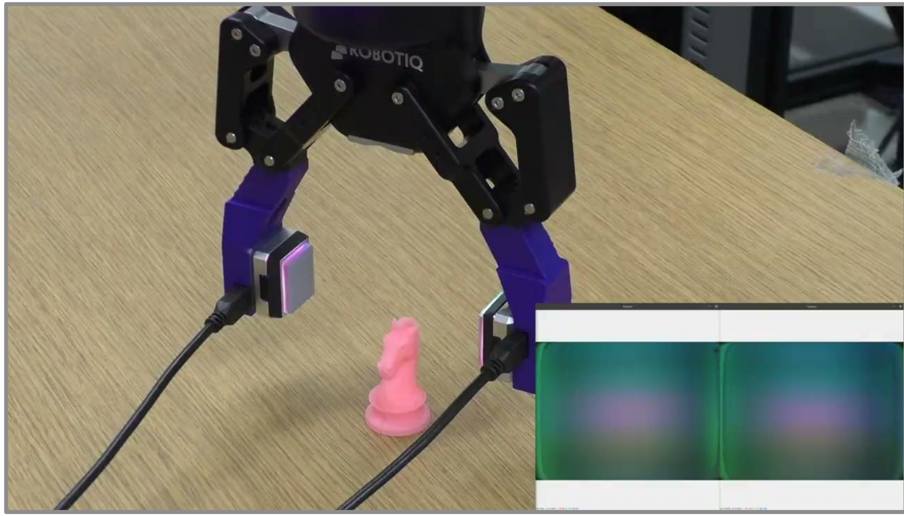
# DeepRob

[Student] Lecture 21

By Miles Priebe, Nirmal Raj, and Adam Imdieke

**Tactile Perception for Robot Grasping and Manipulation**

University of Michigan and University of Minnesota

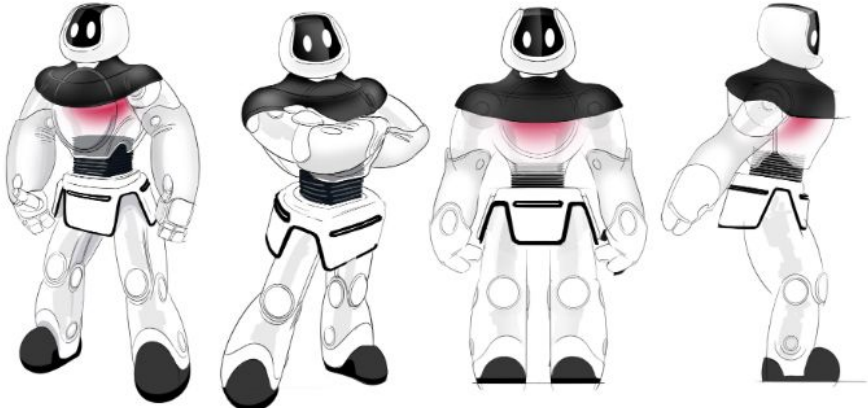



Gelsight grasp demo

# Agenda

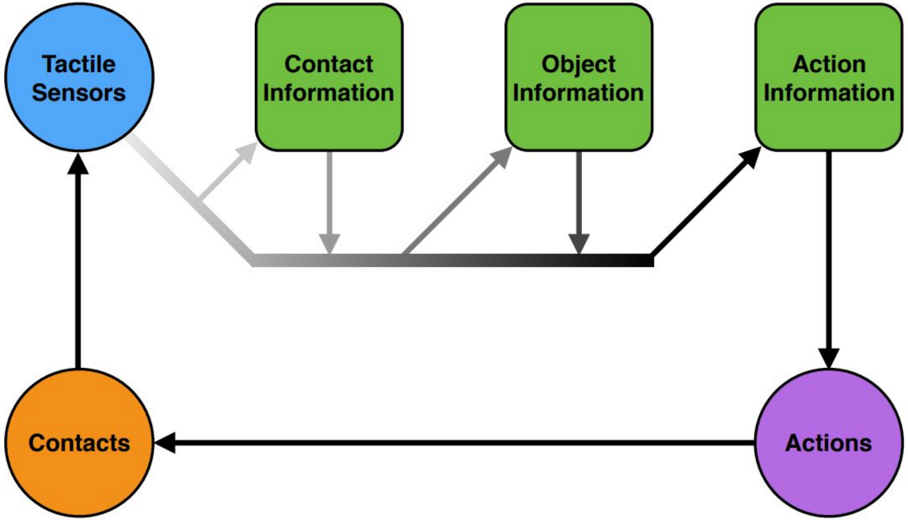
- Tactile perception
- Signal categories
- Types of sensors
- Haptic vs Tactile sensing
- Gelsight
- Tac2Pose
- Tacto
- Tactile sensing for Deep Learning

Punyo



# What is Tactile Perception?

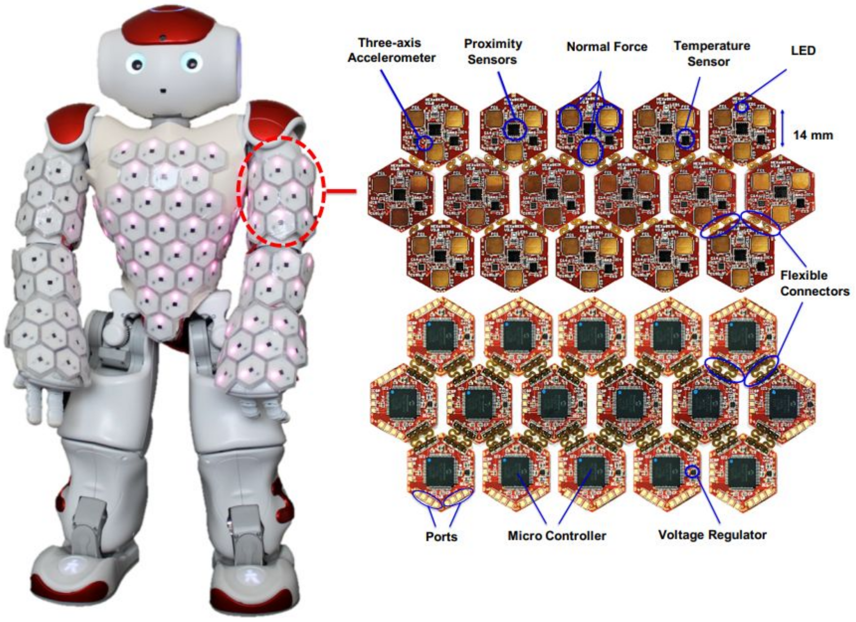
- Key sensor modality for robots
- Provide a rich and diverse set of data signals about...
  - Contact
  - Objects
  - Actions




A Review of Tactile Information: Perception and Action Through Touch, Li et al. , 2020

# Sensor-Level Signals

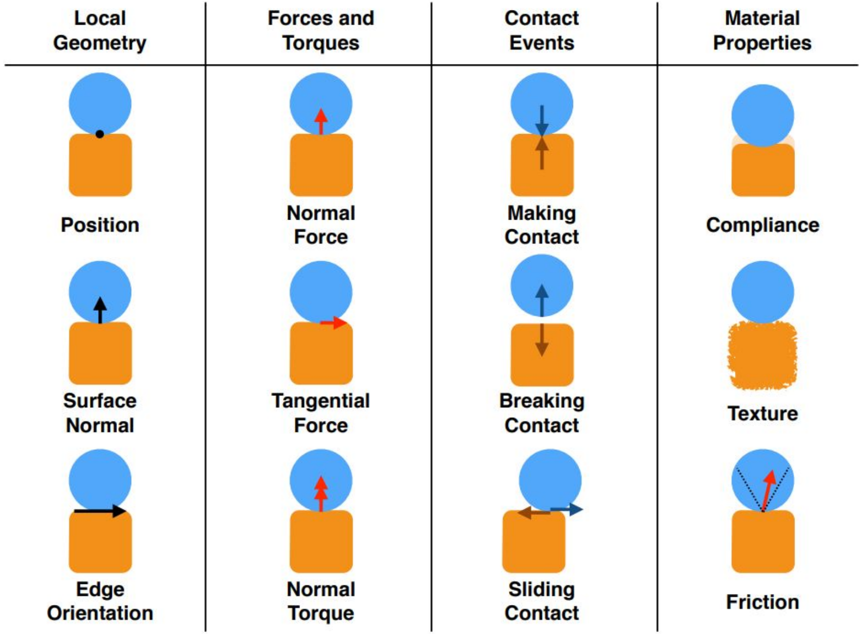
- Normal and tangential force
- Vibration
- Thermal
- Pretouch proximity
- Sensor coverage




