



Transparent 9DTact: Tactile–Vision 3D Contact Reconstruction and Force Estimation

Oliver Young¹ Zhang Tao (PhD Student)¹

¹The Chinese University of Hong Kong ²University of California, Santa Cruz
Supervisor: Prof. [Hongming Ren]

UC SANTA CRUZ

Abstract

We integrate a standard camera workflow with tactile sensing in a transparent 9DTact, maintaining full RGB image/video capture while estimating contact depth and force. From frames, we create object masks using traditional OpenCV segmentation, SAM2, or their intersection. Pixel brightness/colour changes are rectified, background-removed, and mapped to depth via a calibrated Pixel→Depth LUT, yielding deformation fields. We experimentally cross-reference these deformations with ground-truth forces to learn simple, interpretable mappings (mean/max/sum depth → force) and to visualize per-pixel force. The current poster emphasizes robust rectification, consistent masking, and honest evaluation: we summarize strong qualitative results and highlight challenges (mask semantics, photometric sensitivity, alignment) that inform the next iteration of calibration and benchmarking.

Motivation and Background

- 9DTact is a low-cost, effective tactile sensor for shape reconstruction and force estimation with broad applications in dexterous manipulation.
- Aim: advance 9DTact to sense approaching objects (pre-contact awareness) while preserving accurate contact shape and force mapping after impact.
- Approach: combine camera-first perception (RGB, SAM2) with calibrated photometric depth to keep the system a useful camera and a tactile sensor.
- Focus: robust rectification, consistent mask semantics, calibrated Pixel→Depth, and interpretable (simple) force models.

