

Embodied Vision

Tactile Sensing

Tsung-Wei Ke

Spring 2026



We Have Talked About Physically Modeling of the 4D World...

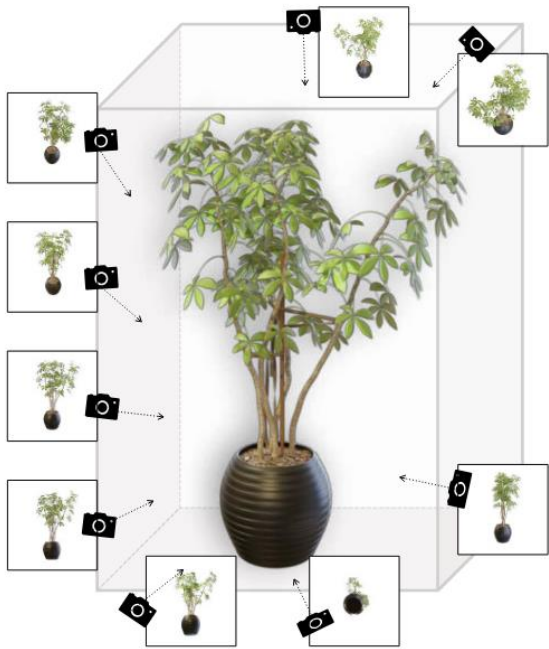


Image formation and stereopsis



3D / 4D Modeling



Physical Modeling

What We Discussed is like Watching Movies from a Box



But, Human Has Other Types of Sensing

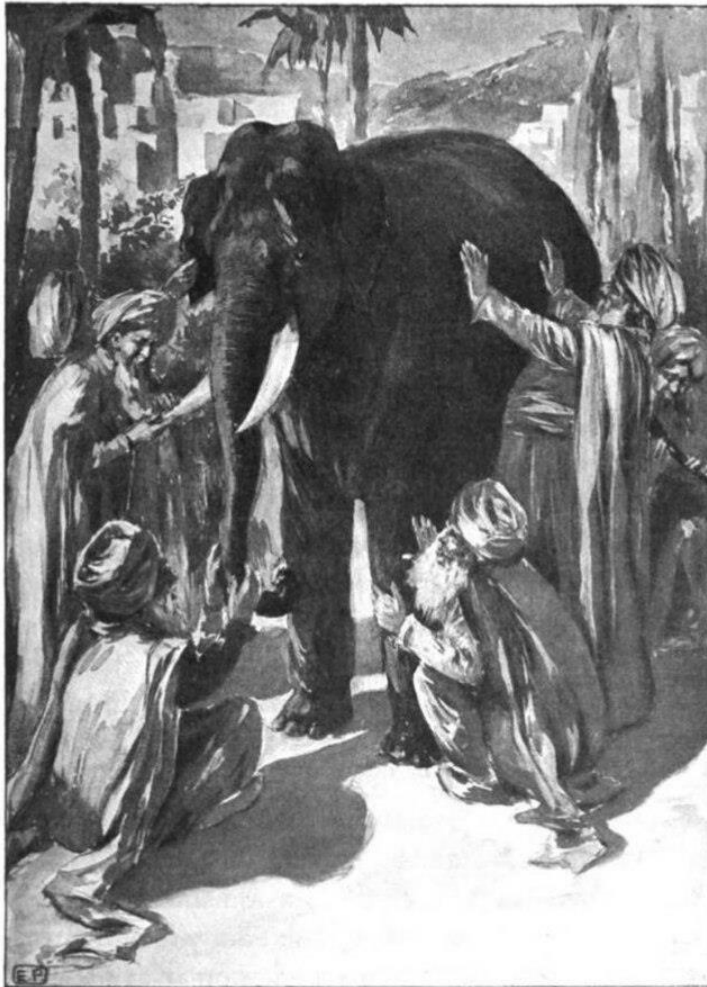


Image source: Wikipedia



<https://youtu.be/FQhn00FArs8?si=yHawxIFmQPakg3X8>

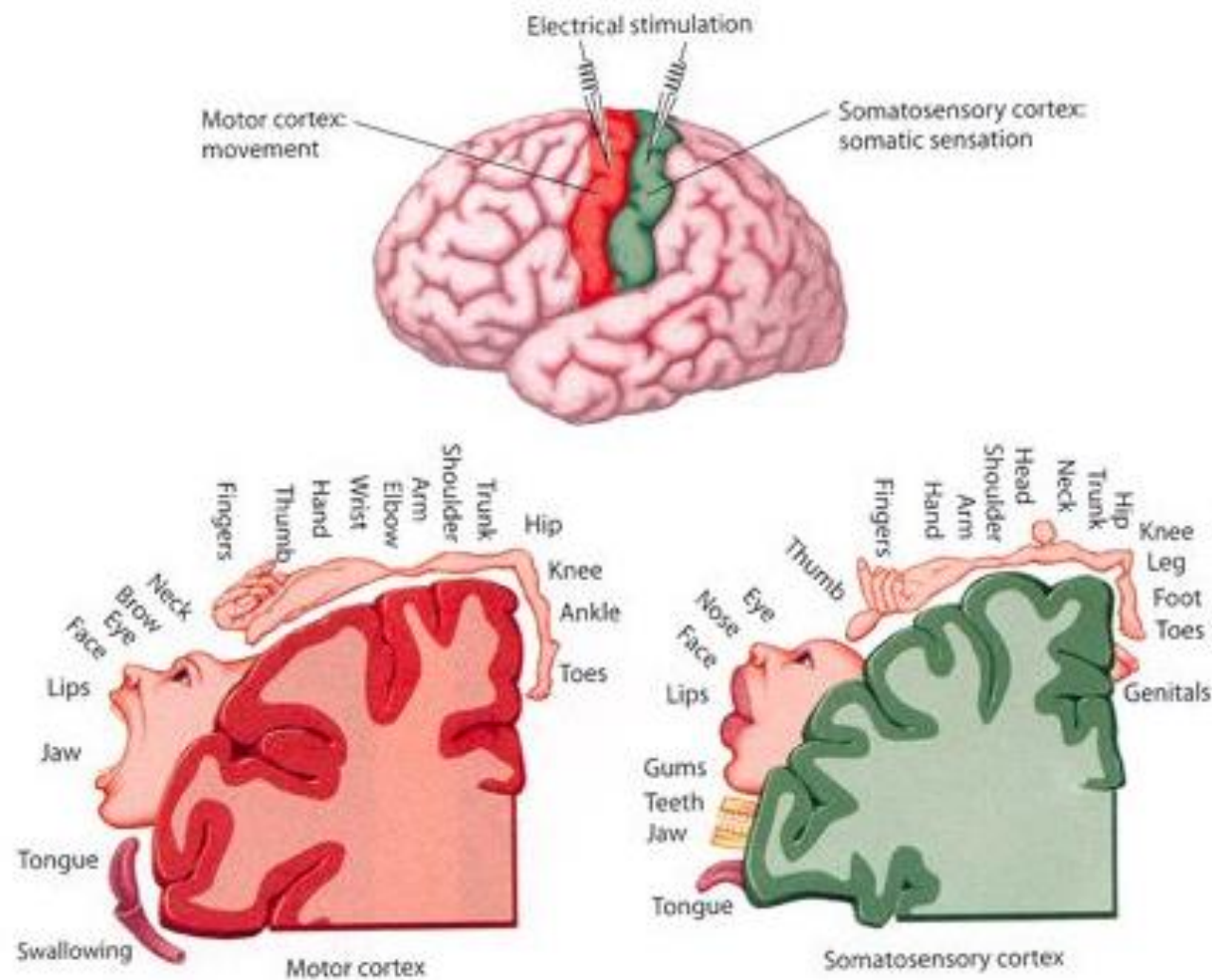
A Visual, Aural and Physical World



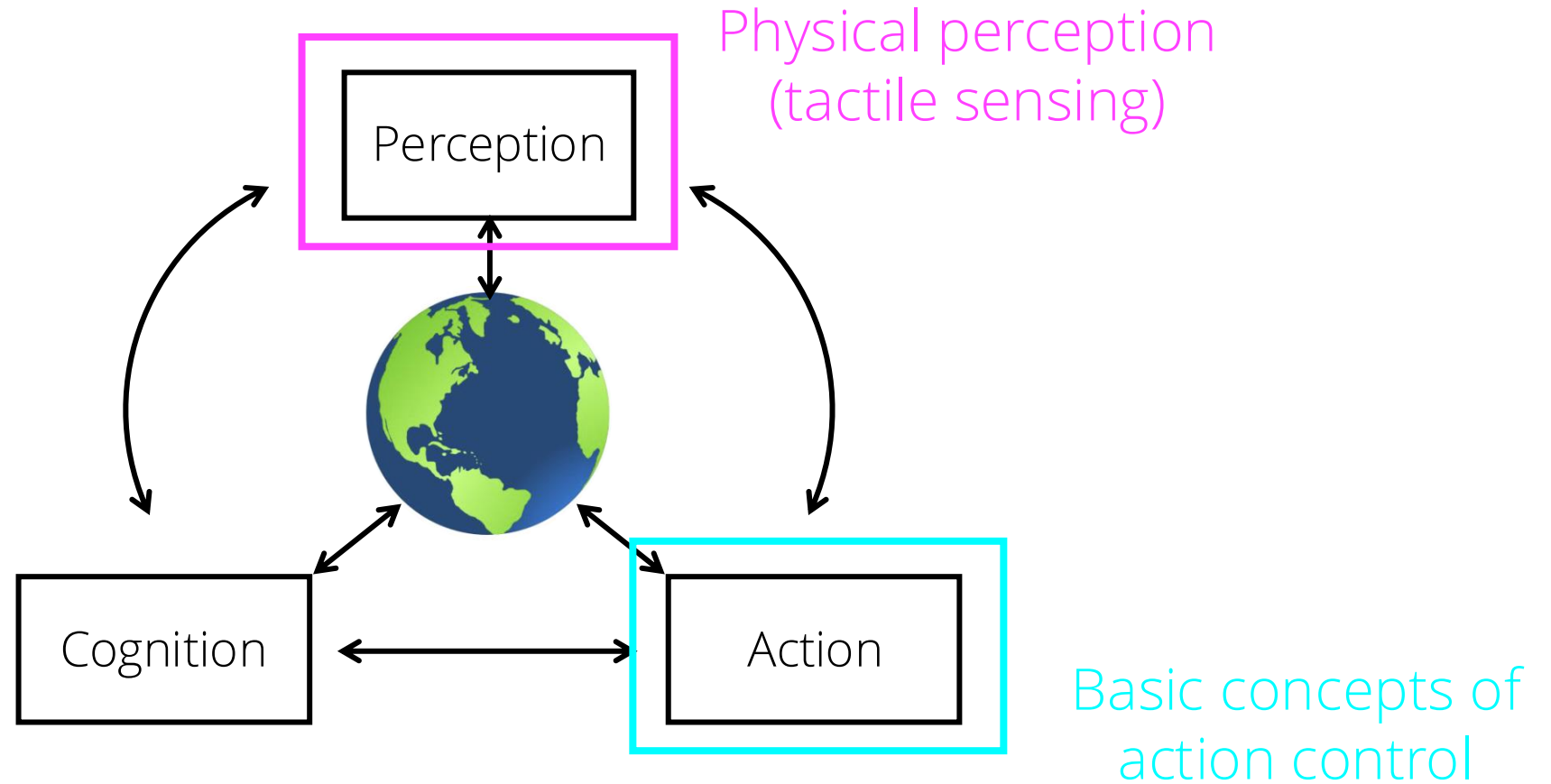
Meanwhile, Human Acts Not Just Watches

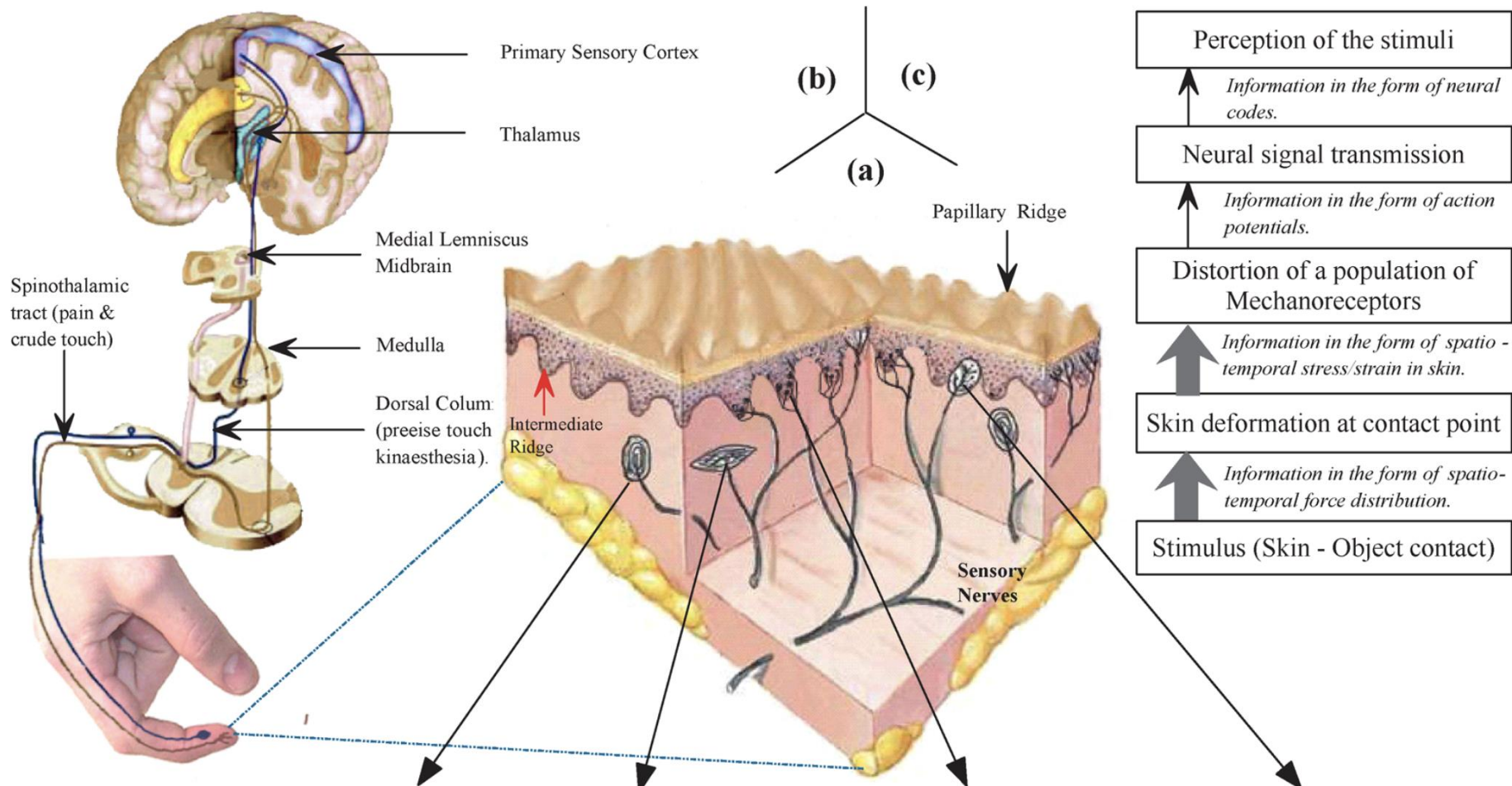


Our Brain is Created to Perceive and Sense



In the Following Weeks, We'll Talk About





Classification Basis	Pacinian Corpuscle	Ruffini Corpuscle	Merkel Cells	Meissner's Corpuscle
Type	FA II	SA II	SA I	FA I
Adaptation Rate	Fast	Slow	Slow	Fast
Spatial Acuity (mm)	10+	7+	0.5	3-4
Vibration/rapid indent. threshold	Best(μm)	40	8	2
	Mean(μm)	300	30	6
Stimuli Frequency (Hz)	40-500+	100-500+	0.4-3	3-40
Conduction Velocity (m/s)	35-70	35-70	40-65	35-70
Effective Stimuli	Temporal changes in the skin deformation	Sustained downward Pressure; Lateral skin stretch; Skin slip.	Spatial deformation; Sustained pressure; Curvature, edge, corners.	Temporal changes in skin deformation
Sensory Function	High frequency vibration detection; Tool use.	Finger position; Stable grasp; Tangential Force; Motion direction	Pattern/form detection; texture perception; Tactile flow perception.	Low frequency vibration & motion detection; Grip control; Tactile flow perception.