A Review of Tactile Information: Perception and Action Through Touch, Li et al. , 2020

# Contact-Level Signals

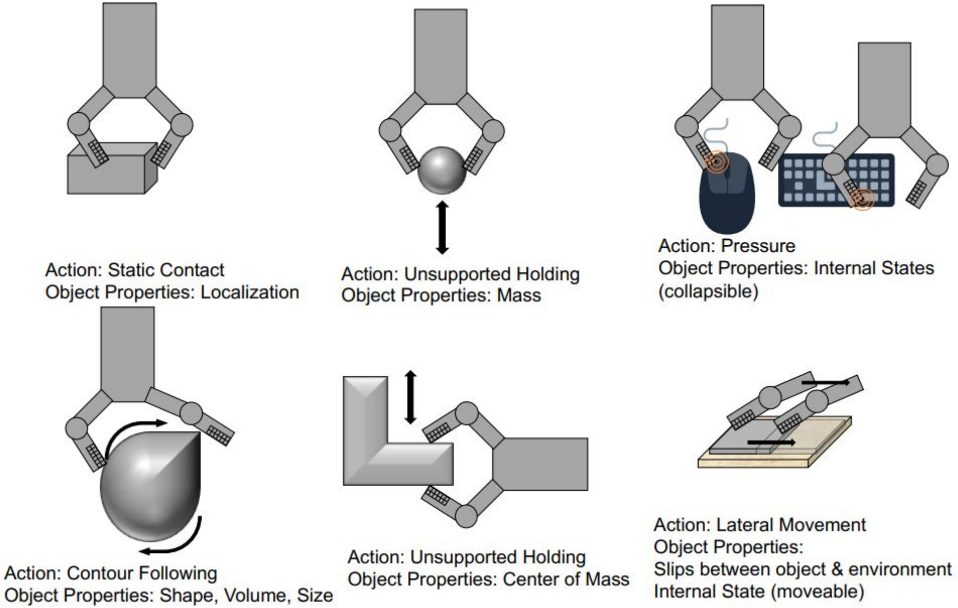
- Contact geometry
- Force and torque
- Contact events
- Material properties




A Review of Tactile Information: Perception and Action Through Touch, Li et al. , 2020

# Object-Level Signals

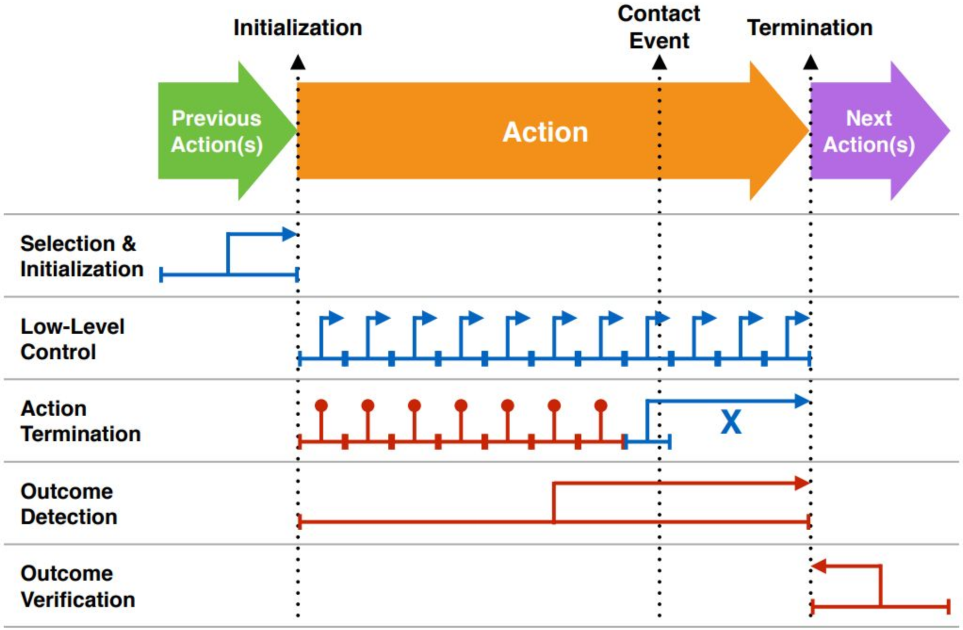
- Object localization
- Shape
- Mass and dynamics
- Contents of containers




A Review of Tactile Information: Perception and Action Through Touch, Li et al. , 2020

# Action-Level Signals

- Action selection and initialization
- Tactile feedback for low-level control
- Action termination
- Action outcome detection
- Action outcome verification

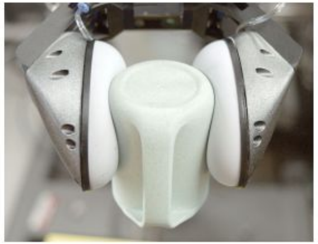



A Review of Tactile Information: Perception and Action Through Touch, Li et al. , 2020

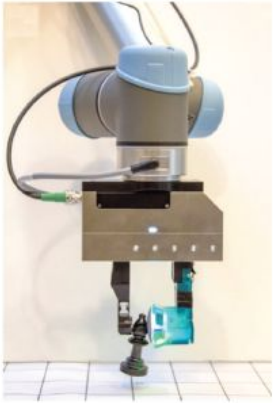
# Sensors

Facebook AI Digit

**Visuotactile**



TRI Punyo SoftBubble

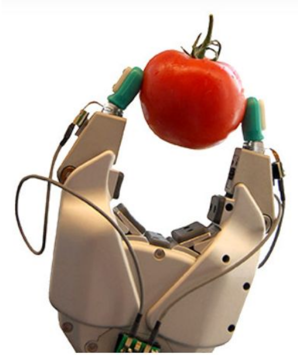


GelSight Mini

**Tactile**

Xela Robotics uSkin

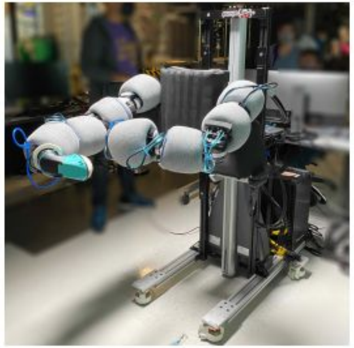
SynTouch BioTac

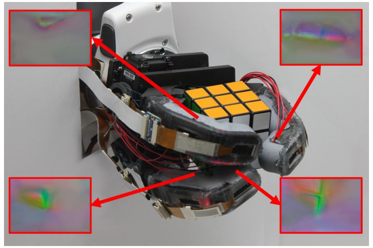



https://digit.ml/
https://arxiv.org/abs/2004.03691
https://xelarobotics.com/en/integrations
https://syntouchinc.com/robotics/#:~:text=core%2C%20enhancing%20durability.-,The%20sensor%20that%20paved%20the%20way%20to%20machine%20touch,capabilities%20of%20the%20human%20fingertip.
https://www.mdpi.com/1424-8220/17/12/2762#