Hardware & Setup

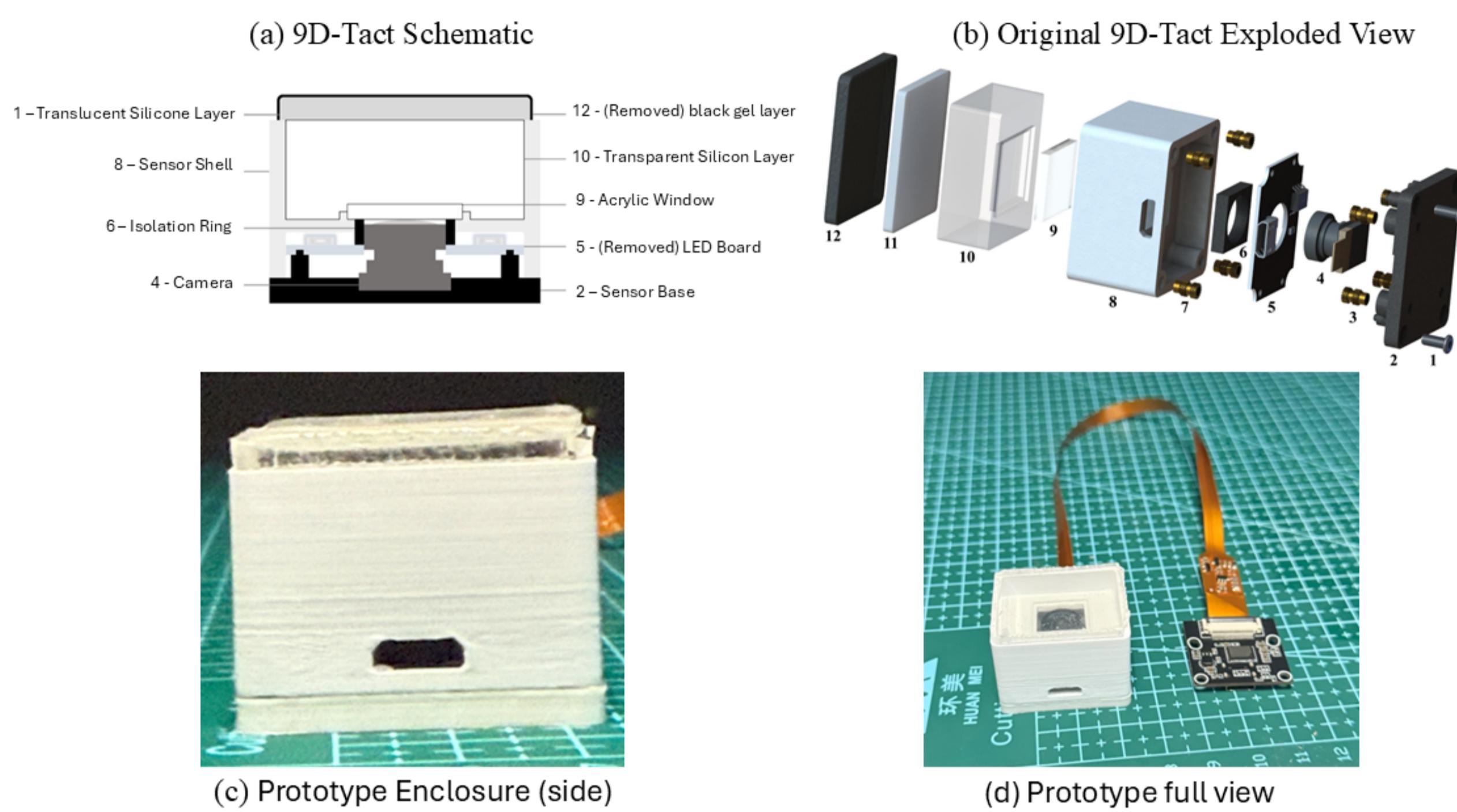


Figure 1. (a) 9D-Tact schematic highlighting the optical path and numbered parts. (b) Original 9D-Tact exploded view (numbers match labels). (c) Prototype enclosure (side) showing the transparent + translucent gel stack on an acrylic window. (d) Prototype full view with camera board and FFC.

Context (what changed)

- Goal: enable on-sensor imaging/video while preserving calibrated depth sensing.
- Change vs. legacy: remove the opaque black gel (12) and LED board (5); use a transparent (10) + translucent (1) silicone stack over the acrylic window (9).
- Housing/optics: shell (8), isolation ring (6), sensor base (2), camera (4).
- Benefit: RGB frames for contact visualization and per-frame processing; compatible with Pixel→Depth LUT pipeline.

Mini-legend (numbers → parts)