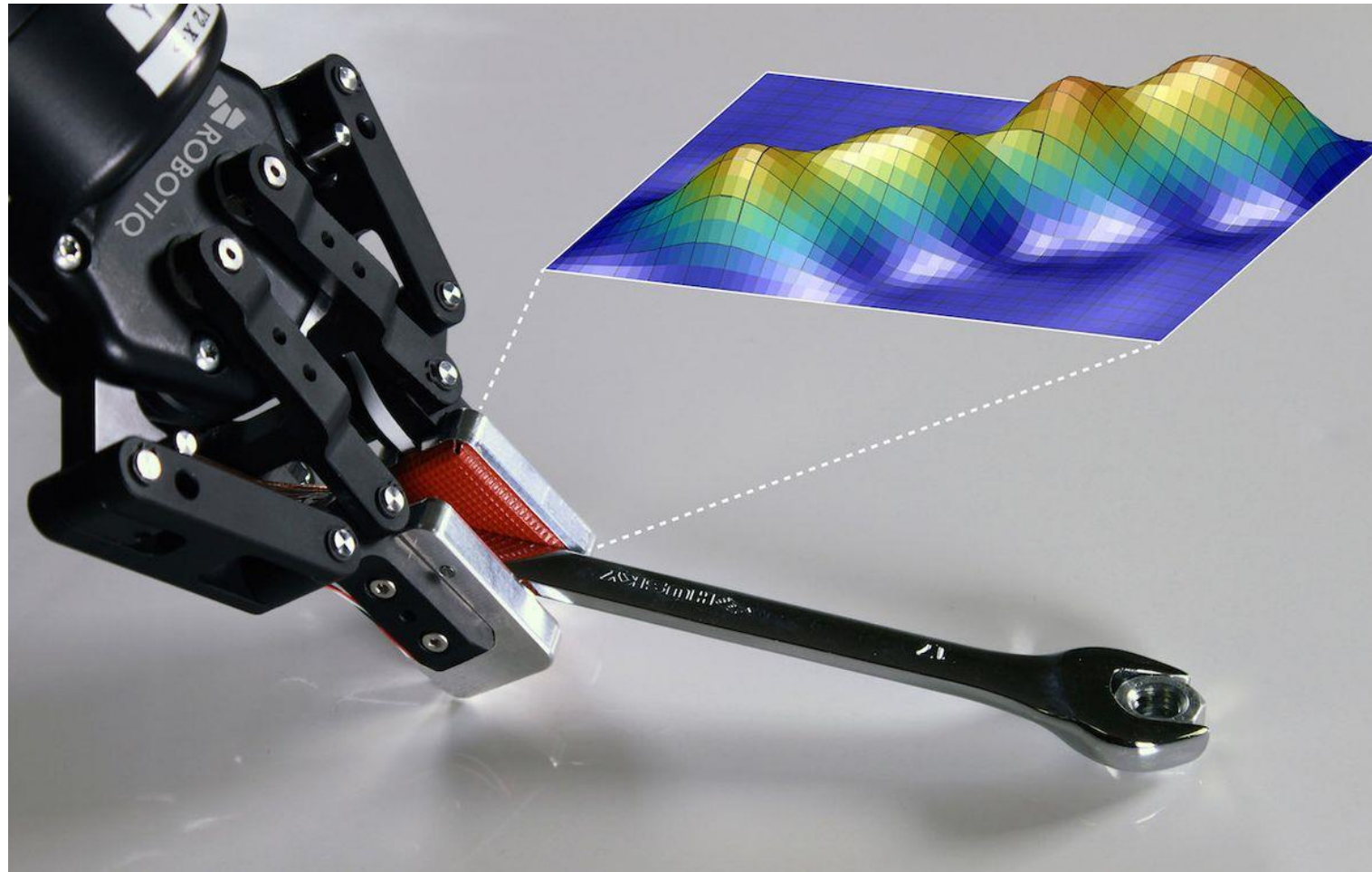
What Information Can be Derived from Tactile Sensing?

- Contact vs. no contact
- Force (normal, shear)
- Temperature
- Texture
- Friction
- Object pose or motion
- Detailed Geometry of an Object
-

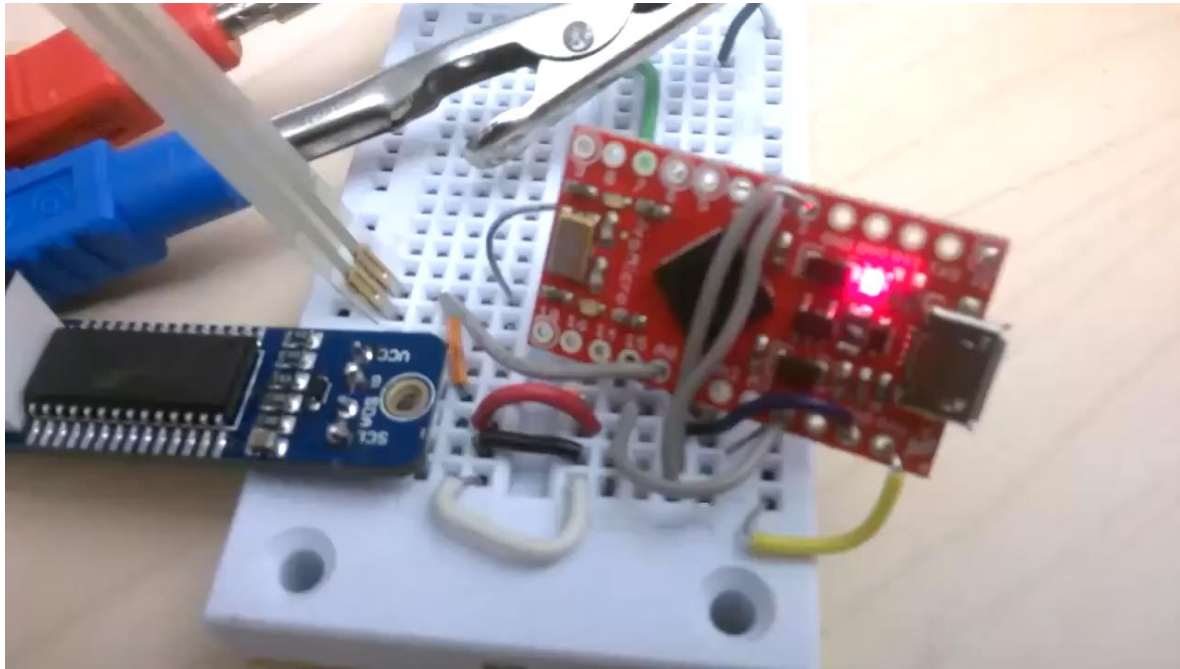
What Can We Do with Tactile Information?

- Visionless action
- Stable grasp
- Fine-grained or dynamic action
- Safe robot control
- ...

How to Equip Robots with Tactile Sensing?



Low-Cost Force-Sensing Resistor



Video source: Wikipedia

Our System
Real World Setup

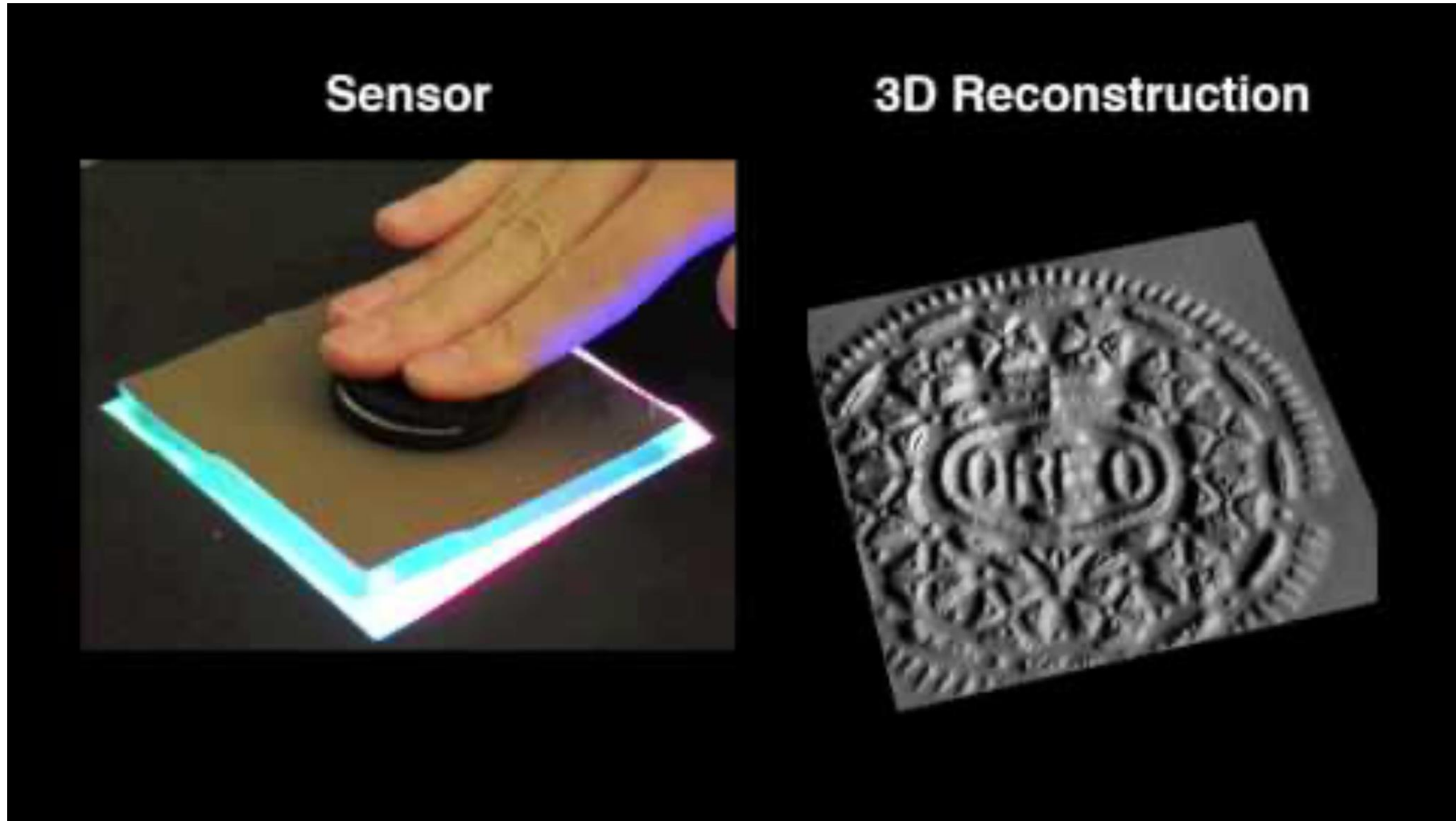
16 Low-cost (\$12 each)
Force Sensing Resistors



Low-Cost Force-Sensing Resistor



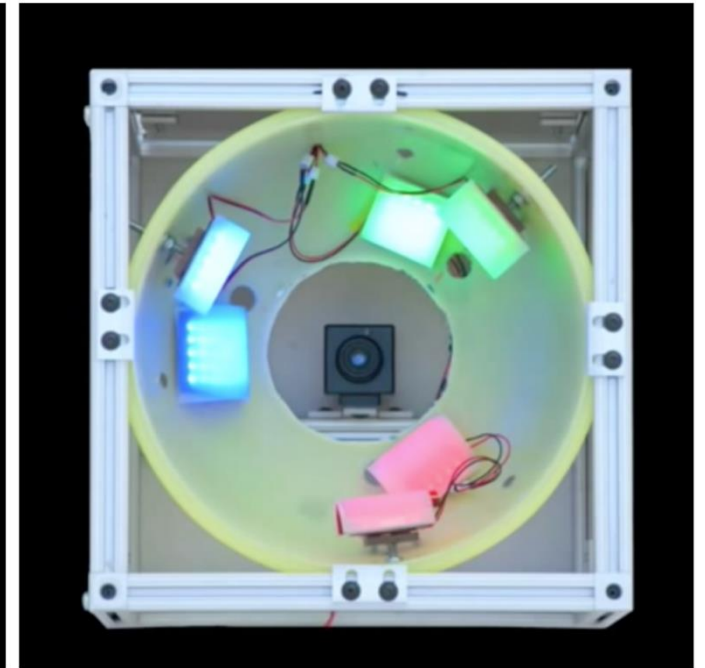
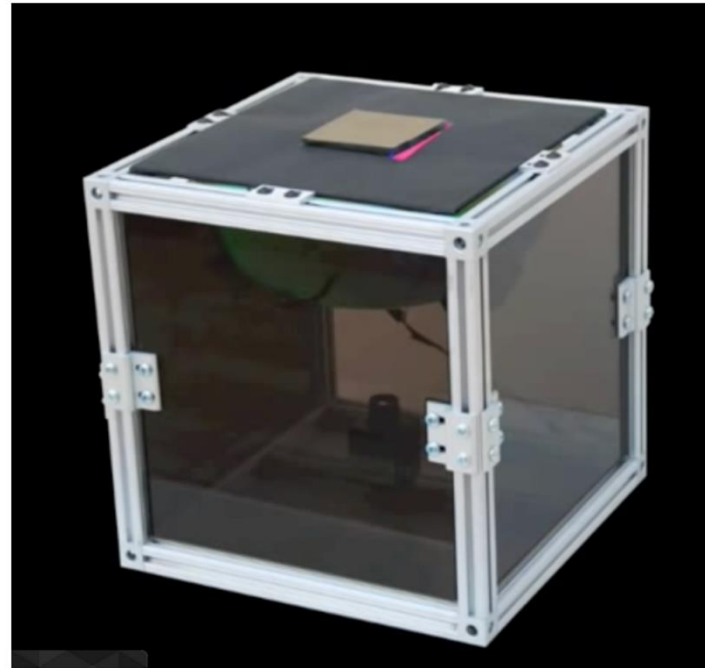
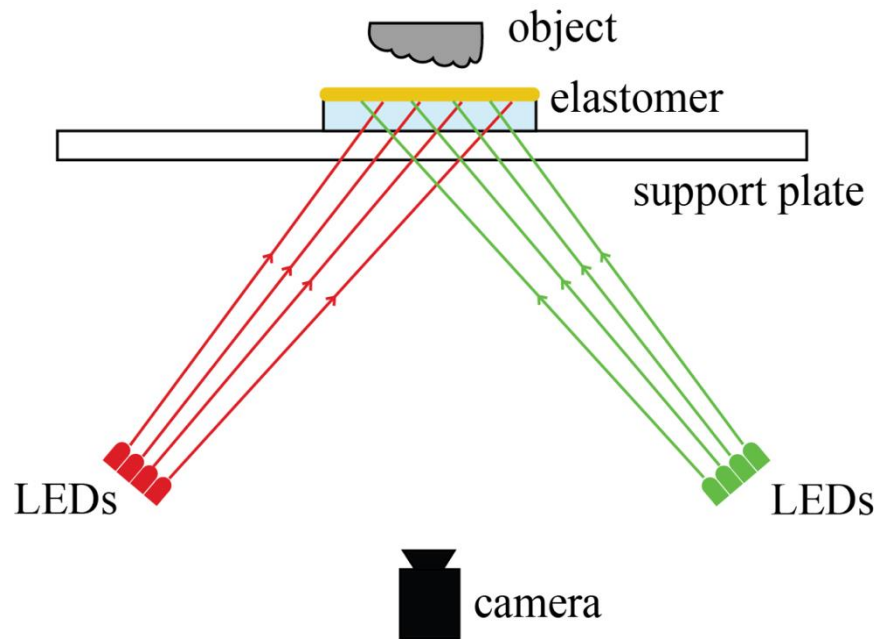
GelSight: Visuo-Tactile Sensor



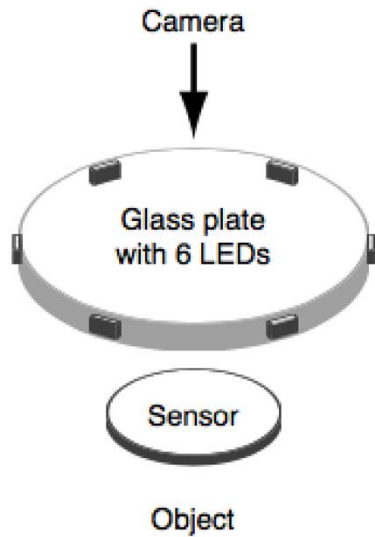
Visuo-Tactile Sensor is Putting a High-Quality Depth Sensor Near the Object Surface



