach, N. Funk, R. Haschke, S. Lemaignan, H. J. Ritter, J. Peters, and G. Chalvatzaki, "Placing by touching: An empirical study on the importance of tactile sensing for precise object placing," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 8964–8971.
- [9] M. Kaboli, K. Yao, and G. Cheng, "Tactile-based manipulation of deformable objects with dynamic center of mass," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, Nov. 2016, pp. 752–757, iSSN: 2164-0580.
- [10] R. Sui, L. Zhang, T. Li, and Y. Jiang, "Incipient Slip Detection Method With Vision-Based Tactile Sensor Based on Distribution Force and Deformation," *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25973–25985, Nov. 2021.
- [11] H. Yussof, J. Wada, and M. Ohka, "Sensorization of Robotic Hand Using Optical Three-Axis Tactile Sensor: Evaluation with Grasping and Twisting Motions," *Journal of Computer Science*, vol. 6, no. 8, pp. 955–962, Aug. 2010.
- [12] M. I. Tiwana, S. J. Redmond, and N. H. Lovell, "A review of tactile sensing technologies with applications in biomedical engineering," *Sensors and Actuators A: Physical*, vol. 179, pp. 17–31, Jun. 2012.
- [13] W. Othman, Z.-H. A. Lai, C. Abril, J. S. Barajas-Gamboa, R. Corcelles, M. Kroh, and M. A. Qasaimeh, "Tactile Sensing for Minimally Invasive Surgery: Conventional Methods and Potential Emerging Tactile Technologies," *Frontiers in Robotics and AI*, vol. 8, 2022.
- [14] Y. Zhu, S. Nazirjonov, B. Jiang, J. Colan, T. Aoyama, Y. Hasegawa, B. Belousov, K. Hansel, and J. Peters, "Visual Tactile Sensor Based Force Estimation for Position-Force Teleoperation," Dec. 2022, arXiv:2212.13007 [cs].
- [15] Y. Ito, Y. Kim, C. Nagai, and G. Obinata, "Shape sensing by vision-based tactile sensor for dexterous handling of robot hands," in *2010 IEEE International Conference on Automation Science and Engineering*. IEEE, 2010, pp. 574–579.
- [16] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, and E. H. Adelson, "Shape-independent hardness estimation using deep learning and a gelsight tactile sensor," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 951–958.
- [17] B. Fang, F. Sun, C. Yang, H. Xue, W. Chen, C. Zhang, D. Guo, and H. Liu, "A dual-modal vision-based tactile sensor for robotic hand grasping," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 4740–4745.
- [18] F. Sun, B. Fang, H. Xue, H. Liu, and H. Huang, "A novel multi-modal tactile sensor design using thermochromic material," *Science China Information Sciences*, vol. 62, pp. 1–3, 2019.
- [19] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile Sensing—From Humans to Humanoids," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 1–20, Feb. 2010.
- [20] 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)*, 2022, pp. 3651–3658.
- [21] W. Li, A. Alomainy, I. Vitanov, Y. Noh, P. Qi, and K. Althoefer, "F-touch sensor: Concurrent geometry perception and multi-axis force measurement," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4300–4309, 2021.
- [22] F. Baghaei Naeini, D. Makris, D. Gan, and Y. Zweiri, "Dynamic-vision-based force measurements using convolutional recurrent neural networks," *Sensors*, vol. 20, no. 16, 2020.
- [23] C. Sferrazza, T. Bi, and R. D'Andrea, "Learning the sense of touch in simulation: a sim-to-real strategy for vision-based tactile sensing," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 4389–4396.
- [24] K. Kamiyama, H. Kajimoto, N. Kawakami, and S. Tachi, "Evaluation of a vision-based tactile sensor," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, vol. 2. IEEE, 2004, pp. 1542–1547.
- [25] 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)*, May 2019, pp. 5418–5424, iSSN: 2577-087X.
- [26] L. Van Duong *et al.*, "Large-scale vision-based tactile sensing for robot links: Design, modeling, and evaluation," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 390–403, 2020.
- [27] L. Zhang, Y. Wang, and Y. Jiang, "Tac3D: A Novel Vision-based Tactile Sensor for Measuring Forces Distribution and Estimating Friction Coefficient Distribution," Feb. 2022, arXiv:2202.06211 [cs].
- [28] W. Yuan, S. Dong, and E. Adelson, "GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force," *Sensors*, vol. 17, no. 12, p. 2762, Nov. 2017.
- [29] 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.
- [30] 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*, vol. 9, no. 5, pp. 926–937, 2022.
- [31] C. Sferrazza, A. Wahlsten, C. Trueeb, and R. D'Andrea, "Ground Truth Force Distribution for Learning-Based Tactile Sensing: A Finite Element Approach," *IEEE Access*, vol. 7, pp. 173 438–173 449, 2019.
- [32] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation."
- [33] GelSight, Inc., "Gelsight mini - datasheet," https://www.gelsight.com/wp-content/uploads/2023/01/GelSight_Datasheet_GSMMini_12.20.22.pdf, accessed: 2024-03-04.
- [34] M. Li, T. Li, and Y. Jiang, "Marker Displacement Method Used in Vision-Based Tactile Sensors—From 2-D to 3-D: A Review," *IEEE Sensors Journal*, vol. 23, no. 8, pp. 8042–8059, Apr. 2023.
- [35] M. K. Johnson and E. H. Adelson, "Retrographic sensing for the measurement of surface texture and shape," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1070–1077.
- [36] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson, "Measurement of shear and slip with a GelSight tactile sensor," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. Seattle, WA, USA: IEEE, May 2015, pp. 304–311.
- [37] J. Li, S. Dong, and E. Adelson, "Slip detection with combined tactile and visual information," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7772–7777.
- [38] X. Jia, R. Li, M. A. Srinivasan, and E. H. Adelson, "Lump detection with a gelsight sensor," in *2013 World Haptics Conference (WHC)*. IEEE, 2013, pp. 175–179.
- [39] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez, "Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1927–1934.
- [40] H. Sun, K. J. Kuchenbecker, and G. Martius, "A soft thumb-sized vision-based sensor with accurate all-round force perception," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 135–145, Feb. 2022.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015, arXiv:1512.03385 [cs].
- [42] G. Dhondt, "Calculix crunchix user's manual version 2.21," https://www.dhondt.de/cex_2.21.pdf, 2023, accessed: 2024-03-05.
- [43] Z. Chen, T. Scheffer, H. Seibert, and S. Diebels, "Macroindentation of a soft polymer: Identification of hyperelasticity and validation by uni/biaxial tensile tests," *Mechanics of Materials*, vol. 64, pp. 111–127, Sep. 2013.
- [44] H. Iqbal, "Harisqbal88/plotneuralnet v1.0.0," Dec. 2018. [Online]. Available: <https://doi.org/10.5281/zenodo.2526396>
- [45] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Jan. 2017, arXiv:1412.6980 [cs].
- [46] Z. Chen, N. Ou, J. Jiang, and S. Luo, "Deep domain adaptation regression for force calibration of optical tactile sensors," *arXiv preprint arXiv:2407.14380*, 2024.