# Tactile Sensing Timeline

- 2008 — BioTac
- 2016 — uSkin
- 2017 — GelSight Mini
- 2020 — SoftBubble
- 2020 — Digit
- 2021 — GelSlim 3.0
- 2022 — 3-Finger Gelsight
- 2023 — Punyo-1





https://arxiv.org/abs/2111.09354
https://arxiv.org/abs/2103.12269
https://news.mit.edu/2023/robotic-hand-can-identify-objects-just-one-grasp-0403

# Haptic vs. Tactile Sensing

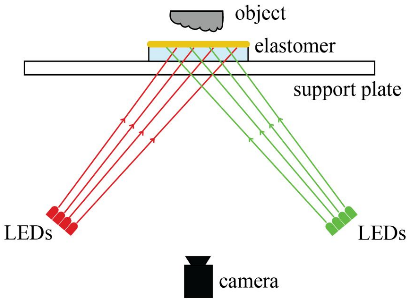
## Haptic

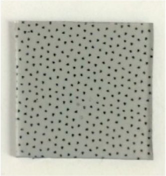
- Overall sensory experiences:
  - Tactile
  - Proprioception
  - Kinesthesia

## Tactile

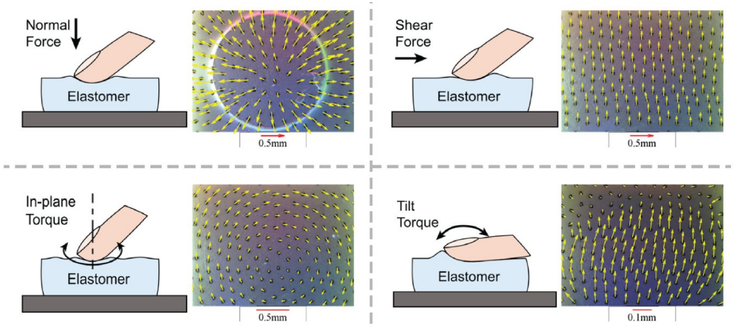
- Detection of physical sensations:
  - Pressure
  - Temperature
  - Texture

# What is Visuotactile Sensing?

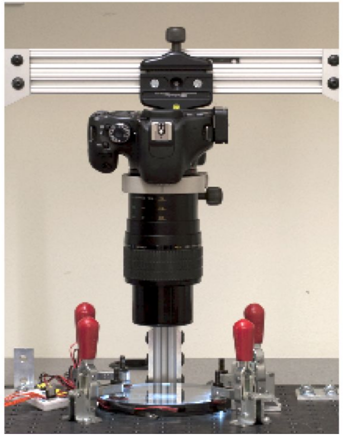
(a)

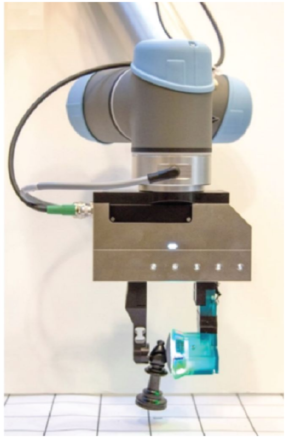


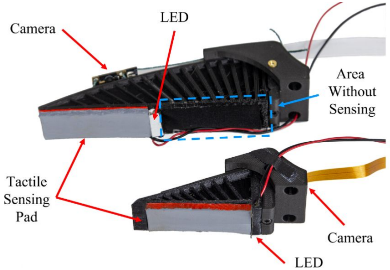

(b)

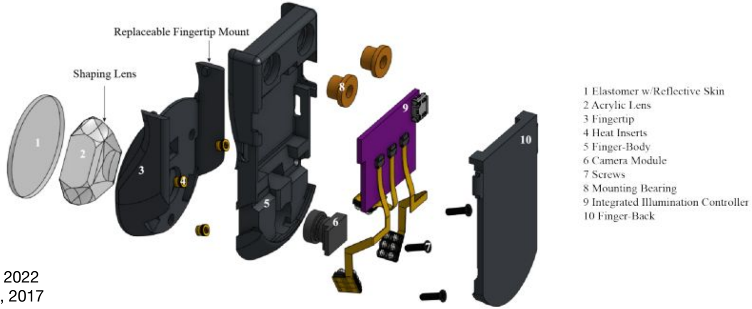
GelSight Fin Ray: Incorporating Tactile Sensing into a Soft Compliant Robotic Gripper, Liu et al. , 2022
GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force,  Yuan et al. , 2017

# Exploring Gelsight Evolution







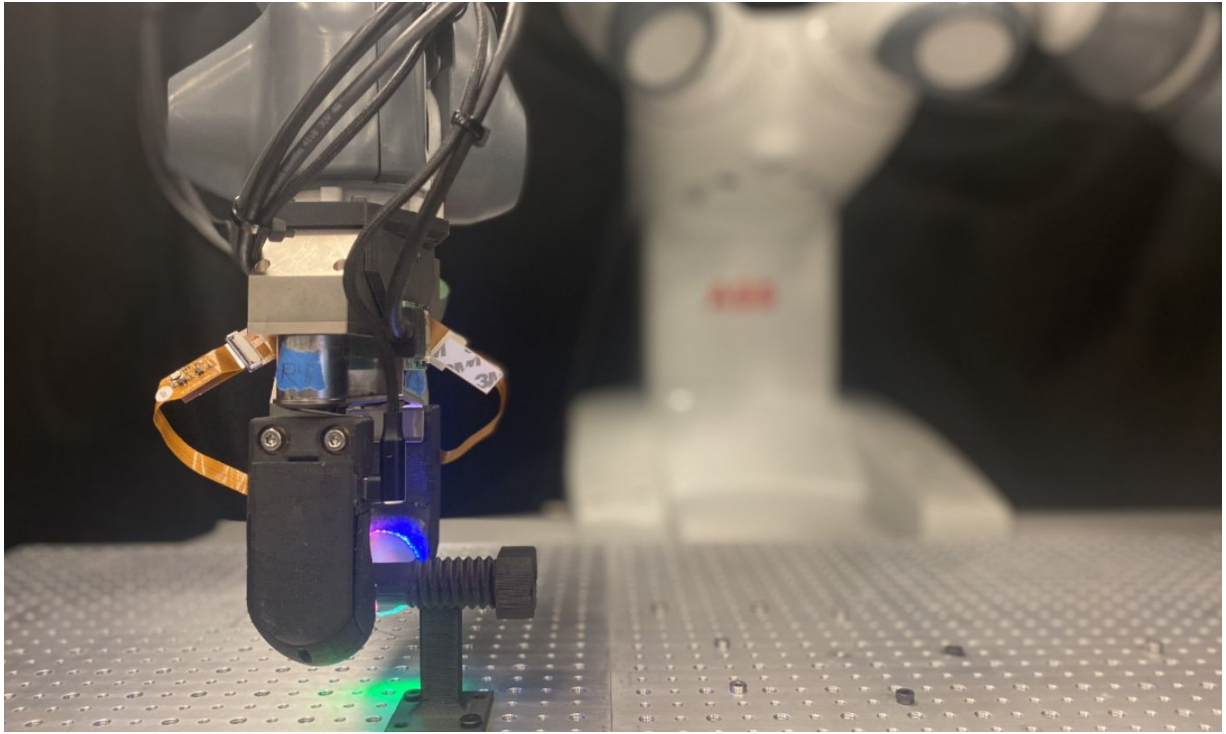




GelSight Fin Ray: Incorporating Tactile Sensing into a Soft Compliant Robotic Gripper, Liu et al. , 2022
GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force, Yuan et al. , 2017