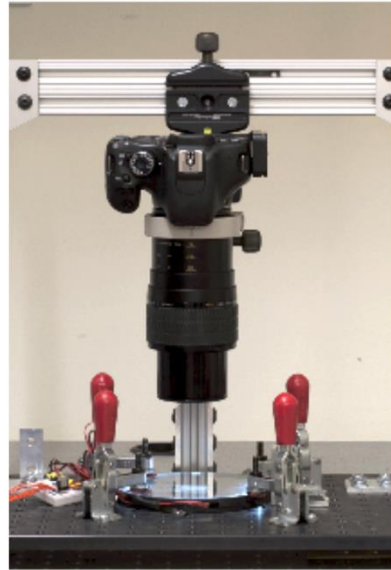
Visuo-Tactile Sensor is Putting a High-Quality Depth Sensor Near the Object Surface



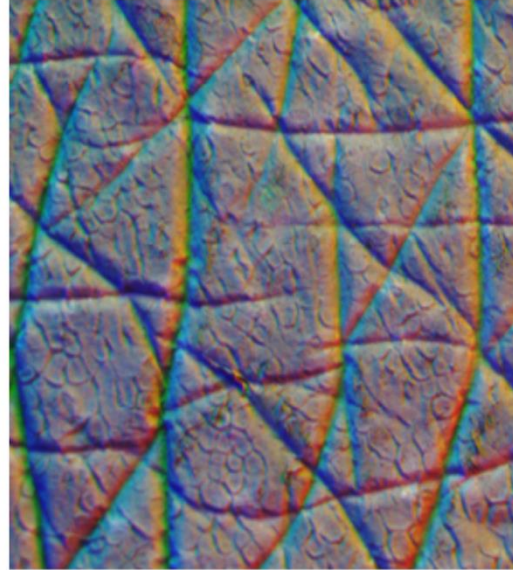
Visuo-Tactile Sensor Facilitates Reconstruction of Detailed Geometry



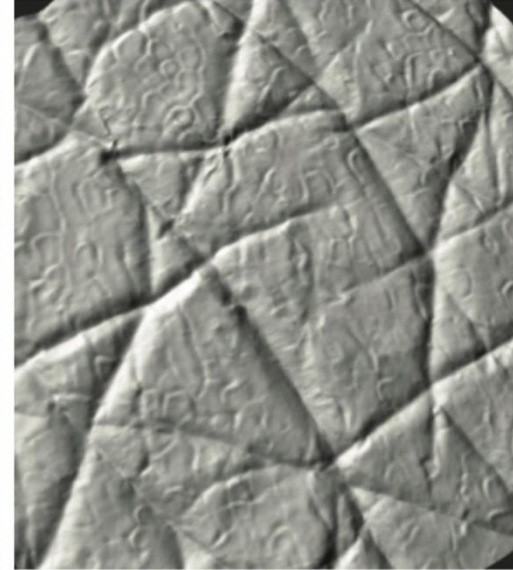
(a)



(b)



(c)

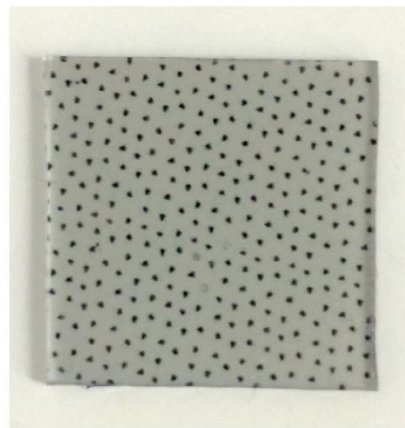


(d)

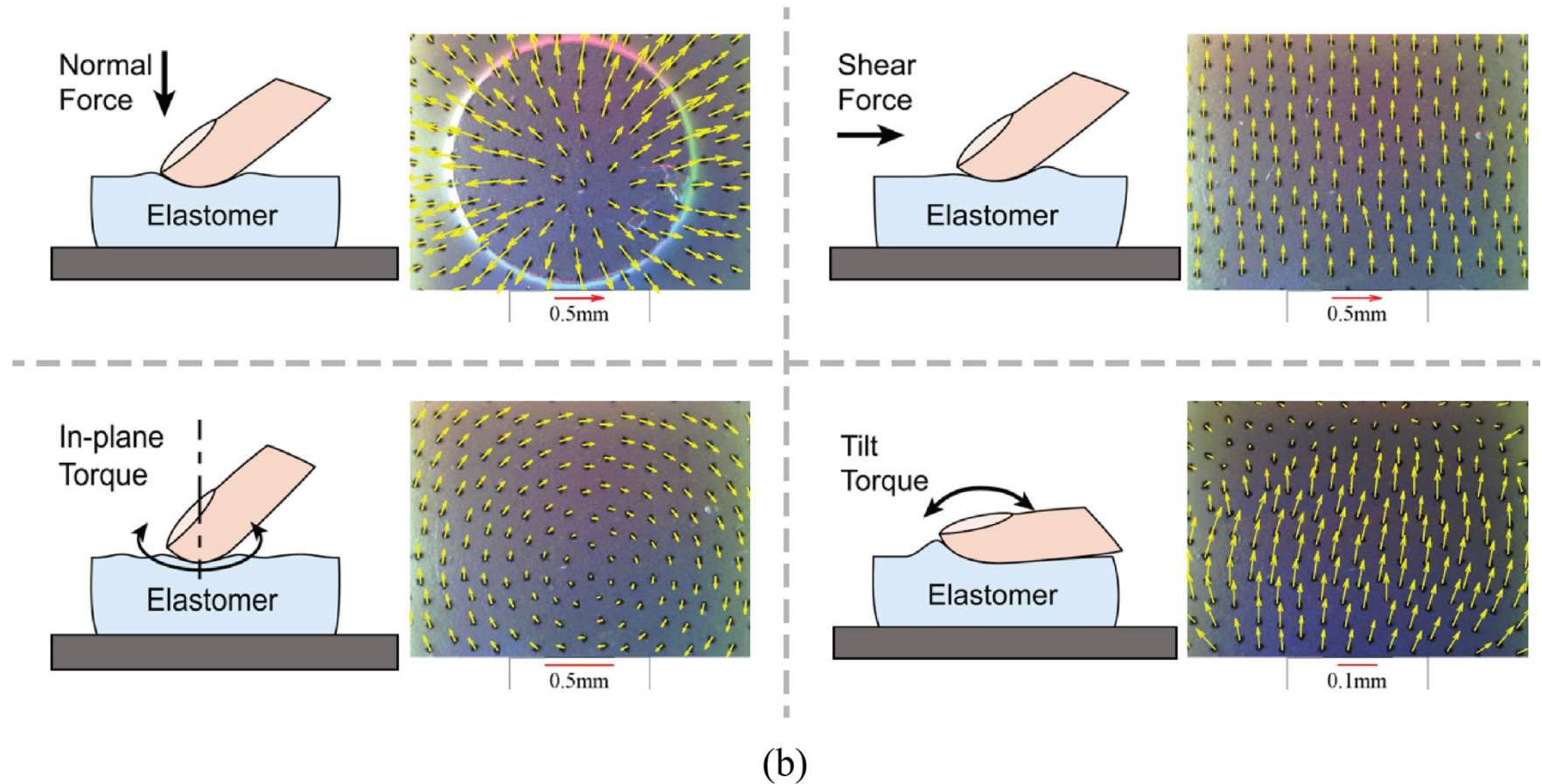


(e)

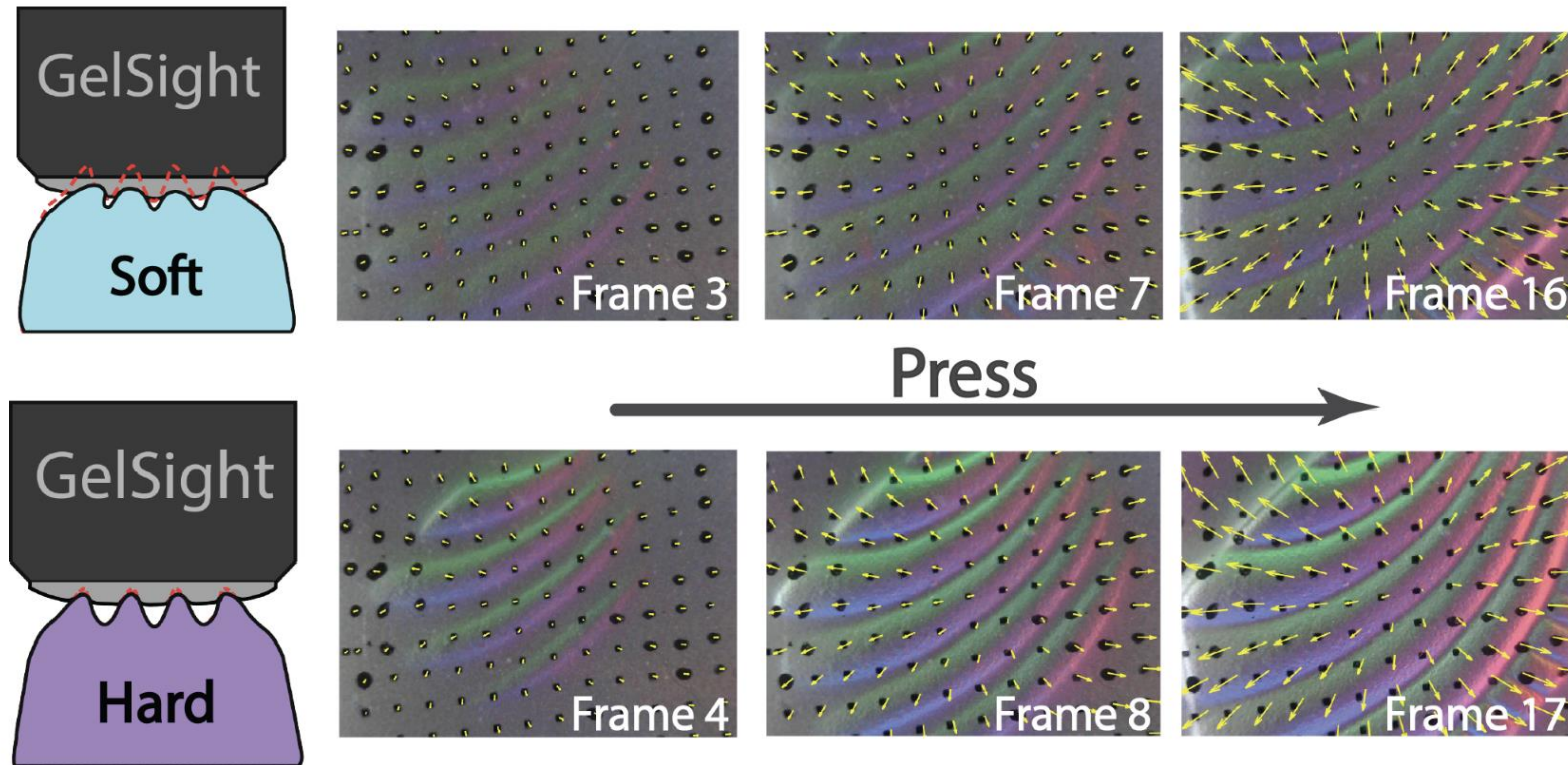
Visuo-Tactile Sensor Facilitates Sensing of Directional Forces



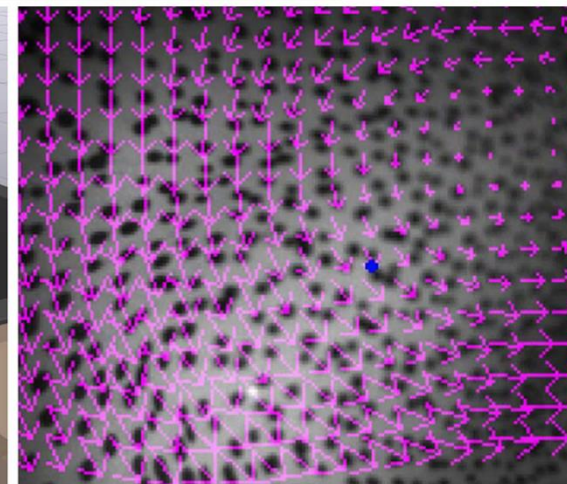
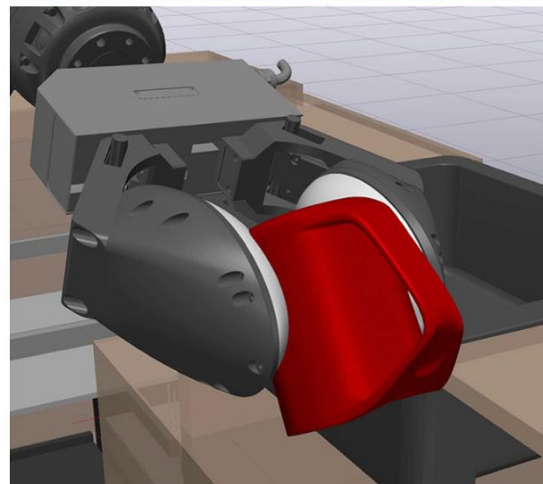
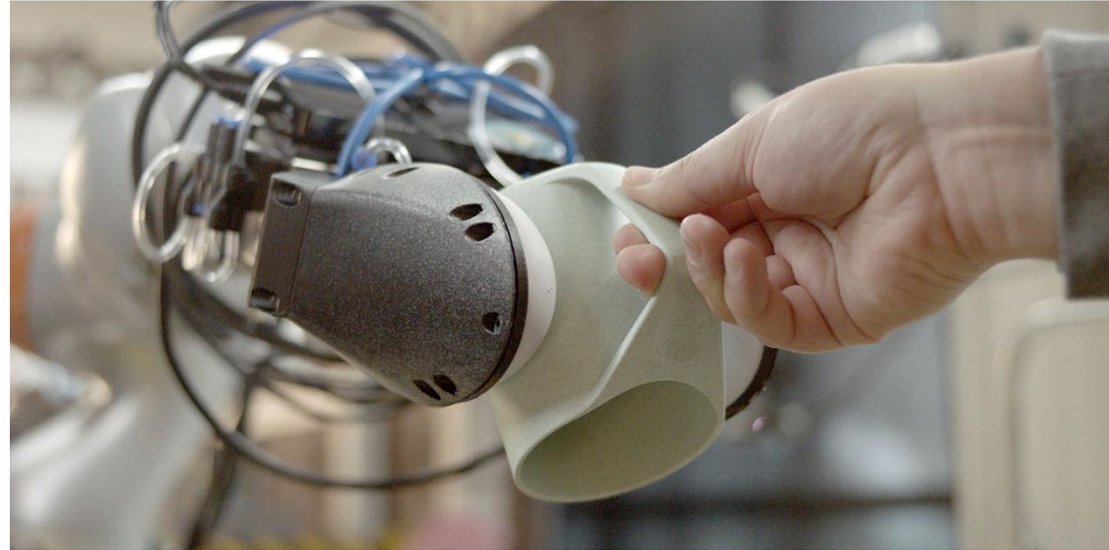
(a)



Visuo-Tactile Sensor Facilitates Understanding Physical Properties of Objects

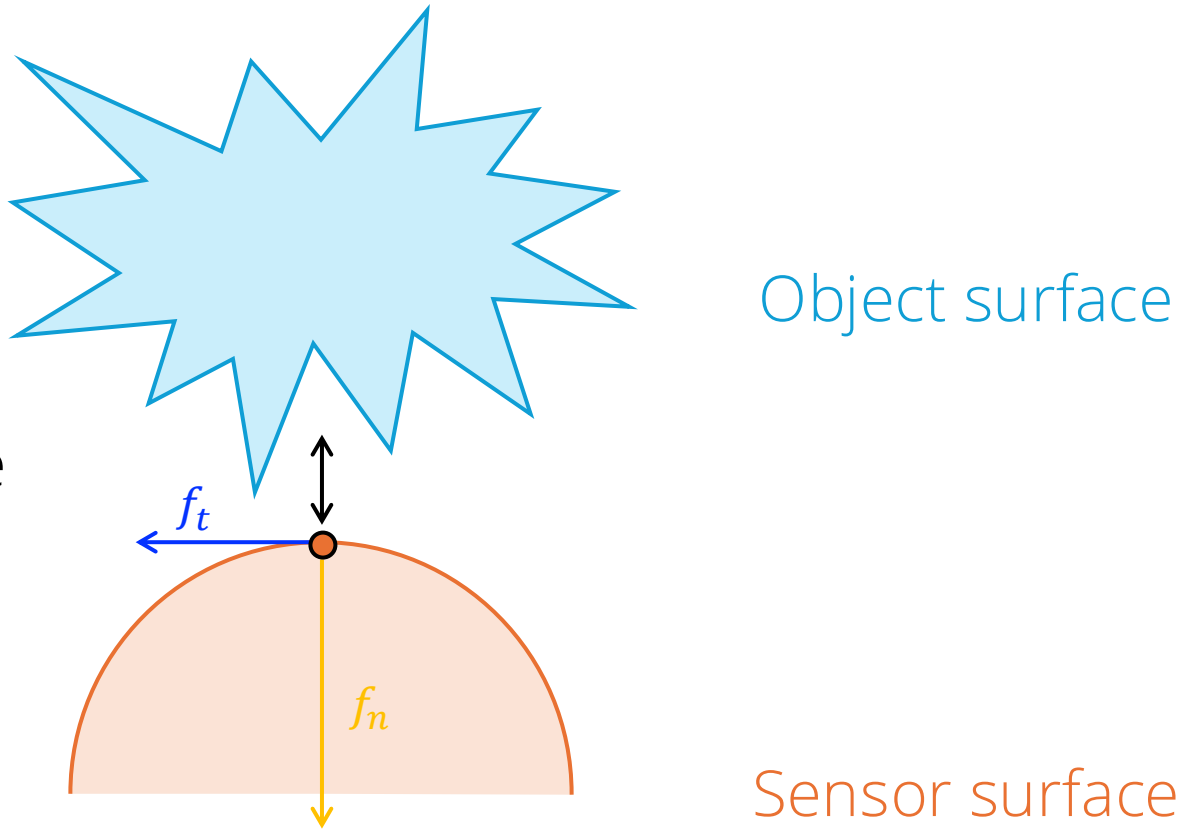


How to Simulate Tactile Sensing?