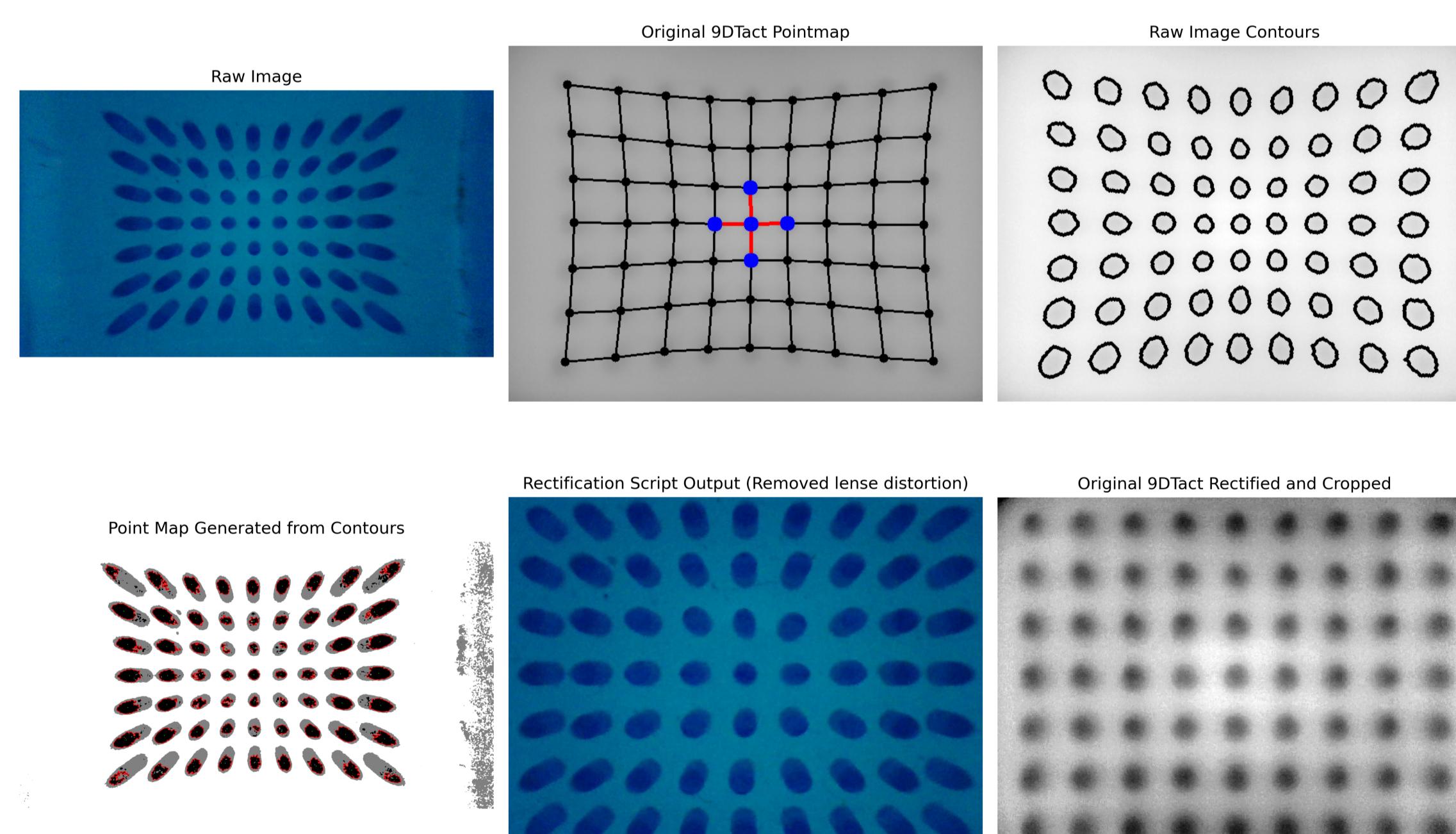
- 1 Translucent silicone layer 2 Sensor base 4 Camera
- 5 (Removed) LED board 6 Isolation ring 8 Sensor shell
- 9 Acrylic window 10 Transparent silicone layer 12 (Removed) black gel layer

Pipeline Overview

- 1) **Rectify & Crop:** apply precomputed maps (row_index, col_index, position_scale) and center-crop to suppress edge artifacts.
- 2) **Contact Mask:** compute a Gaussian high-pass background removal, threshold, and morphology; optionally intersect with SAM2 instance segments; keep the largest connected component.
- 3) **Depth Reconstruction:** convert pixel differences to depth using a calibrated Pixel→Depth LUT and a lighting threshold; smooth with small Gaussian kernels.
- 4) **Force Features:** compute mean/max/sum depth within the contact region and record these statistics.
- 5) **Regression & Visualization:** fit a linear mapping from mean depth (mm) to force (N); render depth maps, point clouds, and a per-pixel force surface.