# Tac2Pose: Tactile Object Pose Estimation from the First Touch.

Maria Bauza,
Antonia Bronars, and
Alberto Rodriguez

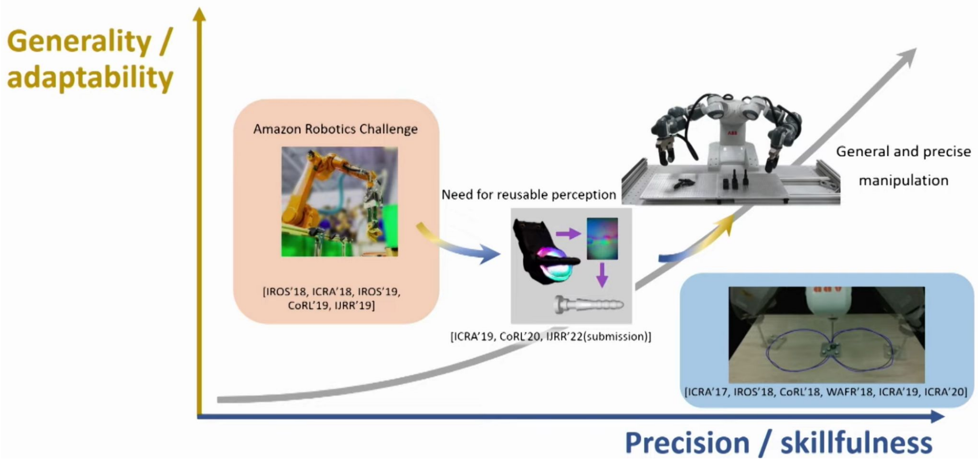
*CORL 2022*

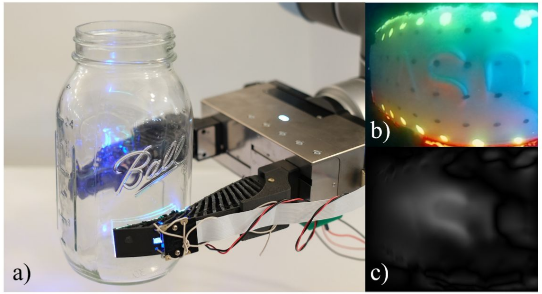


Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Motivation

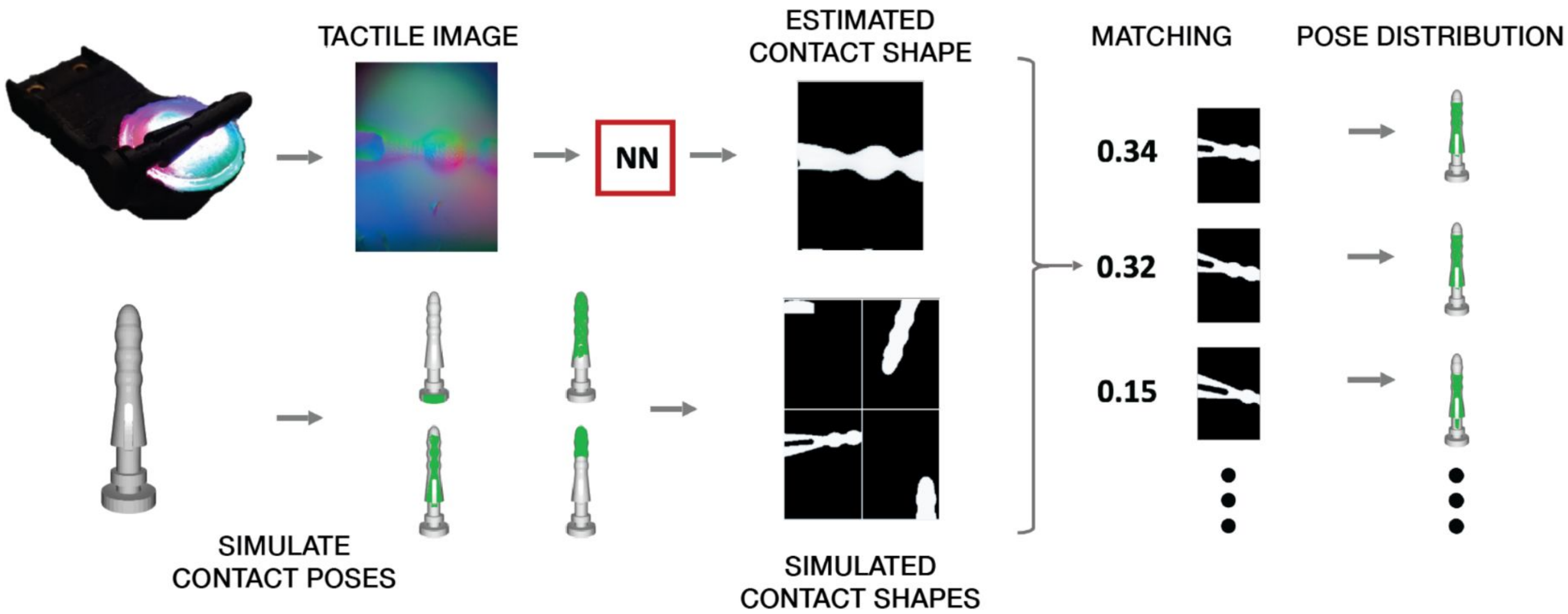
- Close the loop
  - Know the pose of the object
  - React to uncertainty
- Industry needs Precision
  - Often Specialized Solutions
- General solutions often lack Precision




GelSight Fin Ray: Incorporating Tactile Sensing into a Soft Compliant Robotic Gripper, Liu et al. , 2022
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022
[PhD Thesis Defense - Maria Bauza - Visuo-Tactile Perception for Dexterous Manipulation]

# Methods

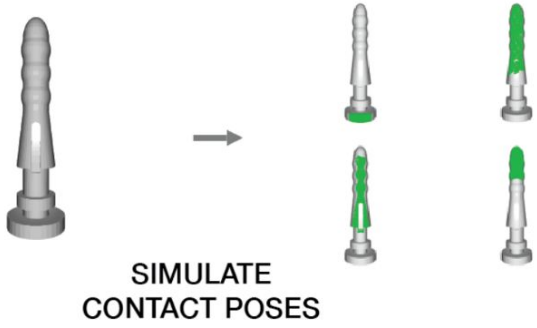



Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Methods

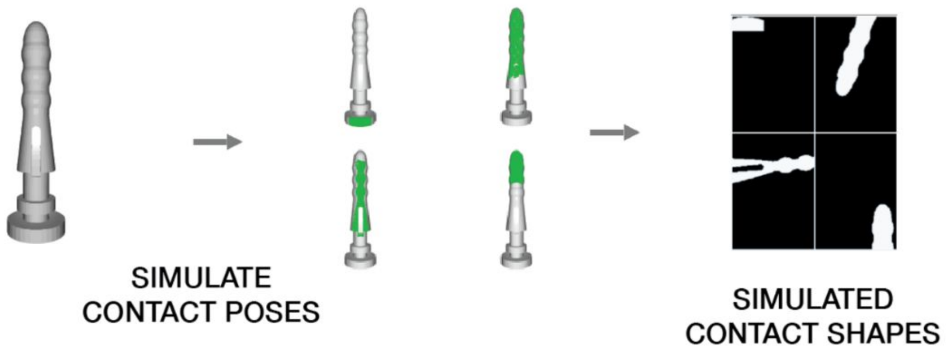
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Methods