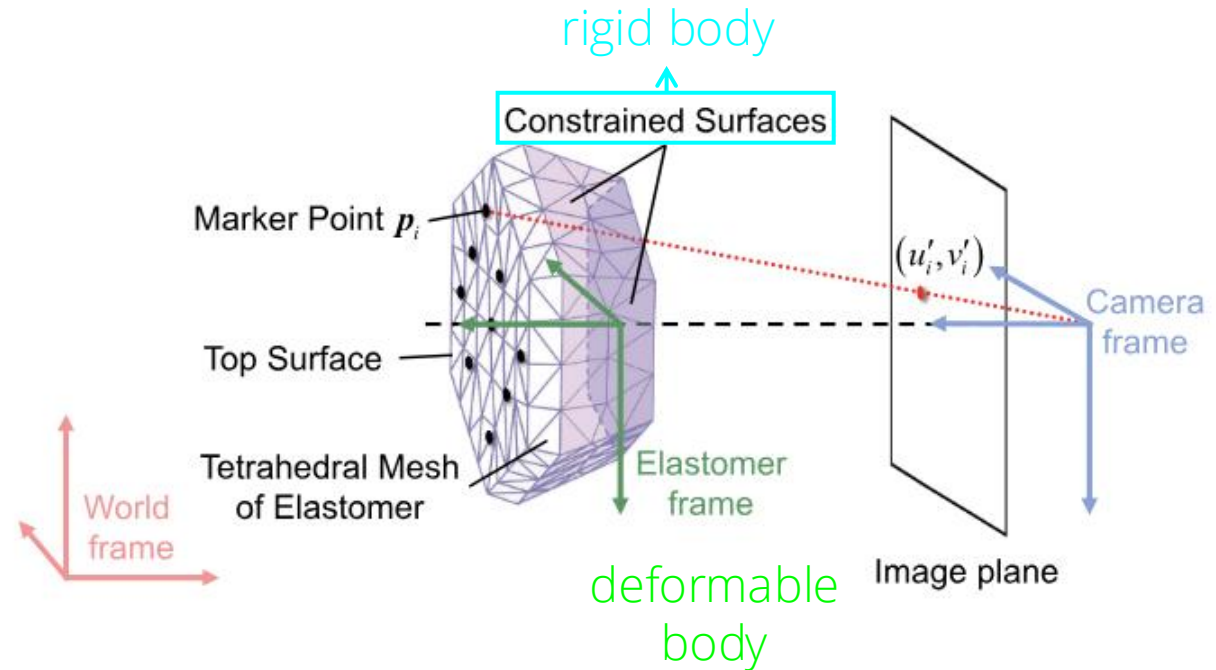
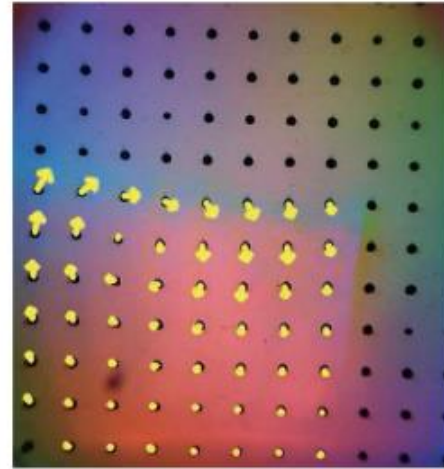
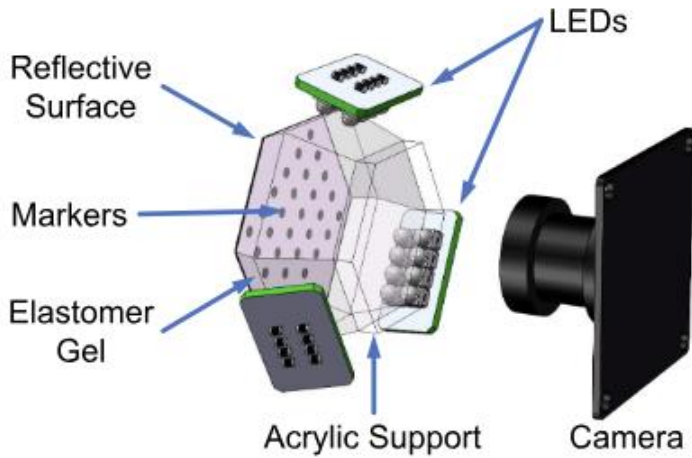
How to Simulate Tactile Sensing?

Tactile sensing requires precise frictional contact modeling at the object-sensor surface



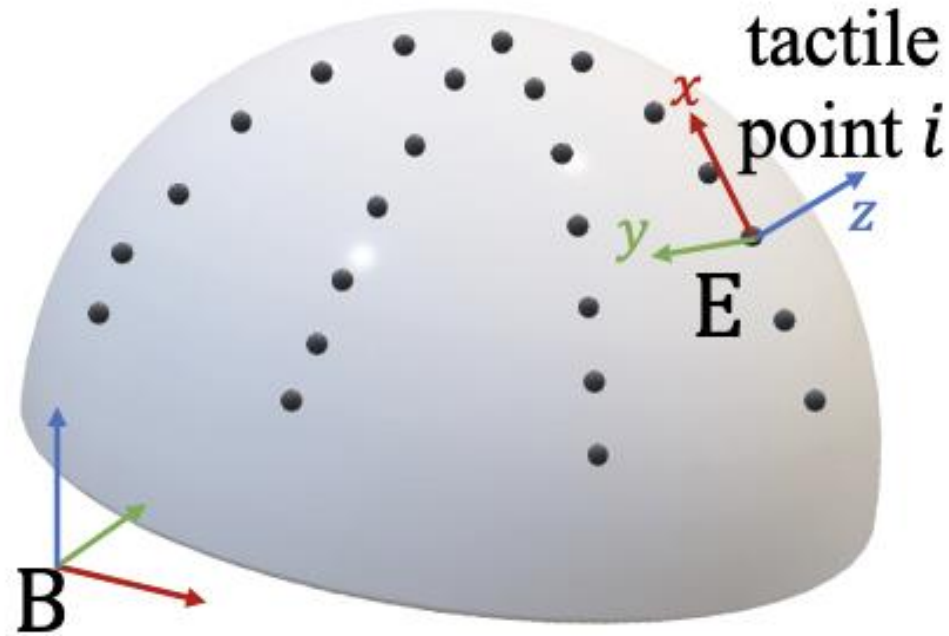
How to Simulate Tactile Sensing?

- A closer look at visuo-tactile sensors

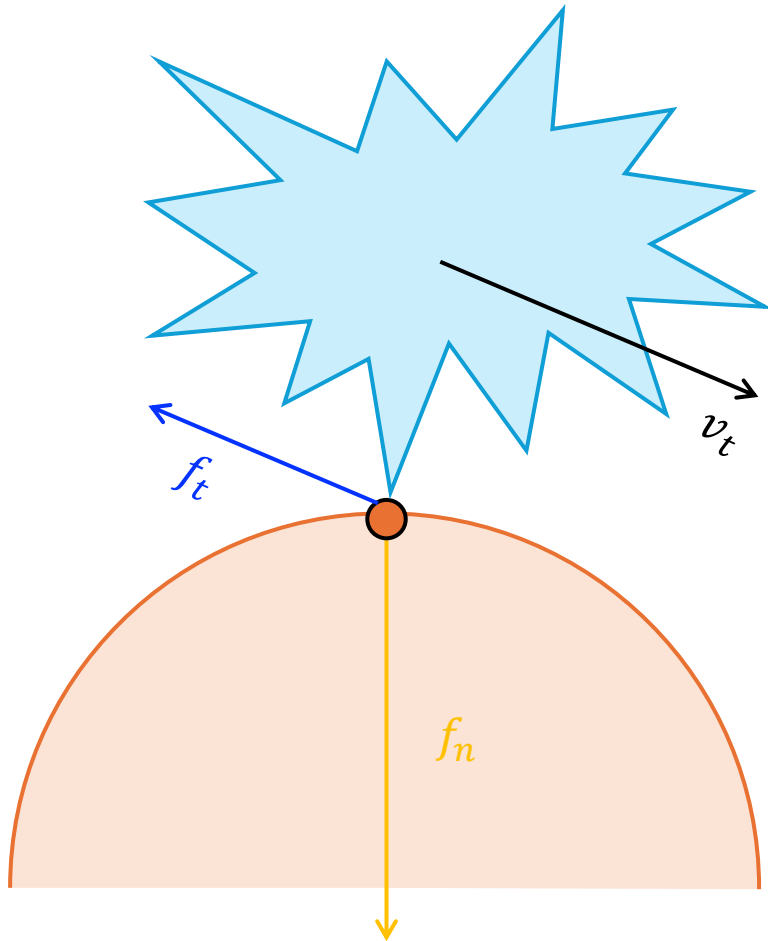


How to Simulate Tactile Sensing?

- A tactile point can be described as an attached particle on the elastomer frame. The base-to-particle transformation is E



How to Model the Motion of Tactile Points?

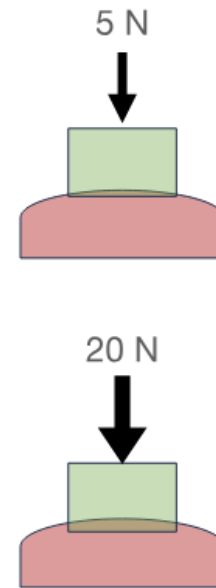
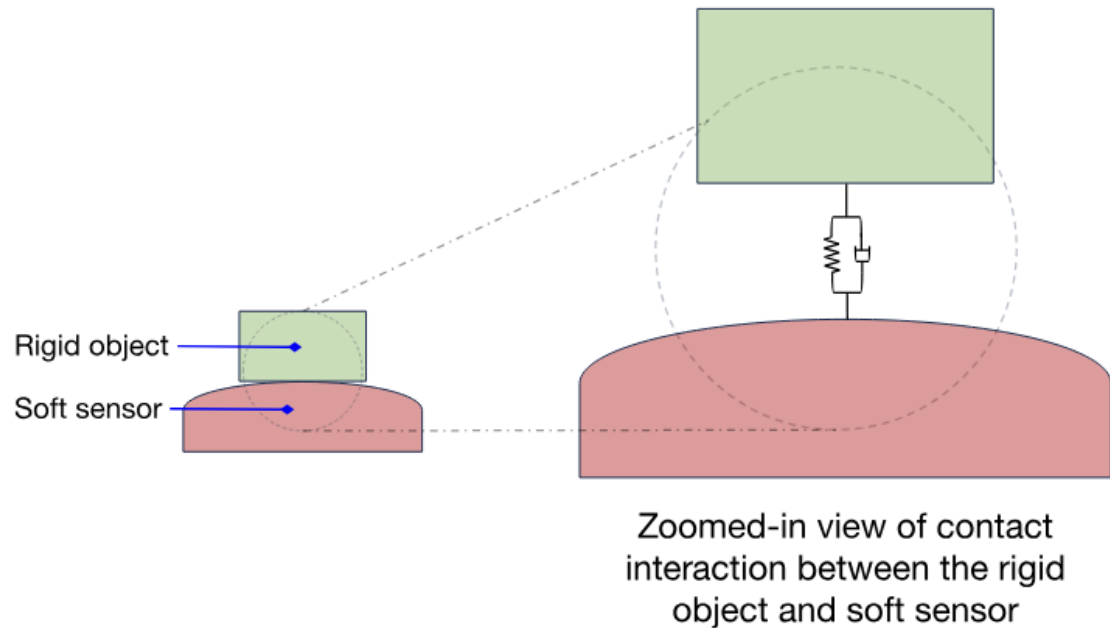


- The motion of a tactile point is determined by the shear force f_t and normal force f_n
- The shear force is induced by frictions, which is determined by relative motion and the normal force between two surfaces:

$$f_t = -\frac{v_t}{\|v_t\|} \min(k_t \|v_t\|, \mu \|f_n\|)$$

- How to calculate the normal force?

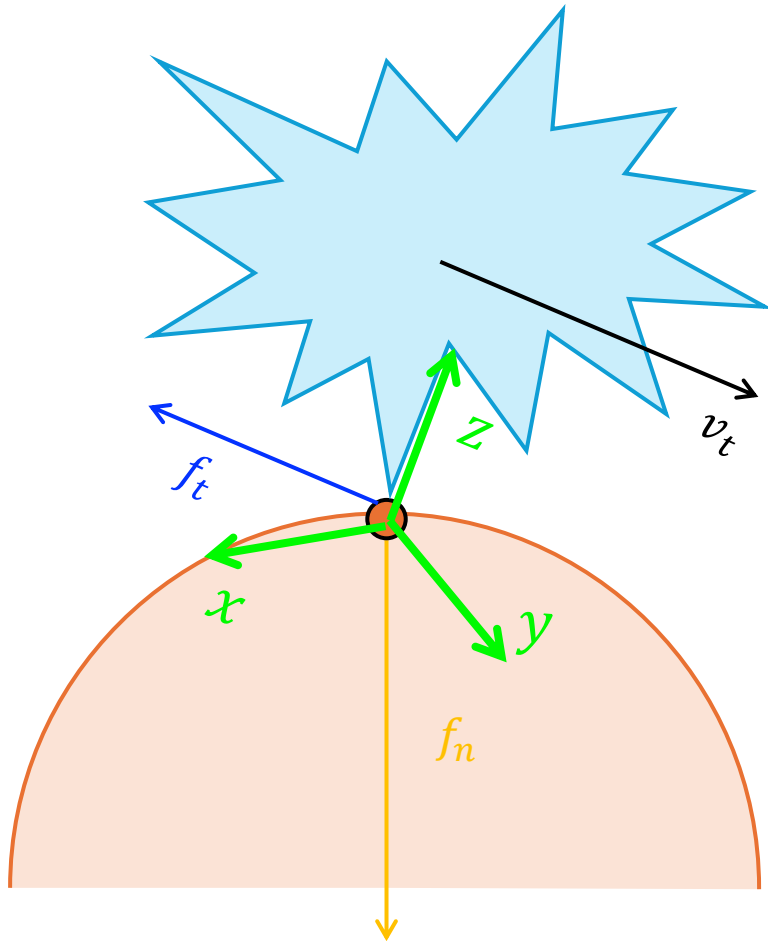
How to Model the Normal Force? Let's Ignore Impenetrability Constraints



- Larger penetration depth d leads to larger normal force
- Let \mathbf{n} denotes normal vector
- Normal force:

$$\mathbf{f}_n = (-k_n + k_d \dot{d}) d \mathbf{n}$$

How to Model the Motion of Tactile Points?