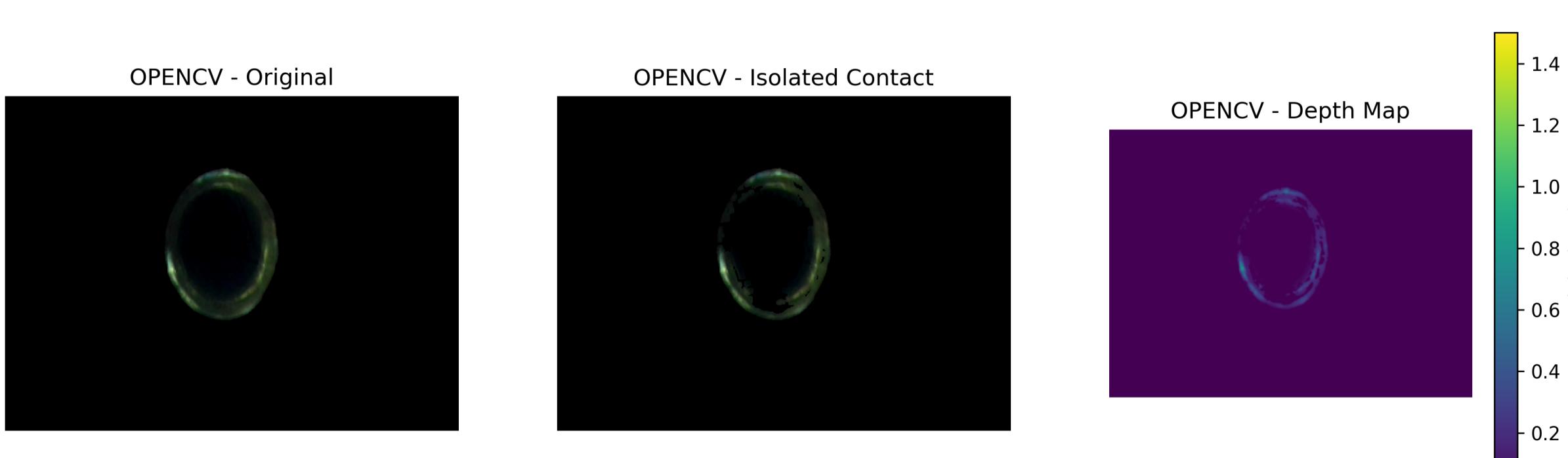
Calibration and Camera Distortion Removal

Rectification aligns pixels to the gel and removes lens distortion before any masking or depth conversion. We leverage SAM2 on RGB frames to separate foreground from background when needed, while maintaining compatibility with the original 9DTact calibration flow for direct comparison. The figure shows the rectified output used to build the dataset.