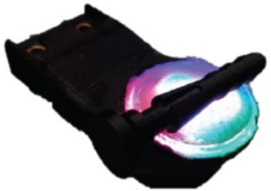

Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

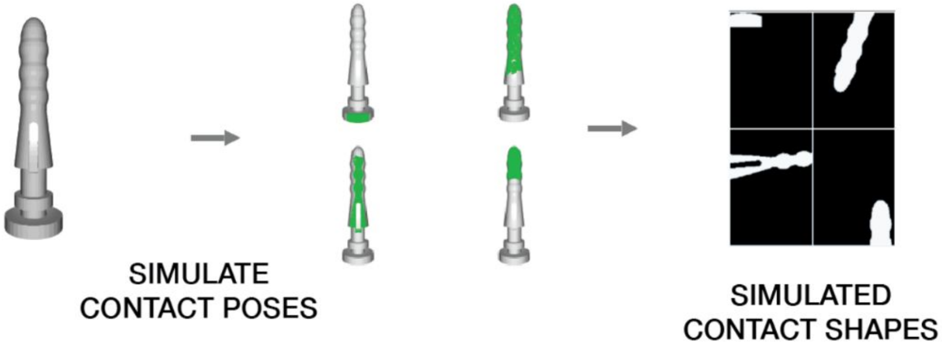
# Methods




Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

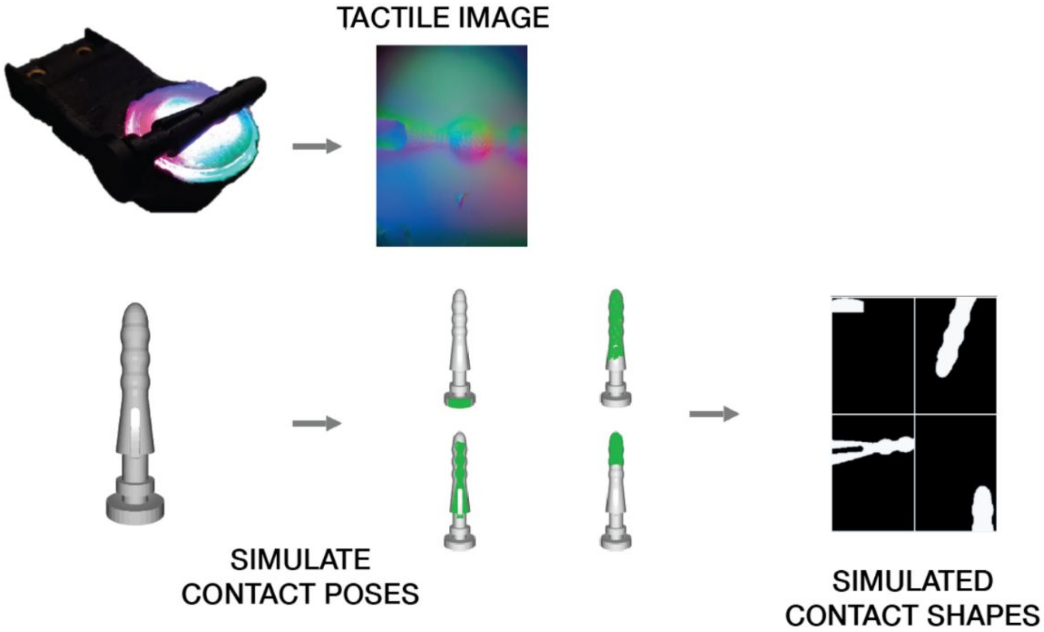
# Methods





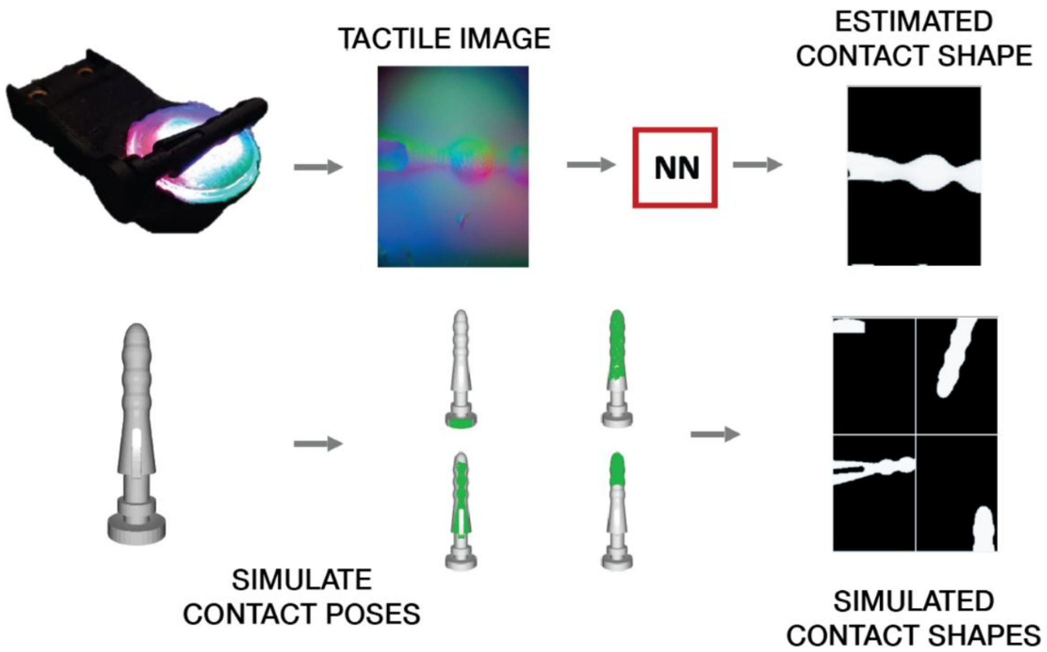

Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Methods



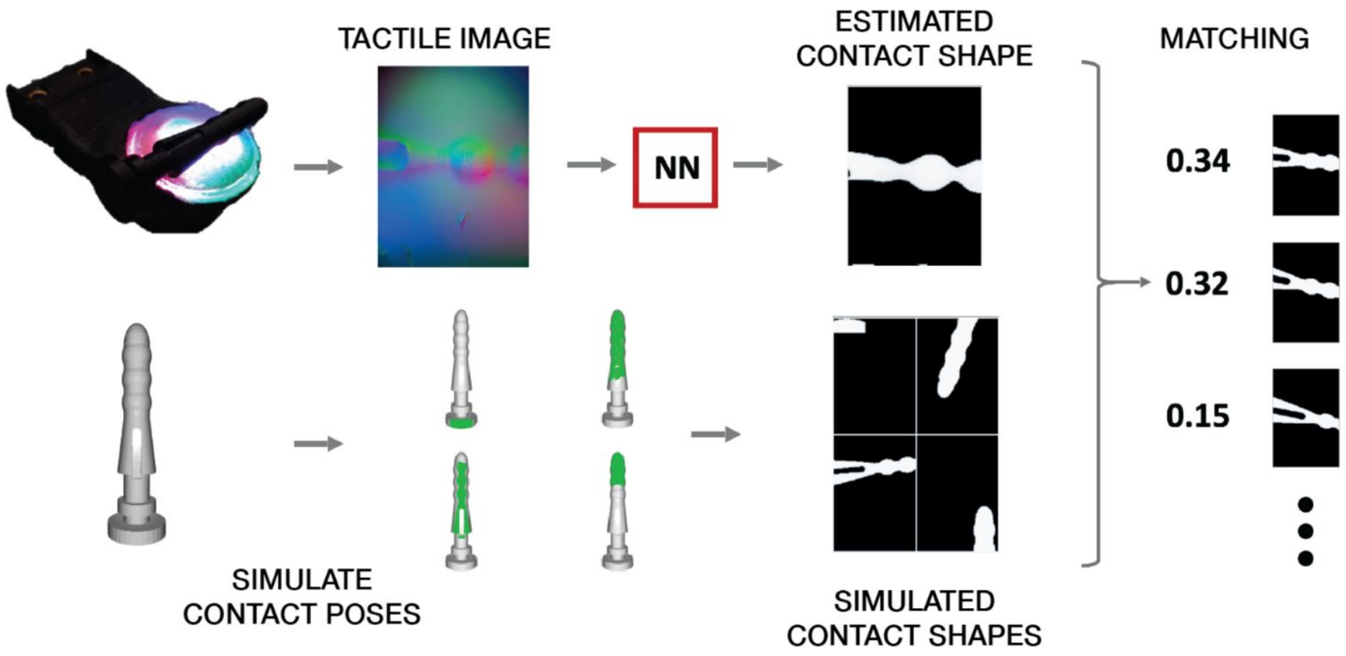

Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Methods



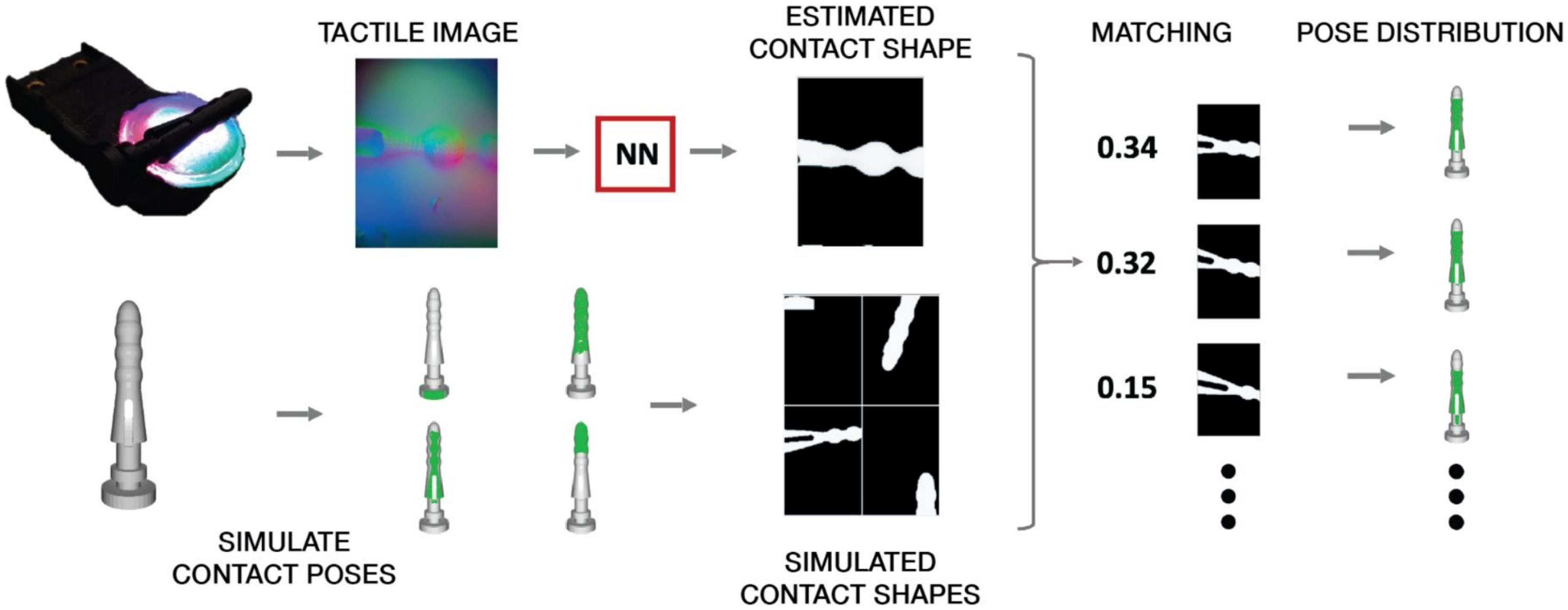

Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Methods



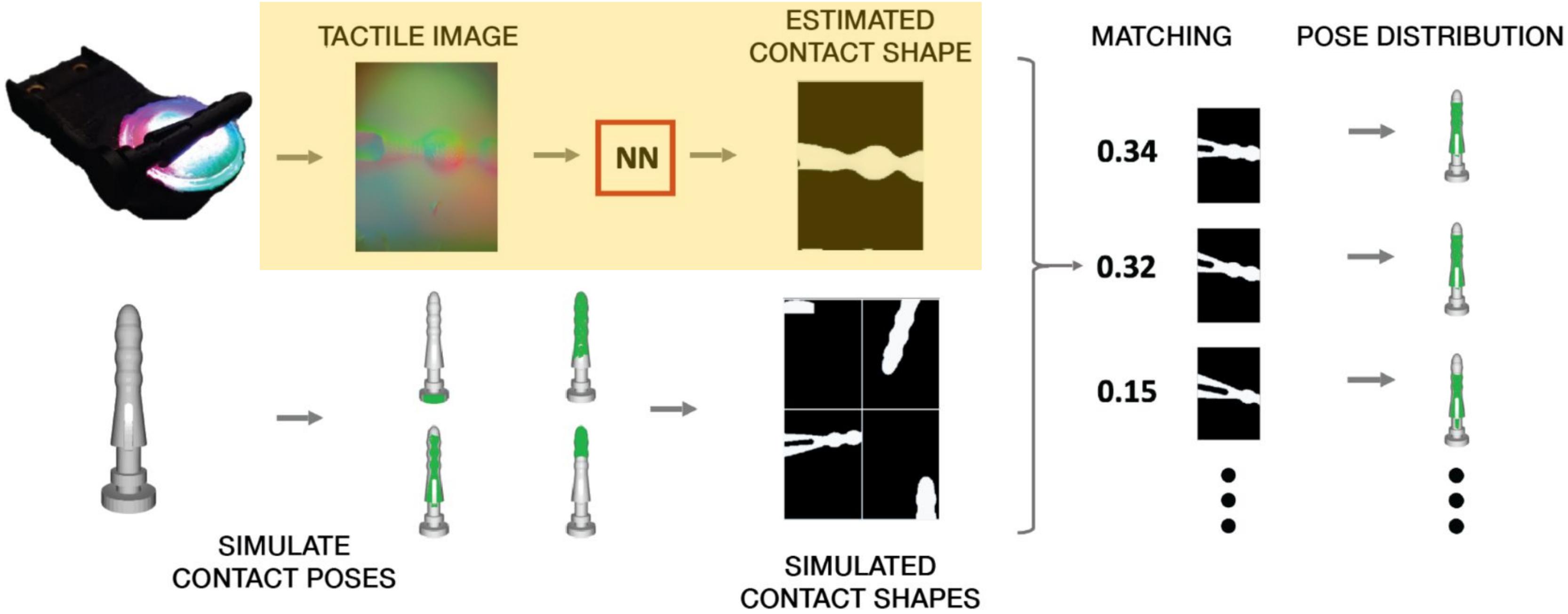

Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Methods




Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Contact Shape Network




Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Contact Shape Network

Known GelSight Pose

- Real image
- Simulated Binary Contact Shape

General Enough to work across sensors

Based on: [Image-to-Image Translation with Conditional Adversarial Networks]

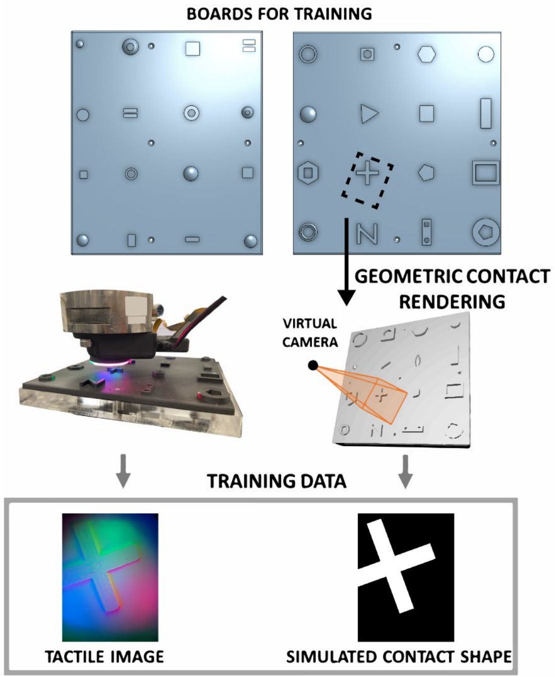
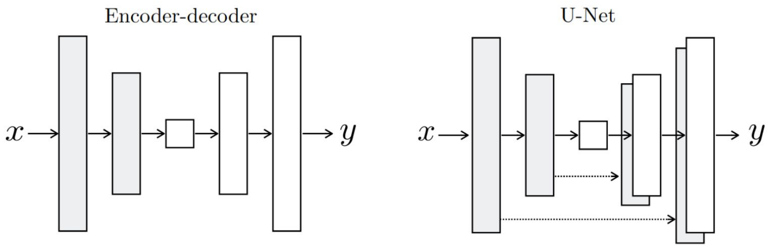



Image-to-Image Translation with Conditional Adversarial Networks, Isola et at. , 2018
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Image-to-Image Translation

- Input: Real image
- Output: Estimated Contact Shape

Trained with ground truth data.