- Normal force:

$$f_n = (-k_n + k_d \dot{d}) d \mathbf{n}$$

- The shear force:

$$f_t = -\frac{v_t}{\|v_t\|} \min(k_t \|v_t\|, \mu \|f_n\|)$$

- Total force:

$$f = f_n + f_t$$

- Local shear and normal force at the tactile point:

$$T_{sx} = f^\top x \quad T_{sy} = f^\top y \quad T_{sz} = f^\top z$$

How to Simulate the Whole Dynamic System?

- Iterative simulation:
 1. Move the robot's end-effector
 2. Detect collision and calculate the shear and normal force on the tactile sensor
 3. Update the sensor observation
 4. Update the dynamic of the robot and objects

How to Simulate the Whole Dynamic System?

- The optimization problem:

$$\underset{\theta}{\text{minimize}} \quad \mathcal{L} = \sum_{t=1}^H \mathcal{L}_t(\mathbf{u}_t, \mathbf{q}_t, \mathbf{v}_t(\mathbf{q}_t))$$

$$\text{s.t.} \quad g(\mathbf{q}_{t-1}, \dot{\mathbf{q}}_{t-1}, \mathbf{u}_t, \mathbf{q}_t) = 0$$

(Equations of Motion)

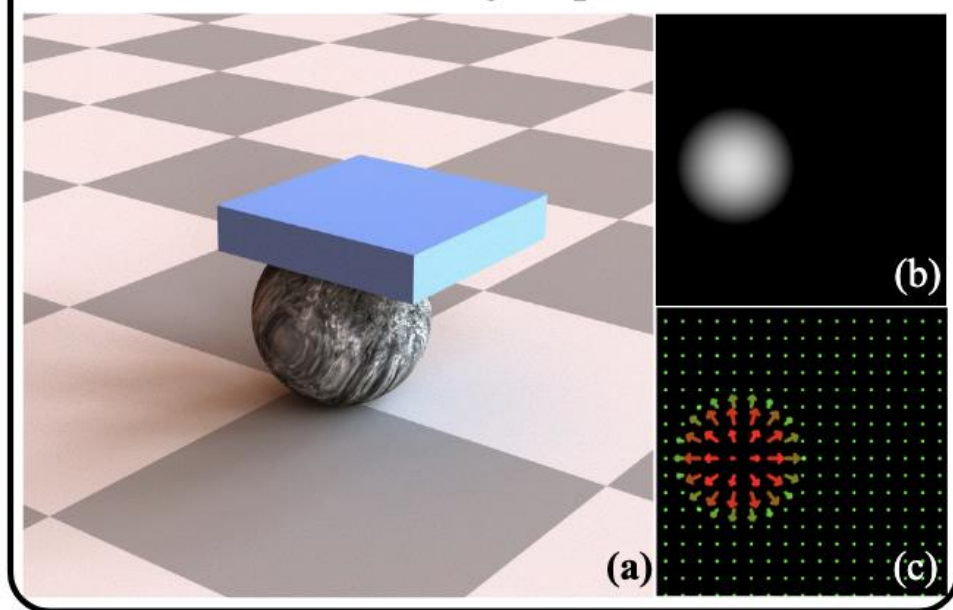
$$\mathbf{u}_t = \pi_{\theta}(\tilde{\mathbf{q}}_{t-1}, \tilde{\mathbf{v}}_{t-1}(\mathbf{q}_{t-1}), T_{t-1}(\mathbf{q}_{t-1}, \dot{\mathbf{q}}_{t-1})).$$

(Policy Execution)

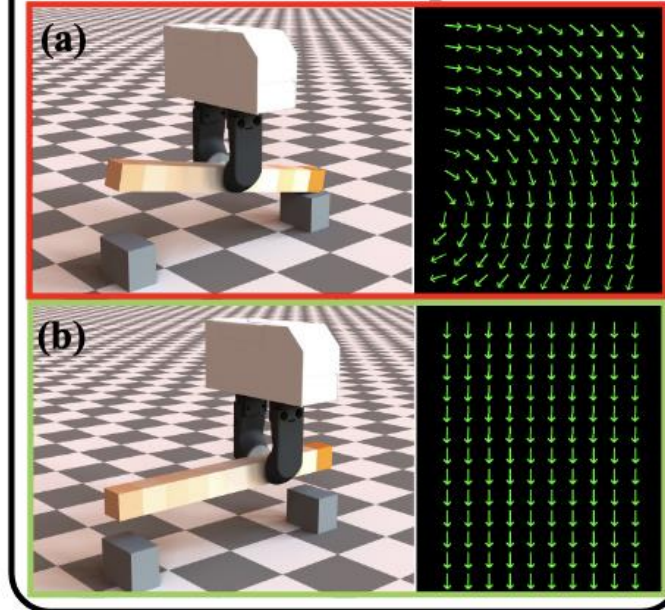
where H denotes the task horizon, \mathcal{L}_t is a step-wise task-dependent cost function, \mathbf{u} is the action, \mathbf{q} is the simulation state and \mathbf{v} is the derived auxiliary simulation variables

Results

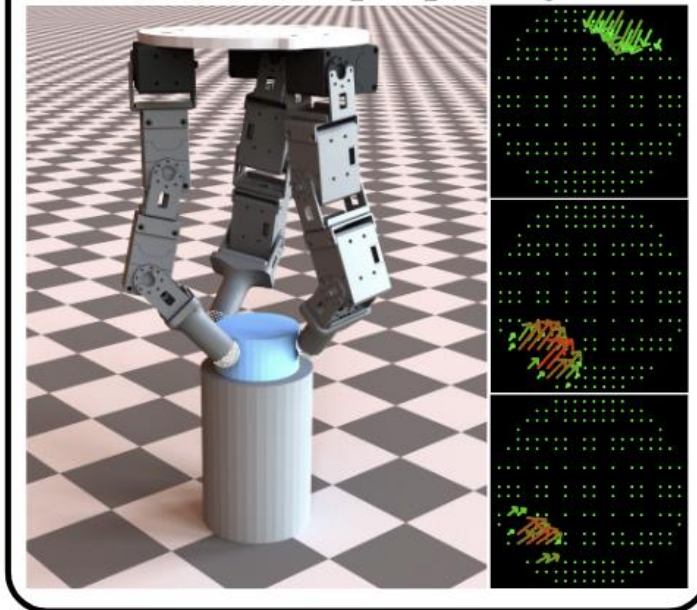
I. Ball Rolling Experiment



II. Stable Grasp Task



III. D'Claw Cap Opening Task





Efficient Tactile Simulation with Differentiability for Robotic Manipulation

Jie Xu[†], Sangwoon Kim[†], Tao Chen[†], Alberto Rodriguez[†], Pulkit Agrawal[†],
Wojciech Matusik[†], Shinjiro Sueda[‡]

[†]Massachusetts Institute of Technology

[‡]Texas A&M University

6th Conference on Robot Learning (CoRL 2022)



However, This is just One Way of Tactile Simulation

TacSL: A Library for Visuotactile Sensor Simulation and Learning

Iretiayo Akinola^{1*}, Jie Xu^{1*}, Jan Carius¹, Dieter Fox^{1,2}, and Yashraj Narang¹

TacEx: GelSight Tactile Simulation in Isaac Sim – Combining Soft-Body and Visuotactile Simulators

Duc Huy Nguyen^{1,3}, Tim Schneider^{1,2}, Guillaume Duret^{1,2},
Alap Kshirsagar¹, Boris Belousov³, Jan Peters^{1,3,4}

¹TU Darmstadt ²École centrale de Lyon ³DFKI ⁴Hessian.AI
duc_huy.nguyen@dfki.de

IPC-GraspSim: Reducing the Sim2Real Gap for Parallel-Jaw Grasping with the Incremental Potential Contact Model

Chung Min Kim¹, Michael Danielczuk¹, Isabella Huang², Ken Goldberg¹

DIFFTACTILE: A PHYSICS-BASED DIFFERENTIABLE TACTILE SIMULATOR FOR CONTACT-RICH ROBOTIC MANIPULATION

Zilin Si^{*,†,1,5}, Gu Zhang^{*,2}, Qingwei Ben^{*,3}, Branden Romero⁴, Zhou Xian¹,
Chao Liu⁴, Chuang Gan^{5,6}

¹CMU RI, ²Shanghai Jiao Tong University, ³Tsinghua University, ⁴MIT CSAIL,
⁵MIT-IBM Watson AI Lab, ⁶UMass Amherst
zsi@andrew.cmu.edu, guzhang.sjtu@gmail.com,
bqw20@mails.tsinghua.edu.cn, brromero@mit.edu,
xianz1@andrew.cmu.edu, chaoliu@csail.mit.edu,
ganchuang@csail.mit.edu

Tactile Gym 2.0: Sim-to-real Deep Reinforcement Learning for Comparing Low-cost High-Resolution Robot Touch

Yijiong Lin¹, John Lloyd¹, Alex Church¹, Nathan F. Lepora¹
Project Webpage: <https://sites.google.com/my.bristol.ac.uk/tactilegym2>

There's no perfect solution yet! Tactile simulation is still an underexplored problem

Why Do We Need Tactile Sensing?

- Robotics: Contact-rich manipulation
- AR / VR: Immersive experience with physical feedback

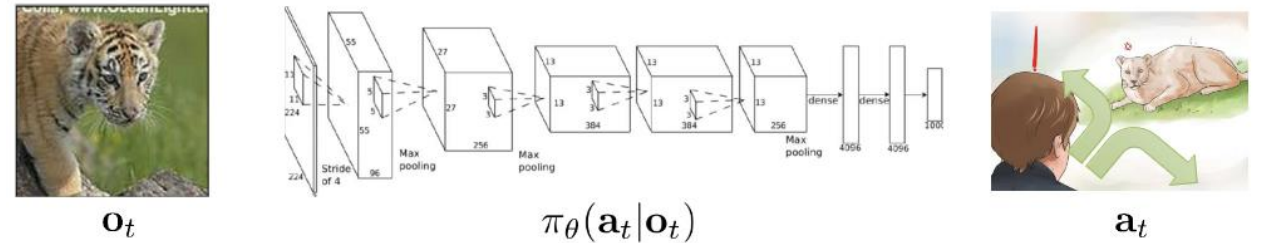
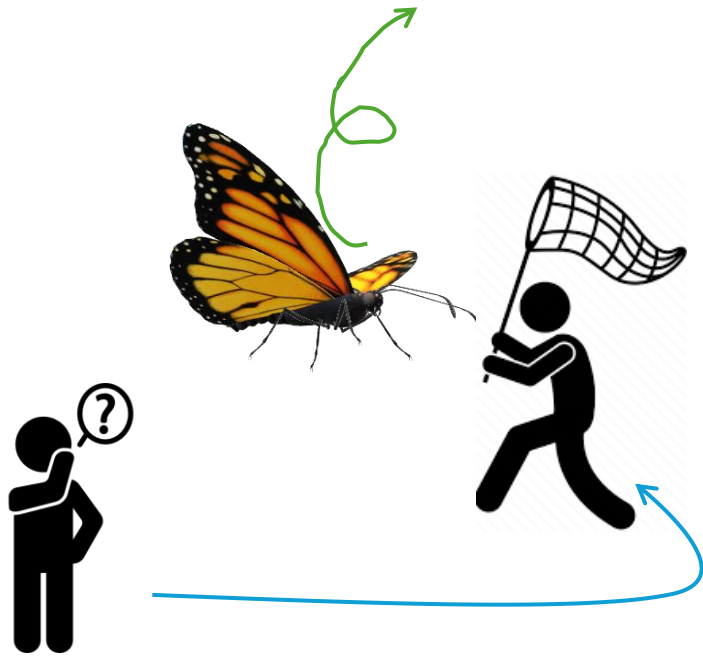


Image source: OpenAI



Image credit Davide Scaramuzza

A Brief Introduction of Robotics



\mathbf{s}_t – state

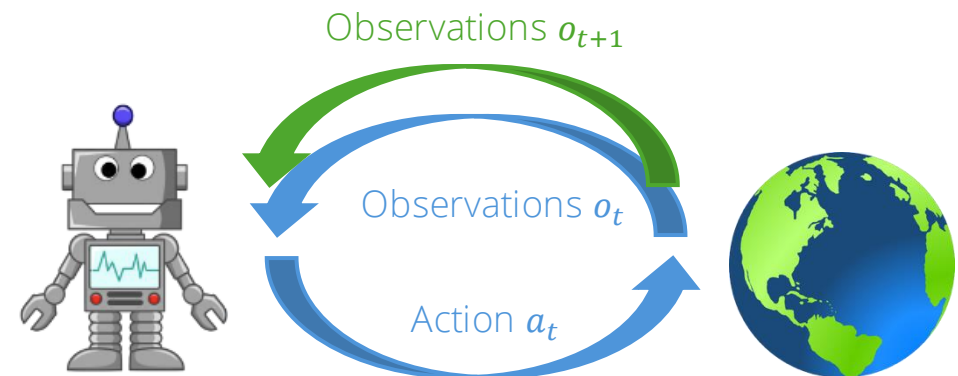
\mathbf{o}_t – observation

\mathbf{a}_t – action

$\pi_{\theta}(\mathbf{a}_t | \mathbf{o}_t)$ – policy

$\pi_{\theta}(\mathbf{a}_t | \mathbf{s}_t)$ – policy (fully observed)

We'll cover more next week!



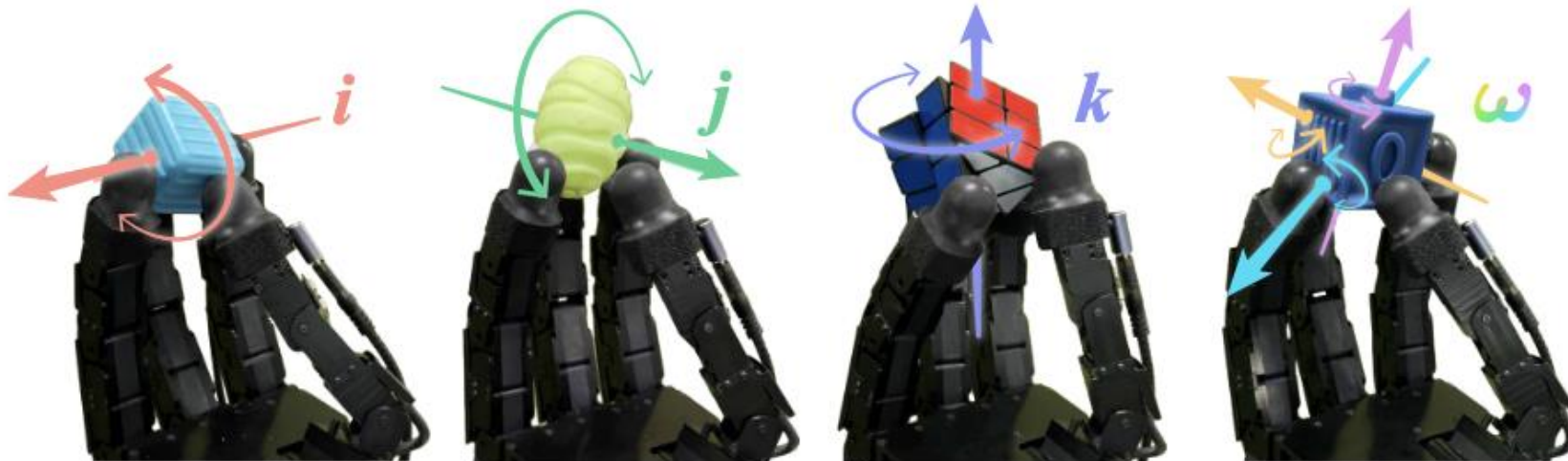
Tactile Sensing Facilitate Robotics

General In-Hand Object Rotation with Vision and Touch

Haozhi Qi^{1,2} Brent Yi¹ Sudharshan Suresh^{2,3}
Mike Lambeta² Yi Ma¹ Roberto Calandra^{4,5} Jitendra Malik^{1,2}

¹UC Berkeley ²Meta AI ³CMU ⁴TU Dresden
⁵ The Centre for Tactile Internet with Human-in-the-Loop (CeTI)