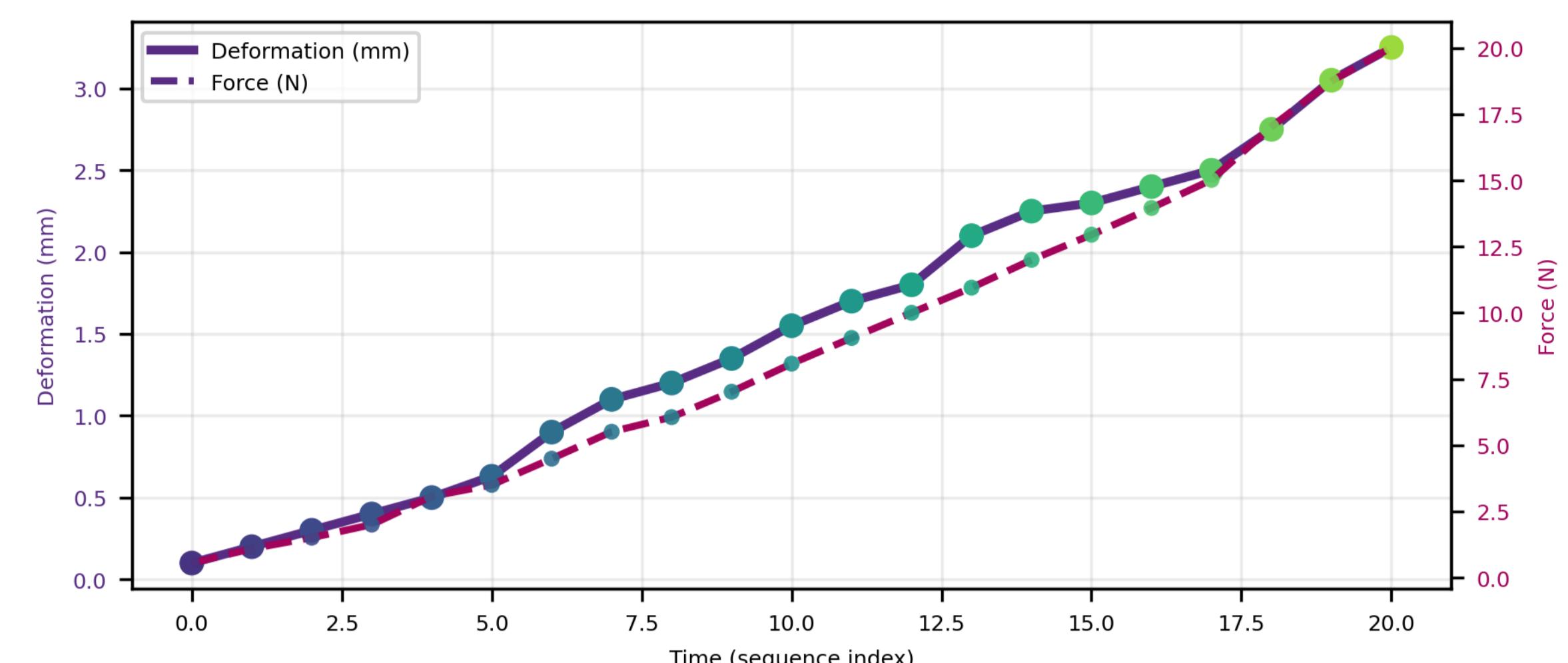
Contact-Region Detection

High-pass + threshold + morphology: compute a Gaussian high-pass $I_{hp} = I - (G_\sigma * I)$ to remove low-frequency background, threshold to isolate deformation, and apply open/close morphology to denoise. Optionally intersect with SAM2 segments to suppress background. Key variables: hp.blur (Gaussian kernel for background, larger \Rightarrow smoother), threshold (binary decision on $|I_{hp}|$), morph_ks (kernel for open/close), and center-crop ratio (ROI for stability).



Force–Deformation over Time

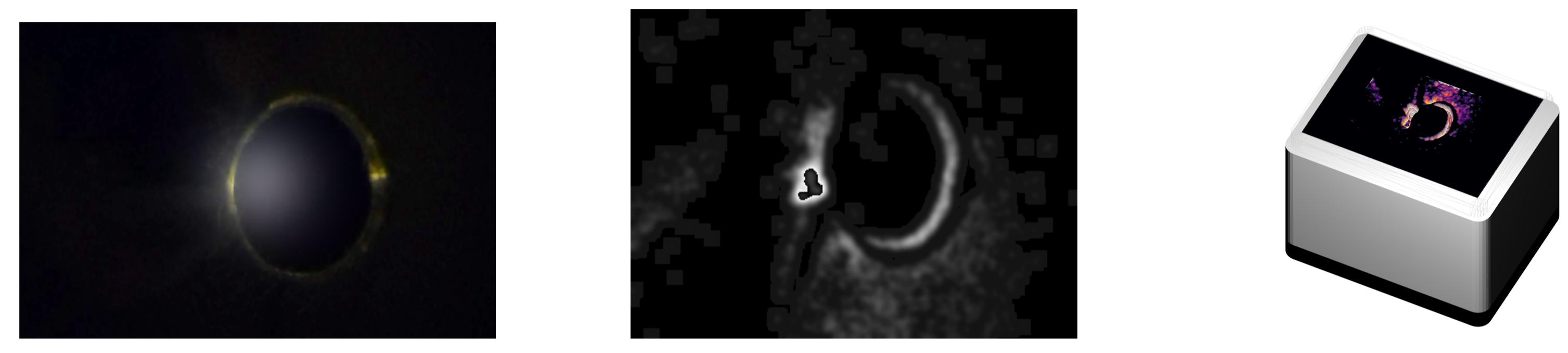
We map applied force against gel deformation over the acquisition sequence (negative Z; start depth 21.25 mm). Deformation is computed as $\delta = d_{start} - d_t$. The plot shows monotonic increase of δ with force and supports a first-order stiffness estimate.



Rigidity (stiffness) estimate: fit a line $F = k\delta + b$ over the quasi-linear region to obtain an effective stiffness k (N/mm). Report k with confidence intervals; check linearity (residuals) and hysteresis by sweeping up/down in force. This constant summarizes the material/system resistance to deformation under normal load and can be compared across sensors or sessions.