Example: Satellite to Map


Example: Fake shoe to convincing fake shoe

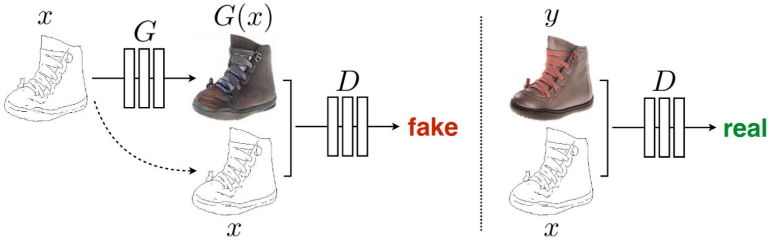


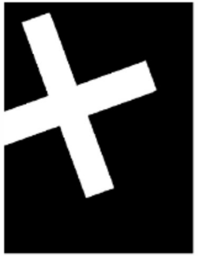


Image-to-Image Translation with Conditional Adversarial Networks, Isola et at. , 2018
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

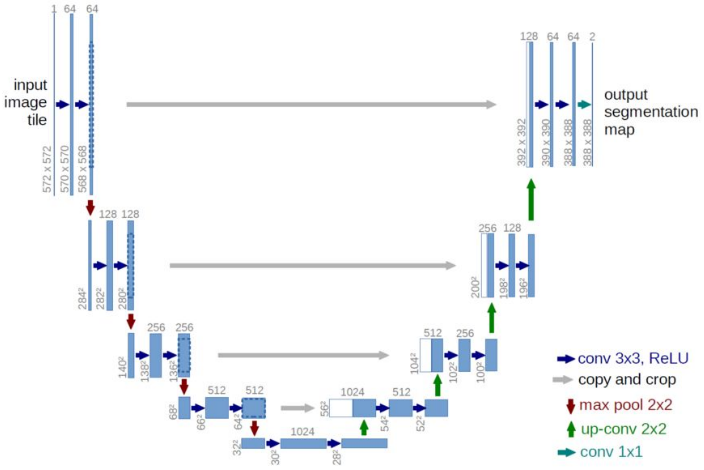
# Image-to-Image Translation

Generator: Creates contact mask
Discriminator: Identifies fake images

TACTILE IMAGE



SIMULATED CONTACT SHAPE



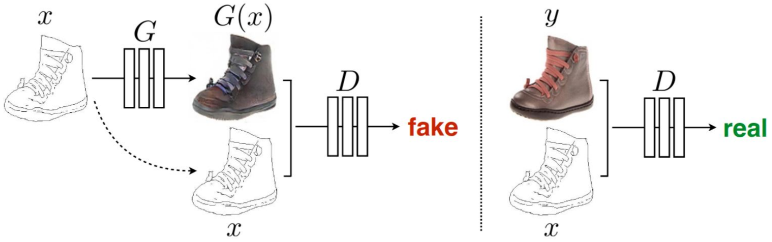
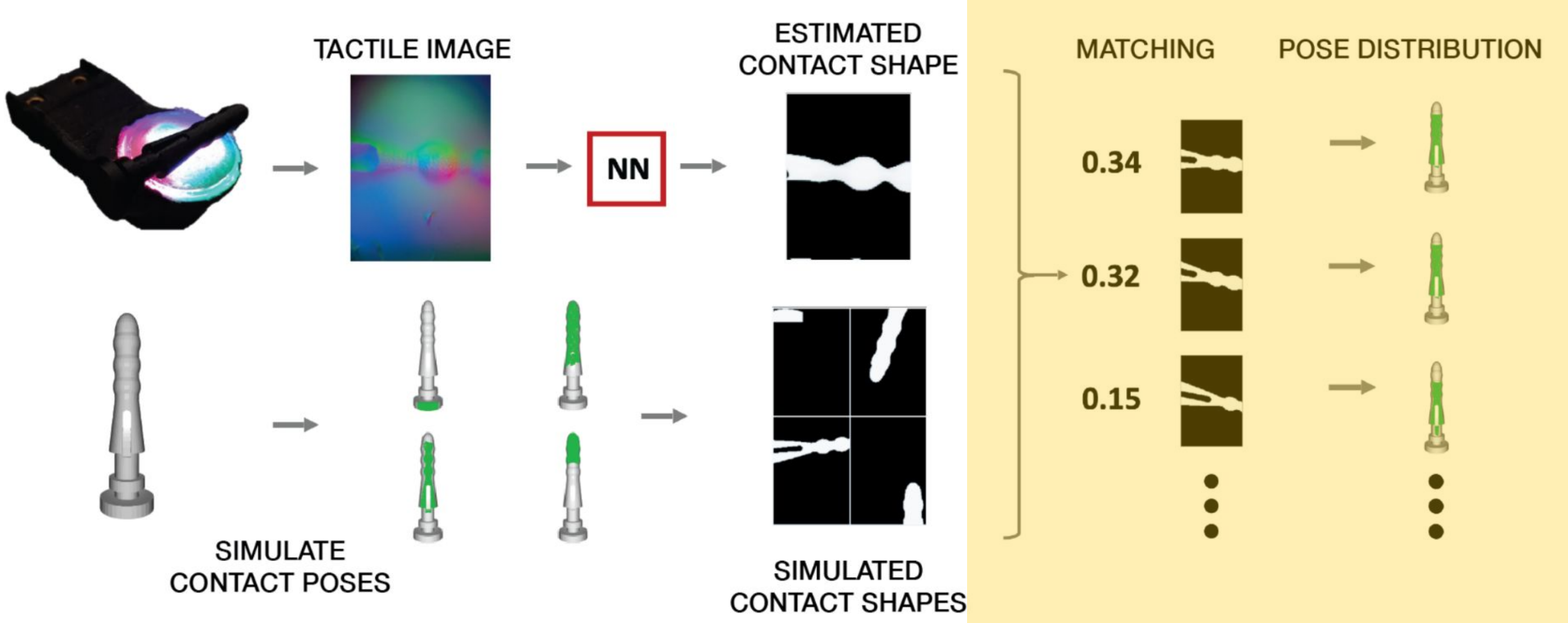



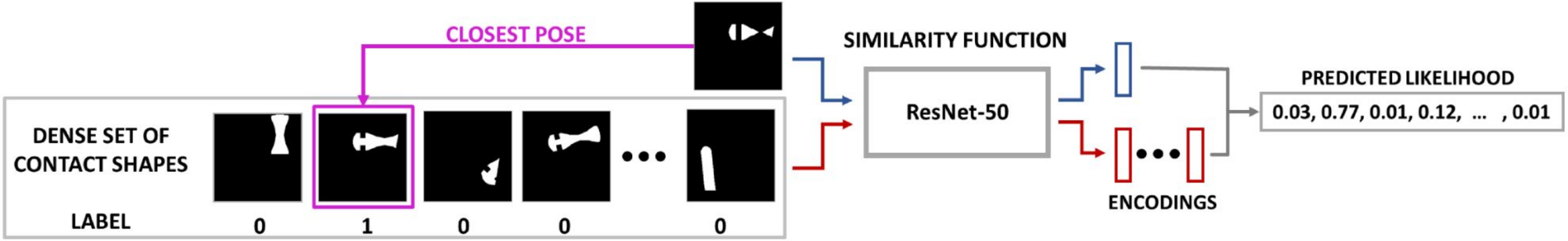

Image-to-Image Translation with Conditional Adversarial Networks, Isola et at. , 2018
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Pose Matching