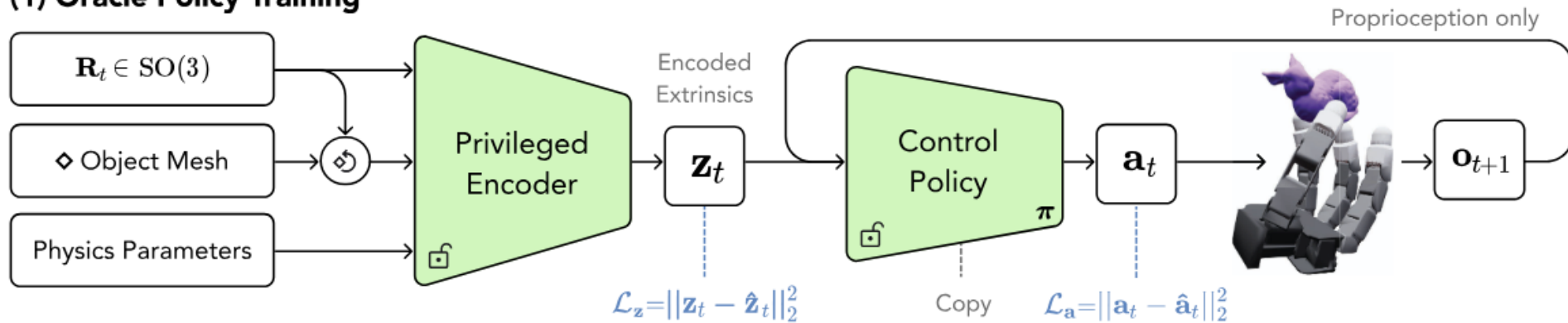
<https://haozhi.io/rotateit/>



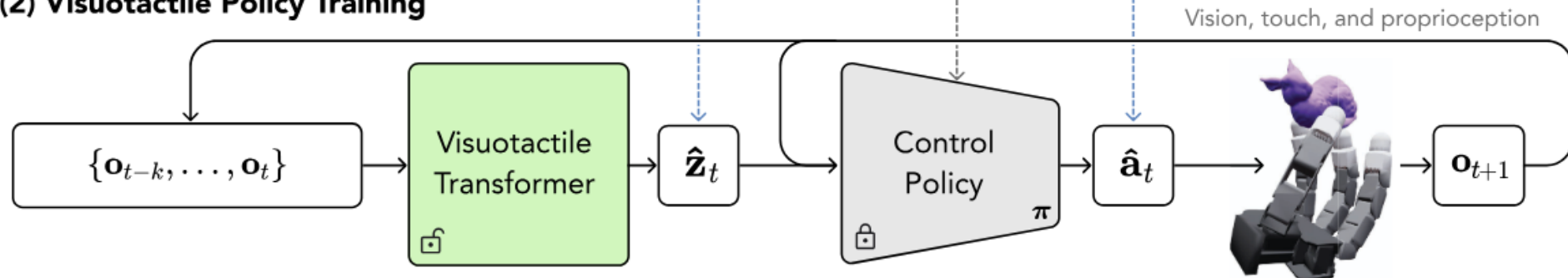
Problem 1: Generalization to Objects of Different Shapes and Properties

- Solution: Learning for Implicit System Identification

(1) Oracle Policy Training



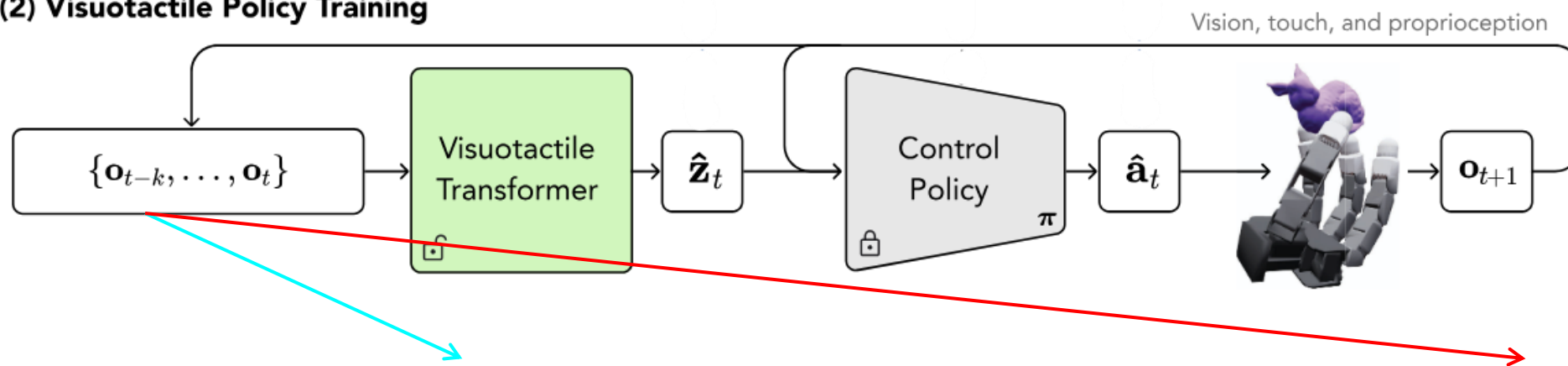
(2) Visuotactile Policy Training



Problem 2: Does Tactile Sensing Help?

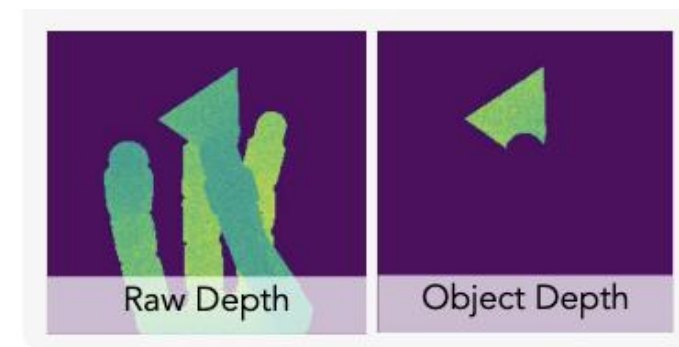
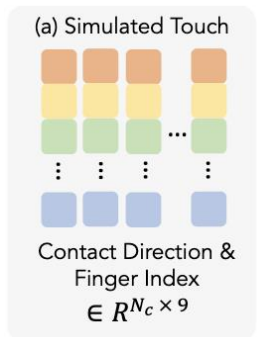
- Solution: A transformer for multi-modal sensor fusion

(2) Visuotactile Policy Training

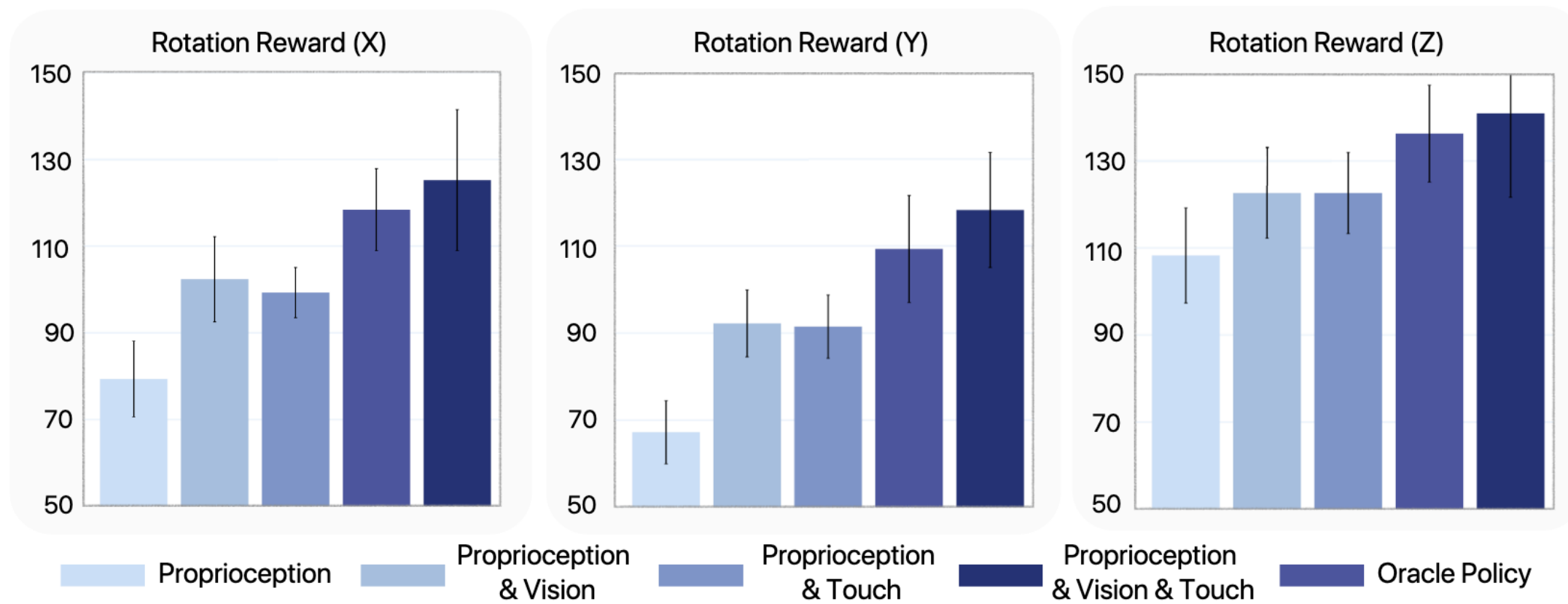


- Simplified tactile observation

- Depth observation



Results: Multi-modal Sensory Input is Better



Results: Fine-grained Tactile Sensing is Better

Touch	<i>x</i> -axis			<i>y</i> -axis			<i>z</i> -axis		
	RotR \uparrow	TTF \uparrow	RotP \downarrow	RotR \uparrow	TTF \uparrow	RotP \downarrow	RotR \uparrow	TTF \uparrow	RotP \downarrow
Full	104.29 \pm 10.29	0.68 \pm 0.04	0.41 \pm 0.02	93.05 \pm 9.28	0.65 \pm 0.01	0.34 \pm 0.03	126.73 \pm 10.11	0.72 \pm 0.03	0.32 \pm 0.03
NoTouch	79.37 \pm 8.72	0.46 \pm 0.03	0.55 \pm 0.02	67.21 \pm 7.25	0.48 \pm 0.02	0.55 \pm 0.03	108.25 \pm 10.92	0.62 \pm 0.01	0.43 \pm 0.02
Binary	80.14 \pm 7.25	0.47 \pm 0.02	0.53 \pm 0.03	66.29 \pm 8.53	0.49 \pm 0.01	0.56 \pm 0.04	110.24 \pm 9.48	0.63 \pm 0.03	0.42 \pm 0.02
ContactLoc	102.36 \pm 9.82	0.65 \pm 0.04	0.41 \pm 0.04	92.22 \pm 7.69	0.64 \pm 0.01	0.36 \pm 0.03	122.60 \pm 10.39	0.73 \pm 0.02	0.35 \pm 0.01

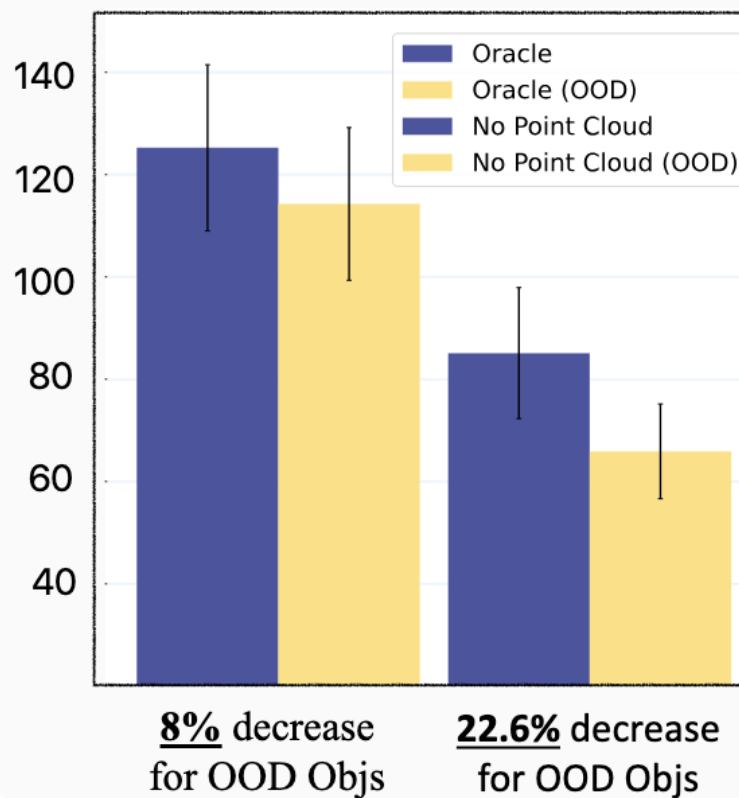
Table 2: **The importance of using a finer tactile information.** We compare *RotateIt* which use contact location (ContactLoc) and its variant of using binary contact (Binary) or full contact (position, normal, and scale) information. All methods are without vision information. Binary contact does not provide additional value compared to NoTouch, since it is already contained in our proprioceptive history. We also find using discretized contact locations can match the performance of using full contact in our task.

Results: Generalization to Unseen Objects

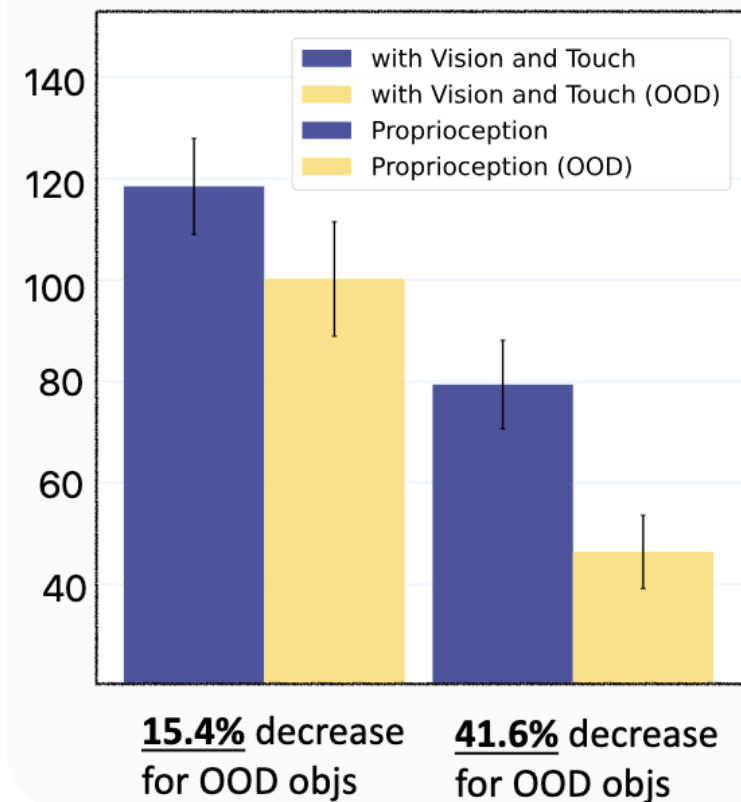
(a) OOD Objects



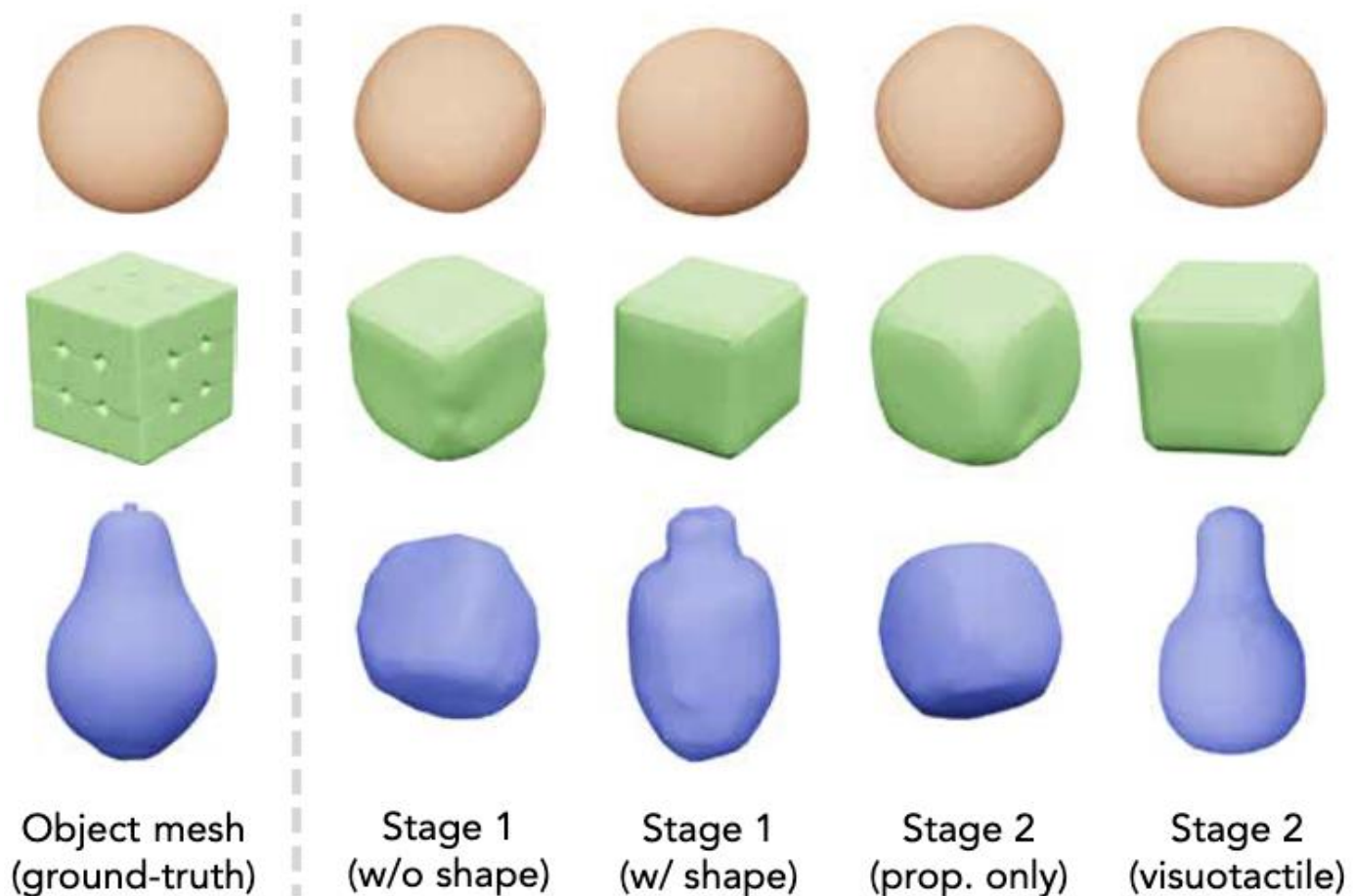
(b) Oracle Policy Comparison
With Point Cloud Without Point Cloud