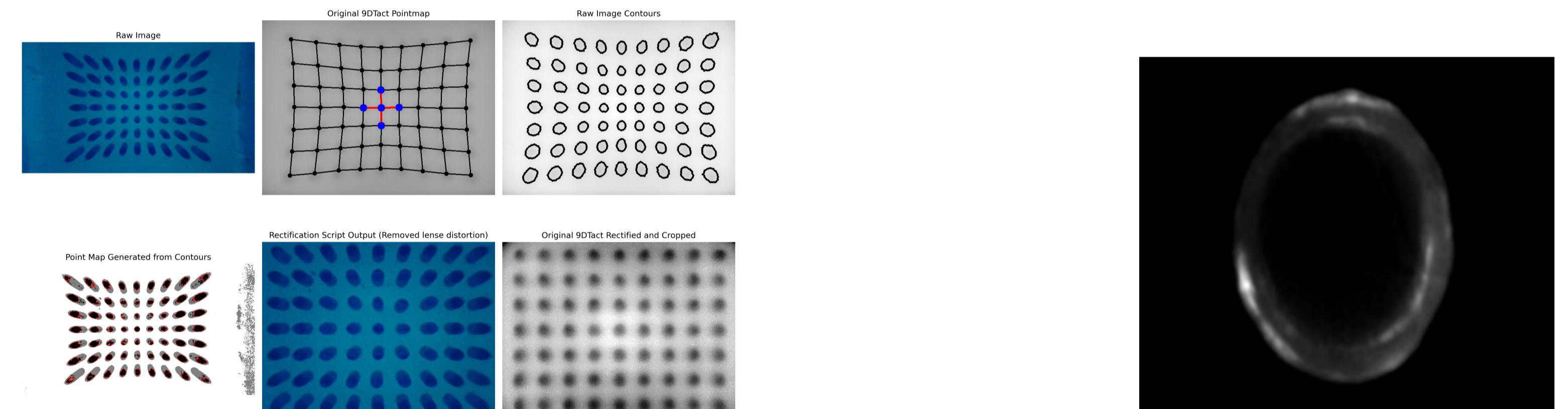
Depth Reconstruction

Use a calibrated Pixel→Depth LUT (Pixel_to_Depth_iterative.npy) built from shallow-to-deep calibration video. For a pixel-difference value g , depth $d(g)$ is computed via piecewise linear mapping gated by a lighting threshold. The depth map is then smoothed with small Gaussian kernels to reduce quantization noise while preserving shape. Outputs include raw, depth, and point-cloud views.