(c) Sensorimotor Policy Comparison
Visuotactile Policy Proprioception Only



Results: Latent Features Encode Object Geometry

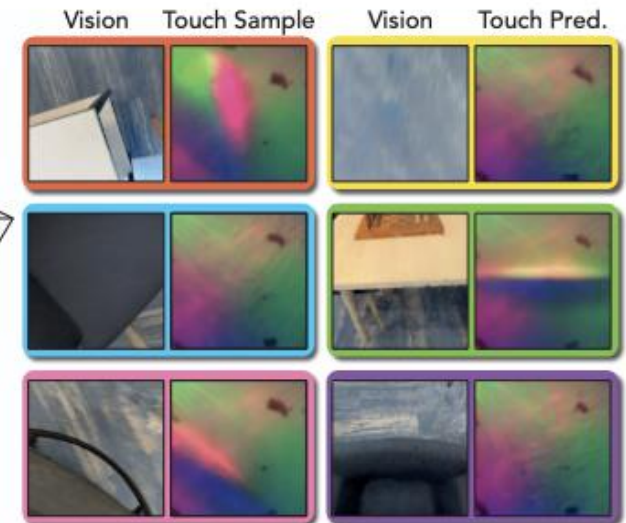


Tactile Sensing Facilitate AR / VR

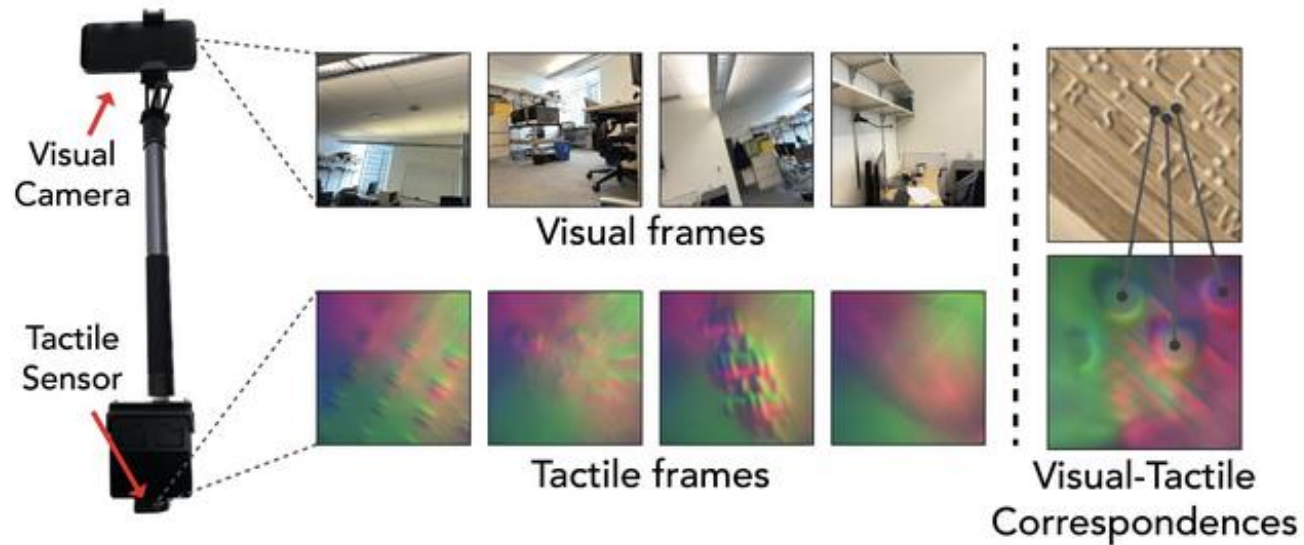
Tactile-Augmented Radiance Fields

Yiming Dou¹ Fengyu Yang² Yi Liu¹ Antonio Loquercio³ Andrew Owens¹

¹University of Michigan ²Yale University ³UC Berkeley

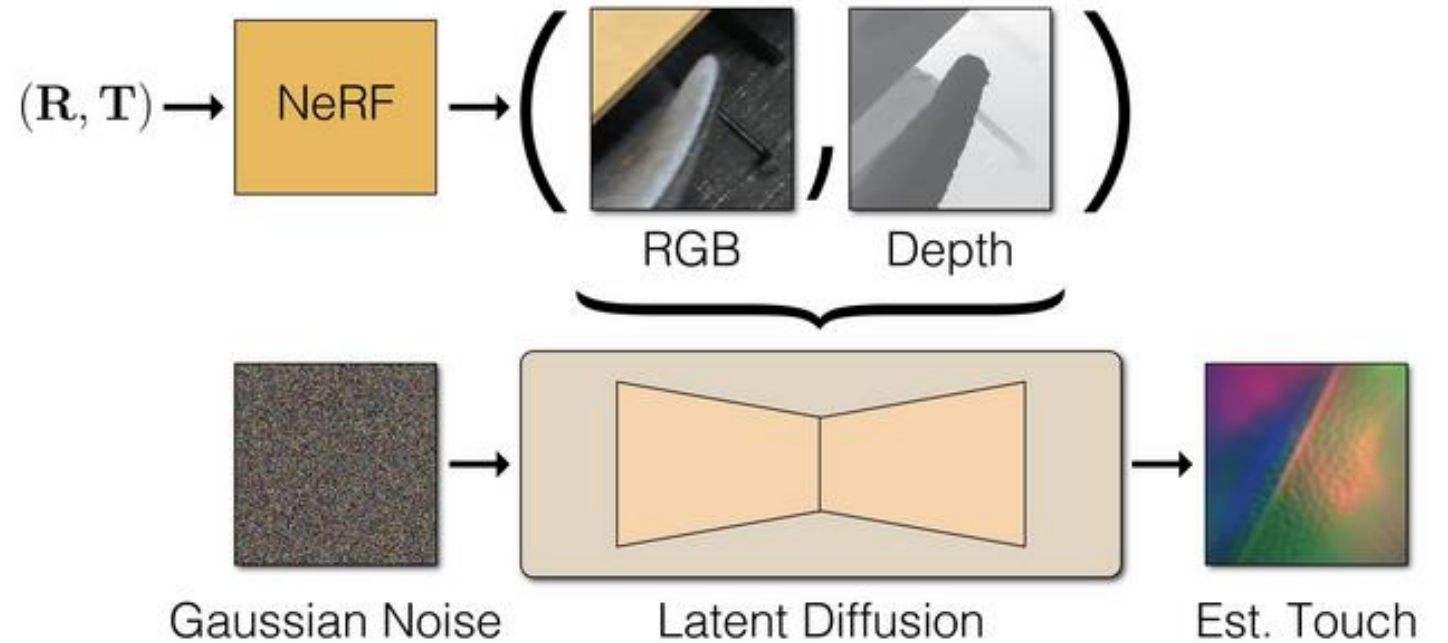
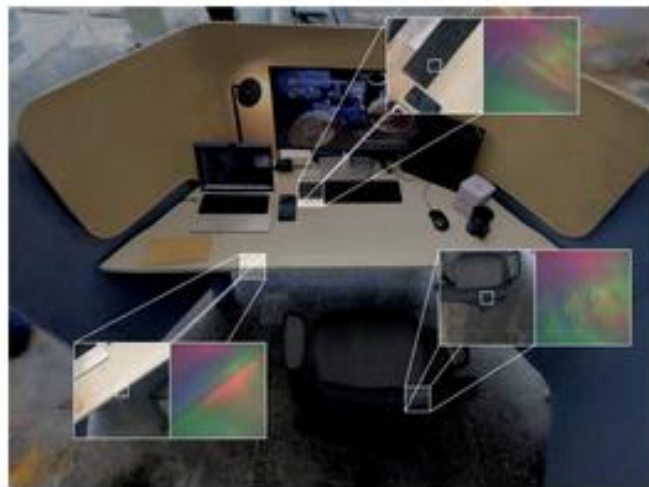


Joint Reconstruction of Geometry and Tactile Sensing



Learning to Impaint Tactile Sensing

- We can extrapolate unseen views by rendering. What about reconstruction tactile sensing of untouched scenes?



3D + Haptic Modeling of the “True” Real World



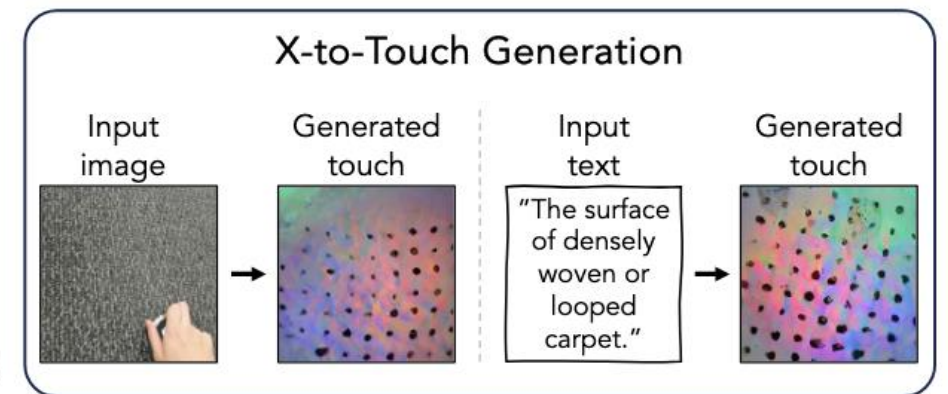
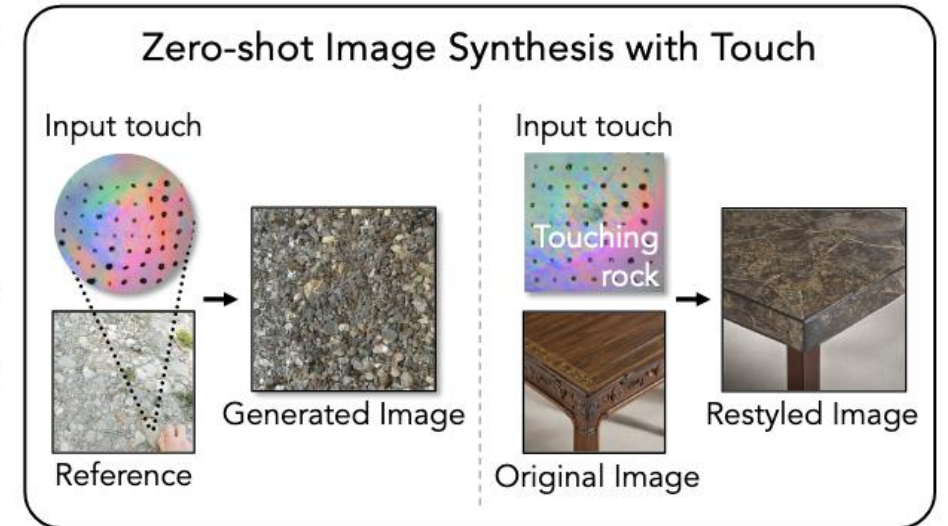
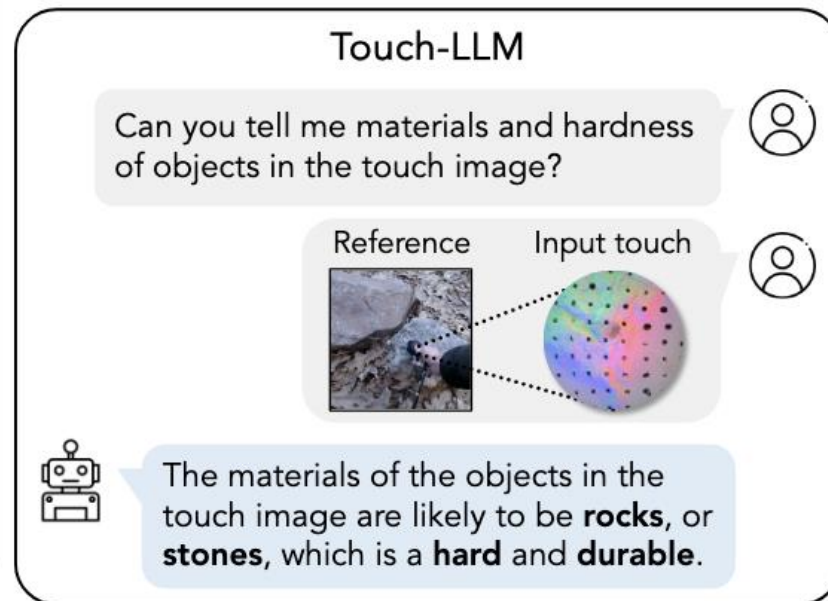
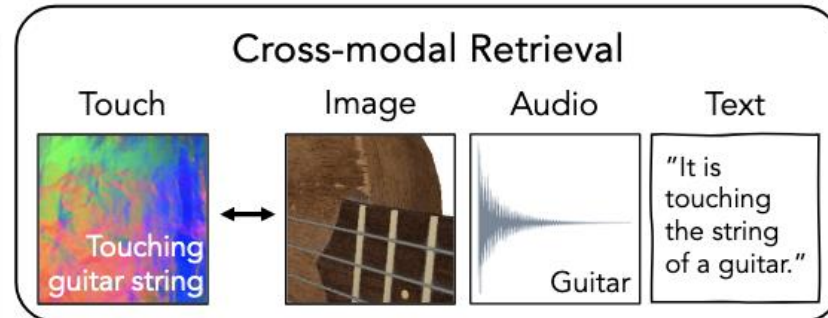
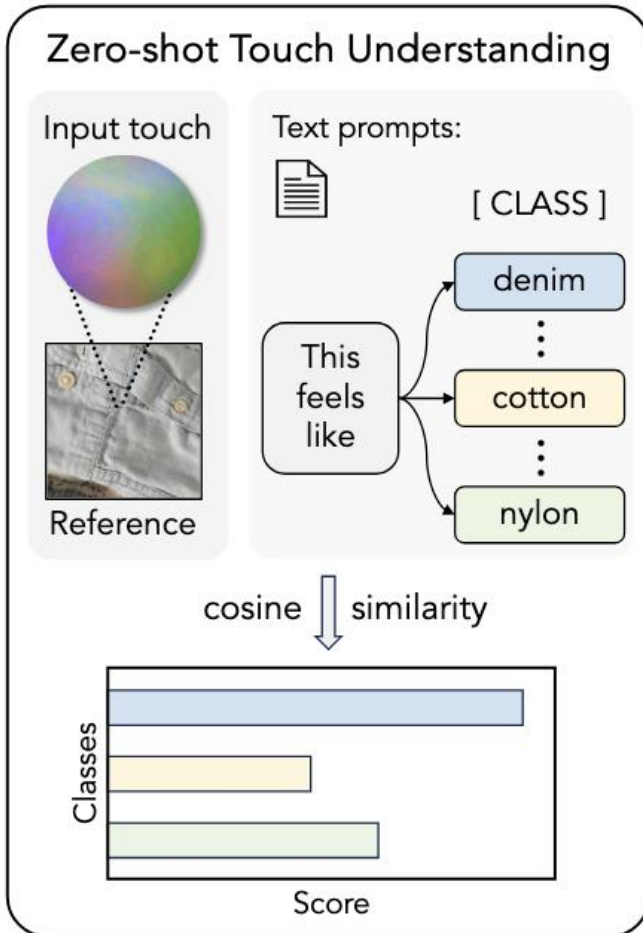
Tactile Sensing and VLMs