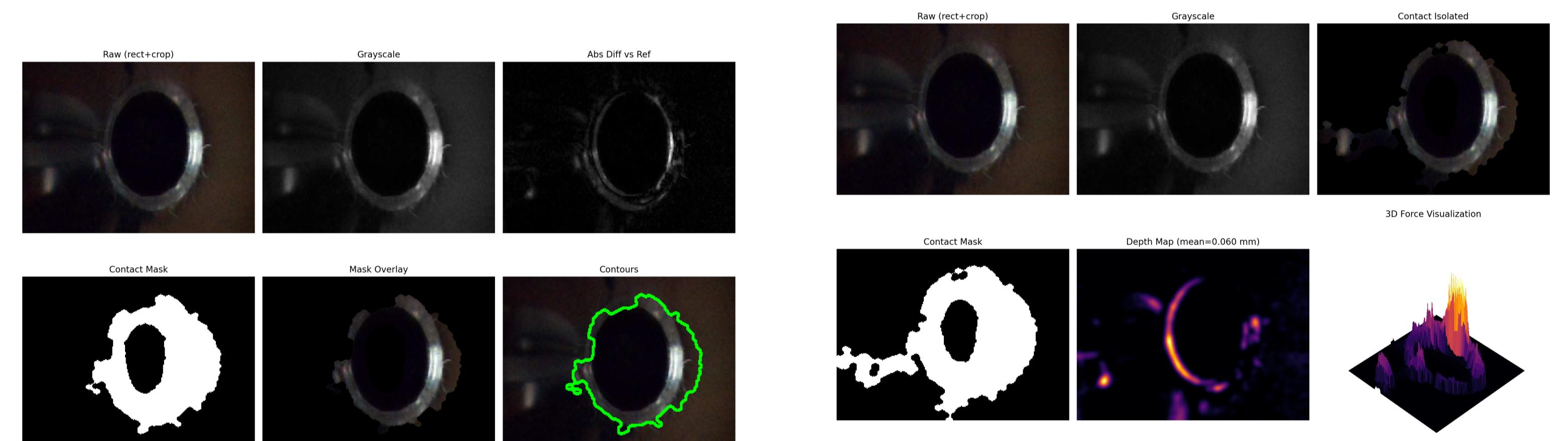
Force–Depth Dataset and Regression

For each labeled image (and sampled video frame), we rectify, mask the contact region, compute depth statistics (mean, max, sum), and record them to a CSV: filename, force_N, mean_depth_mm, max_depth_mm, sum_depth_mm. A simple linear baseline predicts force as $\hat{F} = \alpha\bar{d} + \beta$, where \bar{d} is the mean depth (mm) inside the contact mask. These features support quick parameter sweeps and baseline evaluation; non-linear mappings and learned regressors are topics for future work.



Case Study: 15N Contact and 3-D Force Visualization

The panels below illustrate the full pipeline on a representative 15 N contact image. Left: contact-detection composite highlighting the segmented object and the resulting contact mask. Right: multi-panel figure showing rectified grayscale, contact-isolated view, depth map, and a 3-D force surface derived from brightness within the contact region and normalized to the labeled force.



Notes for interpretation:

- **Contact Mask:** high-pass filtering + threshold + morphology, intersected with SAM-2 segments to suppress background; thresholds tuned for circular objects.
- **Depth Map:** derived via the calibrated Pixel→Depth LUT; brighter hues correspond to larger indentation depths.
- **3-D Force Surface:** per-pixel force $F(x, y)$ computed by normalizing grayscale within the mask to $[0, 1]$ and scaling by the labeled force in newtons; peaks indicate higher local pressure.

Preliminary Results

- Mean depth increases monotonically with applied force across the labeled range (≈ 0.5 – 20 N).
- Intersecting high-pass masks with SAM-2 segments yields clean, near-circular contact regions for axisymmetric indenters.
- 3-D force surfaces highlight pressure peaks and align with deeper basins in the depth map.
- Preliminary video frames show stable contact footprints and reveal transient peaks during load transitions.

Limitations and Future Work

Limitations: sensitivity to lighting, glare and haze in the transparent medium; gel aging/hysteresis; focus on normal force only; current LUTs are sensor/session-specific; cross-object generalization not yet studied. **Future Work:** extend to shear and torque estimation; develop temporal models (optical flow + depth) leveraging video; improve photometric calibration and SAM-2 segmentation; perform on-robot tests; explore non-linear regressors and domain adaptation across sensors and lighting conditions.

Acknowledgements

I thank Prof. Hongliang Ren for continuous support and for providing all resources needed to make this work possible. I am especially grateful to Zhang Tao, who offered exceptional guidance throughout the program and ensured the equipment and environment were ready for experimentation. Without the mentorship of Prof. Ren and Zhang Tao, this project would not have been possible.

References

- Luu, Q.K., Nguyen, D.Q., Nguyen, N.H., Dam, N.P., Ho, V.A. Vision-based Proximity and Tactile Sensing for Robot Arms: Design, Perception, and Control. Project website: <https://quan-luu.github.io/protac-website>.
- Lin, C., Zhang, H., Xu, J., Wu, L., Xu, H. 9DTact: A Compact Vision-Based Tactile Sensor for Accurate 3D Shape Reconstruction and Generalizable 6D Force Estimation. Project site: <https://linchangyi.github.io/9DTact>. arXiv:2308.14277.
- IEEE Xplore document (arnumber 11027485): <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=11027485>.