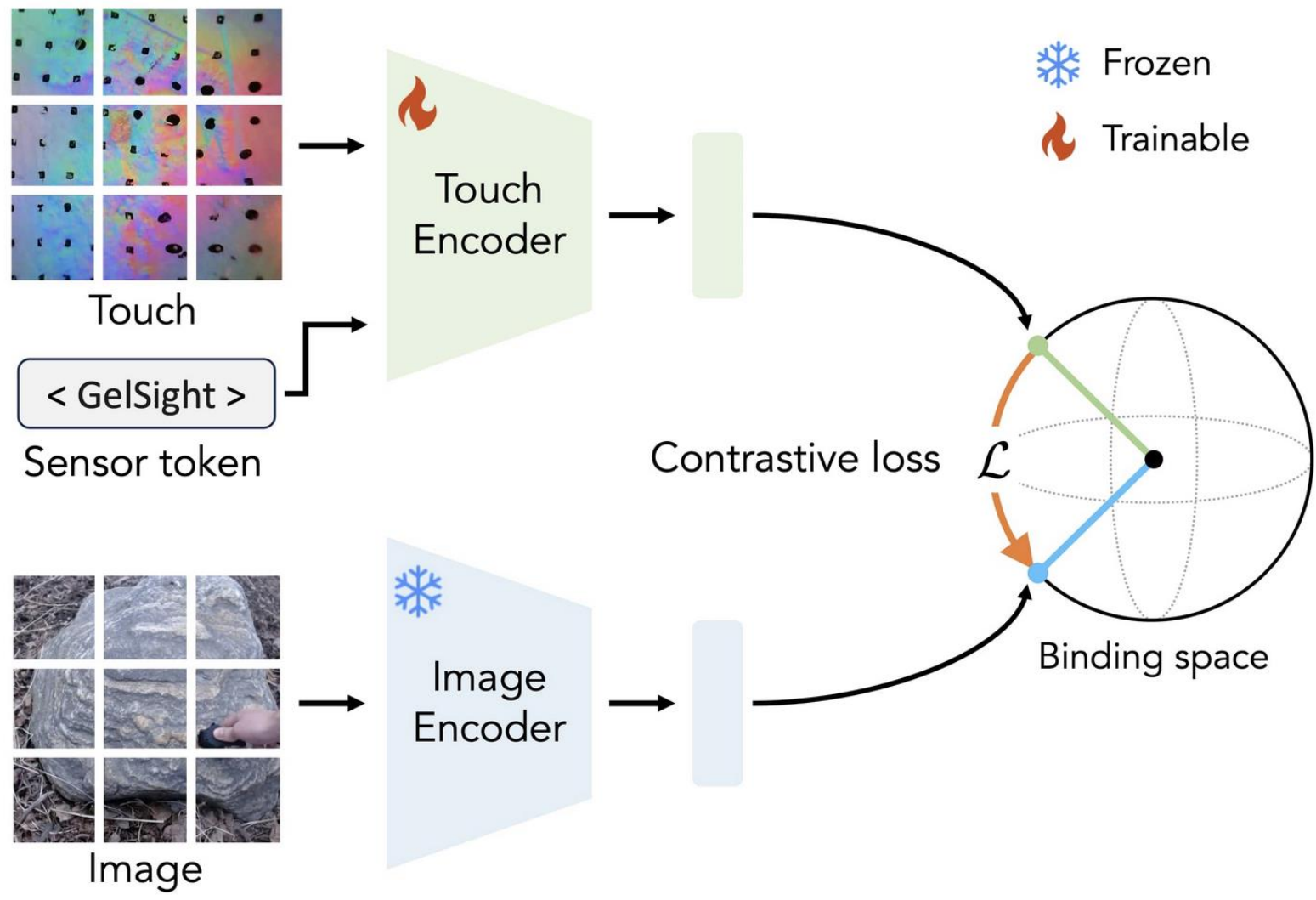
Binding Touch to Everything: Learning Unified Multimodal Tactile Representations

Fengyu Yang^{1*} Chao Feng^{2*} Ziyang Chen^{2*} Hyoungseob Park¹ Daniel Wang¹ Yiming Dou²
Ziyao Zeng¹ Xien Chen¹ Rit Gangopadhyay¹ Andrew Owens² Alex Wong¹

¹Yale University ²University of Michigan

Multi-Sensory Fusion: Binding Touch, Audio, Vision and Language





Embodied LLMs with Tactile Sensing Input

你能告诉我提供的触觉图里的触觉的感觉和质地纹理吗?

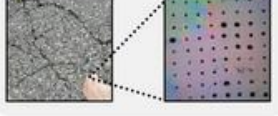
Reference Input touch



触觉图的触觉的感觉和质地纹理是由植物的叶子组成的, 这些植物的叶子呈现出细腻的纹理和柔软的触觉。

Can you describe material of the object presented in touch image?

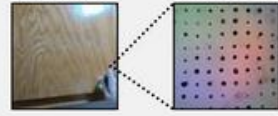
Reference Input touch



The object in the touch image is made of a **rough, grainy** material, which could be a mixture of **sand, small rocks**, or other similar materials.

How easy or difficult is it to hold or grip the object?

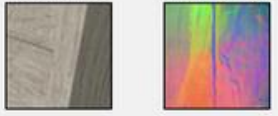
Reference Input touch



The object is likely to be a **flat surface**, which suggests that it might be **difficult** to hold or grip it securely.

Will people get hurt when hitting the point from the touch image? Why?


Reference Input touch



Yes. This is because the point is likely to be a **sharp edge**, which can cause **injuries** when it comes into contact with a person's skin.

What is the material of the object presented in touch image?

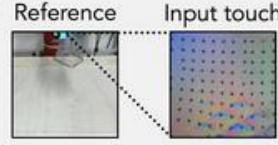
Reference Input touch



The object presented in touch image is likely made of **wood**.

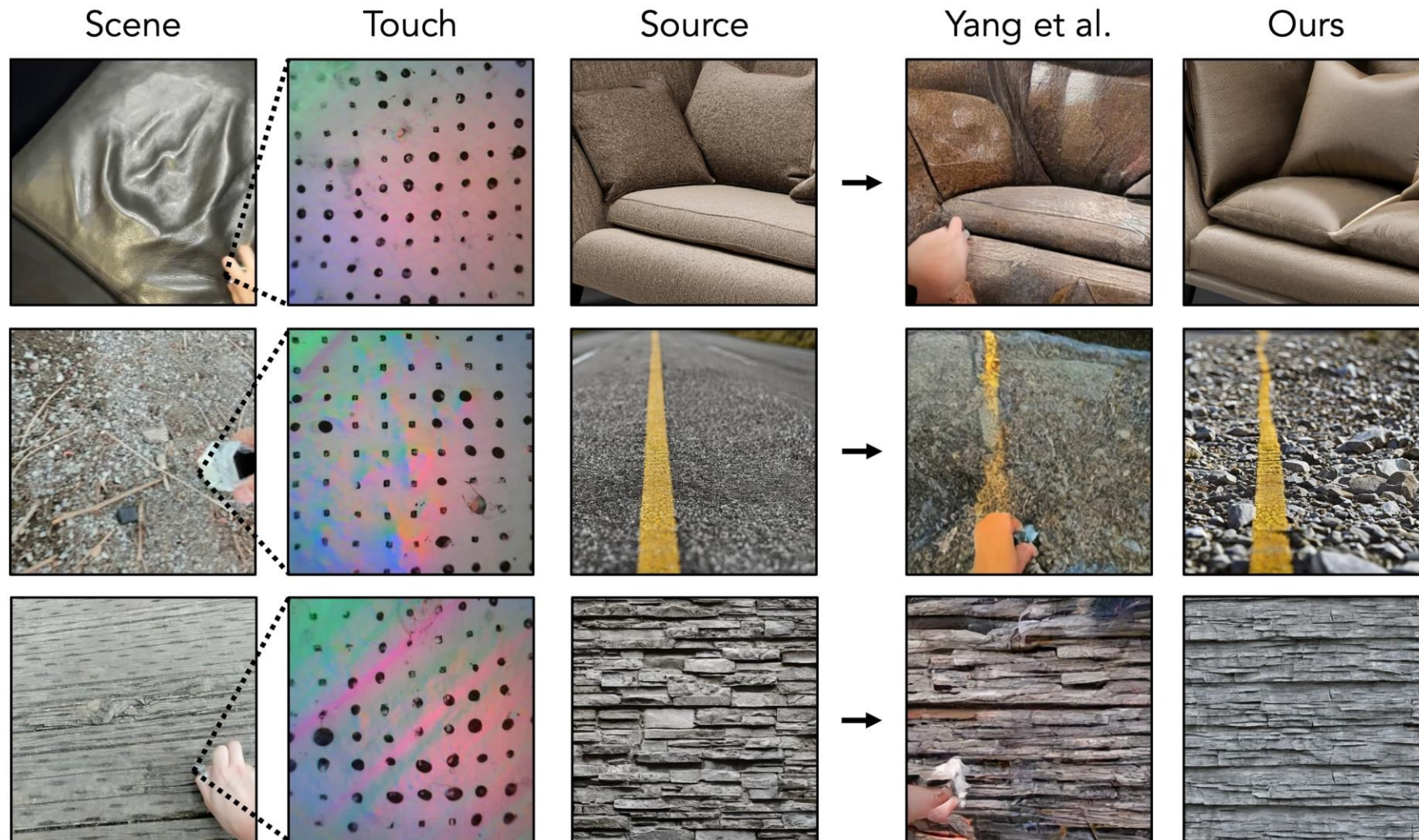
Is this object grasped into the air? Please explain whether the gripper is in contact with the object from the touch image.

Reference Input touch



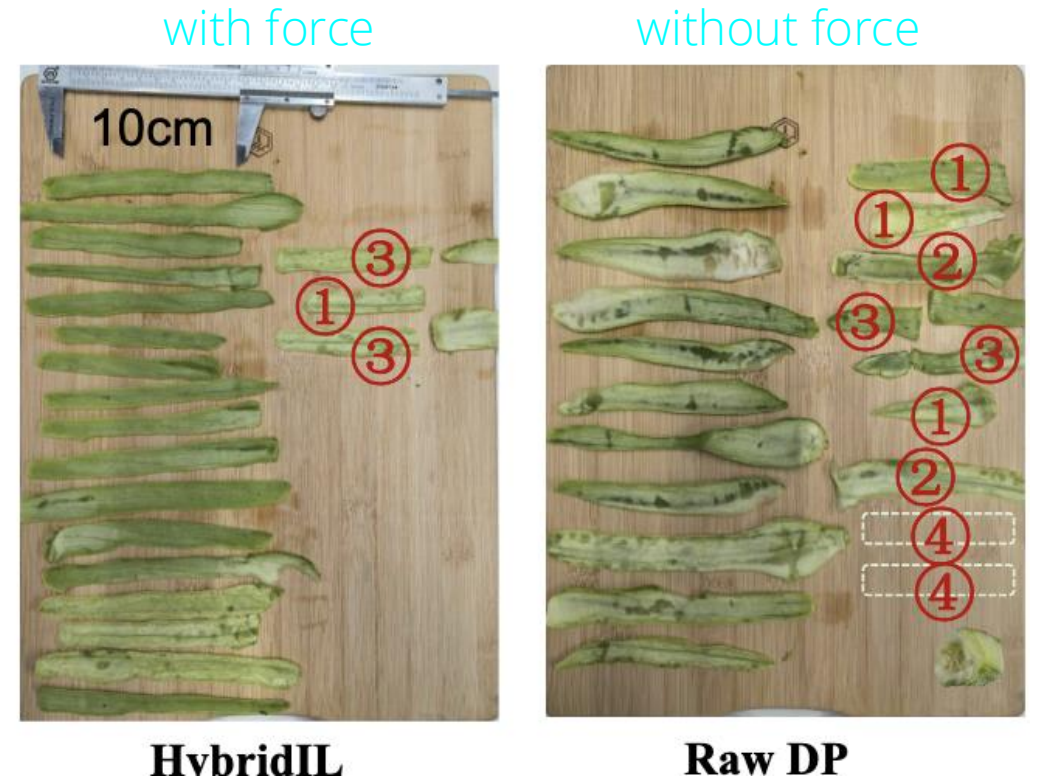
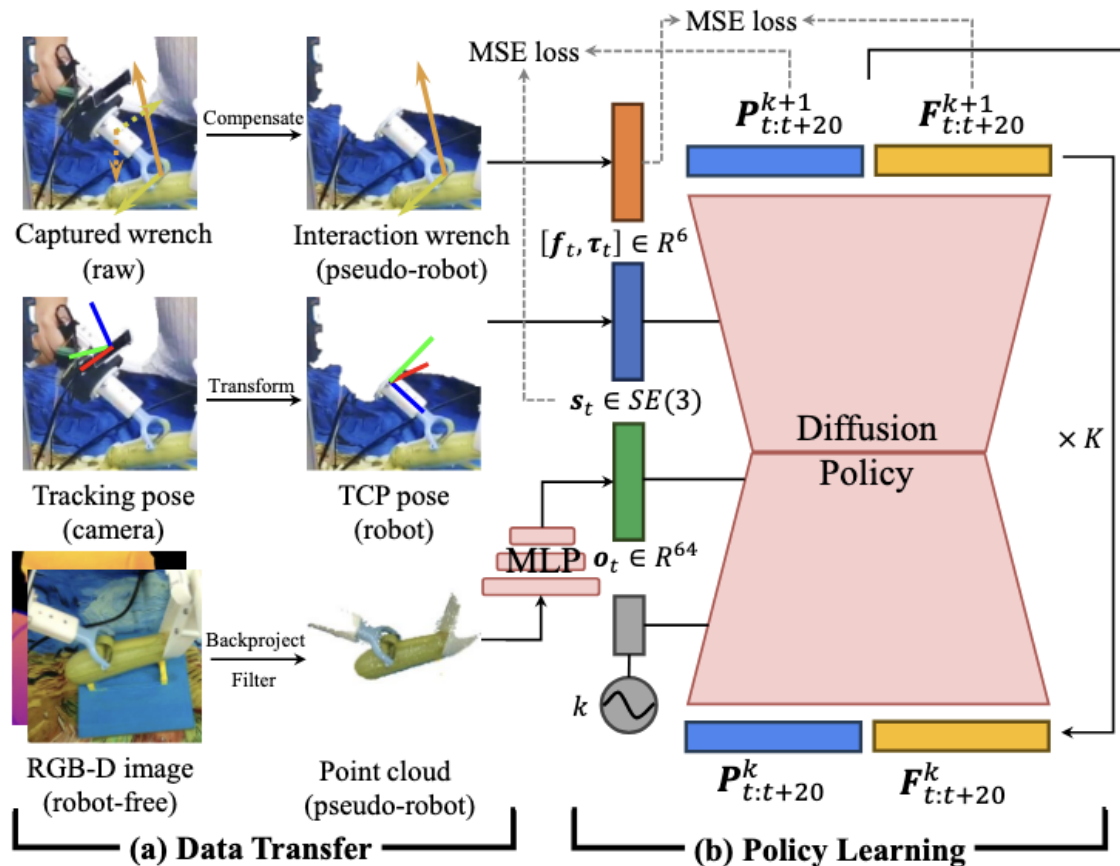
Yes. Because the gripper is still in **contact** with the object.

Touch-Driven Image Stylization



Physical Feedback is Essential for Embodied Agents

- Policies perform safer actions with physical feedback as inputs



Physical Feedback is Essential for AR / VR



What We Will Cover the Next Week

- Robotics
 - Brief introduction of robotics
 - Visual-language model for embodied perception
 - Visual-language-action model

Attendance Form