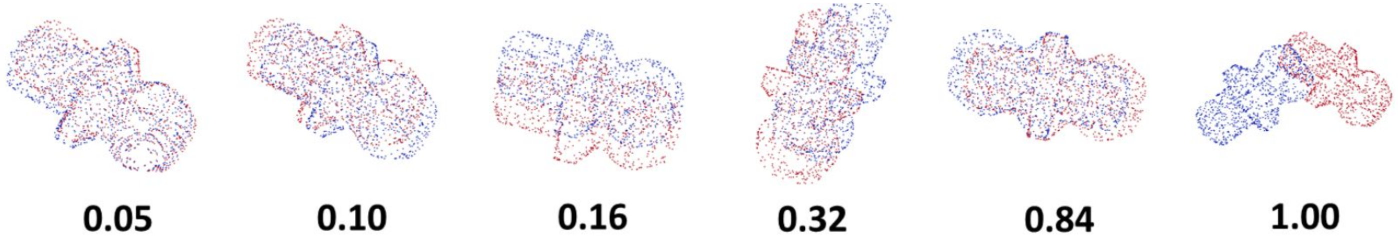



Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Pose Matching



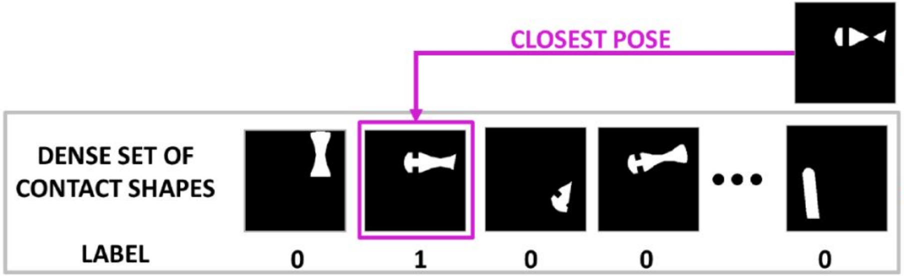

Normalized Pose Errors:

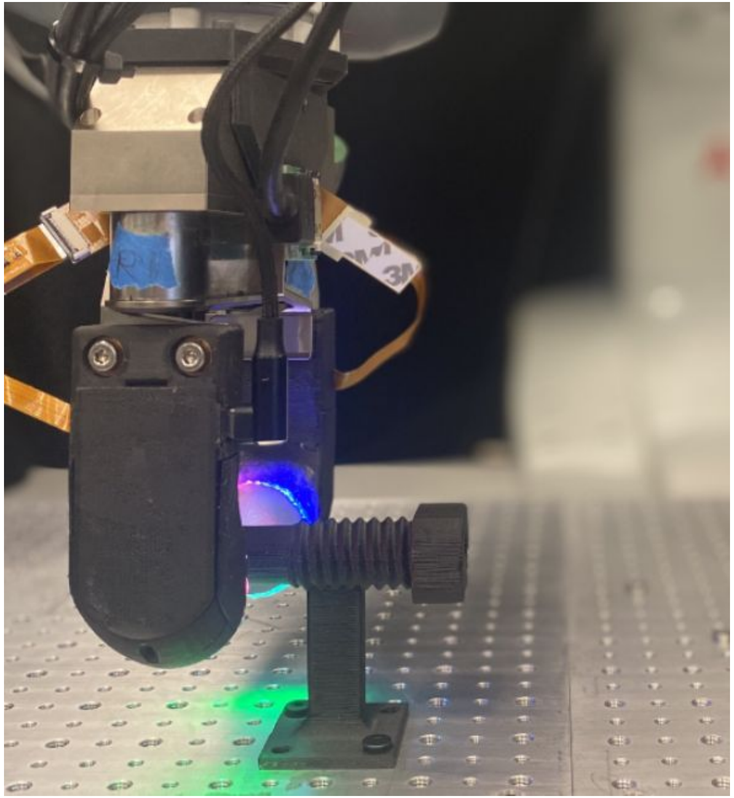



Normalized Pose Error: Original pose error divided by the average error obtained from predicting a random contact pose.

Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Pose Matching



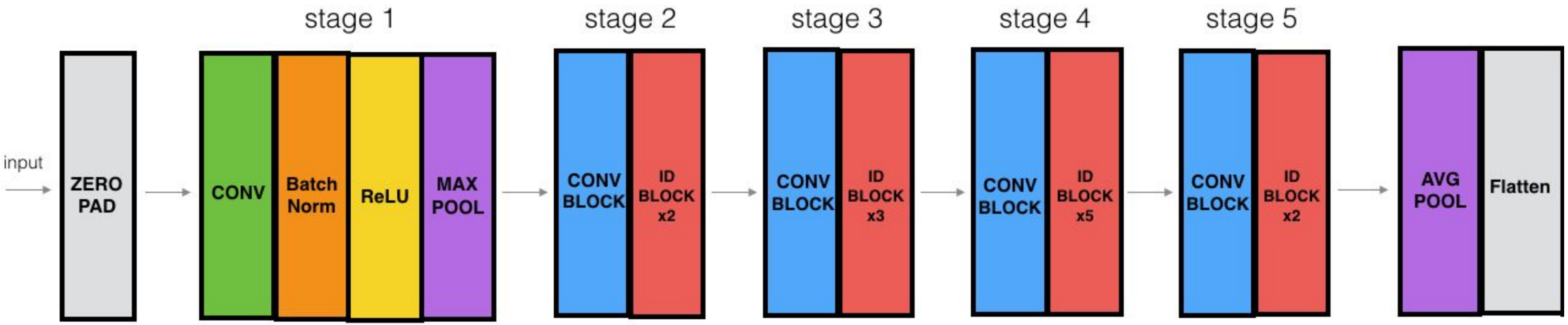
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Pose Matching


Encoder Network:
Pre-Trained ResNet 50

*Removed Linear Layers*

Trained with: [MoCo]



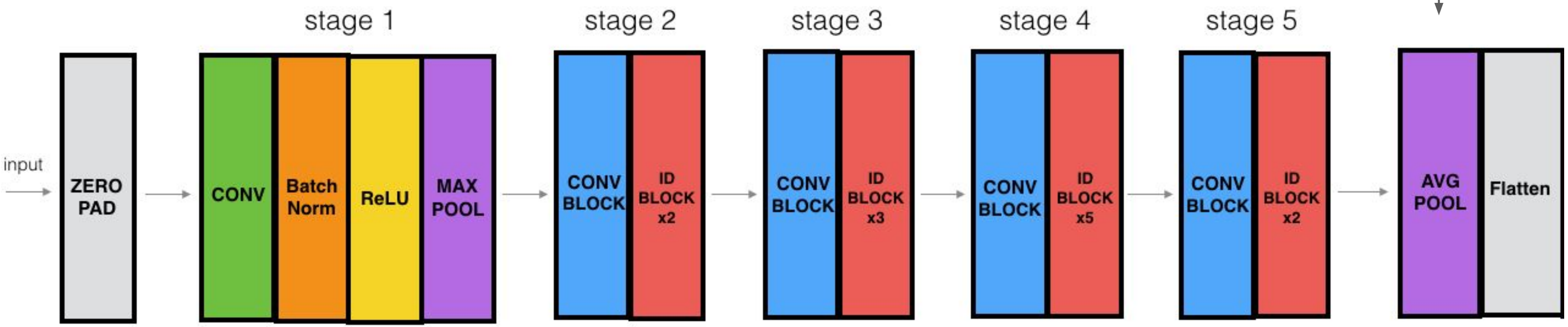

Deep Residual Learning for Image Recognition, He et al. , 2015
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Pose Matching


**It could never be that easy.**

**The designers this put in an AveragePool!**

**The AveragePool is removing critical spatial information.**



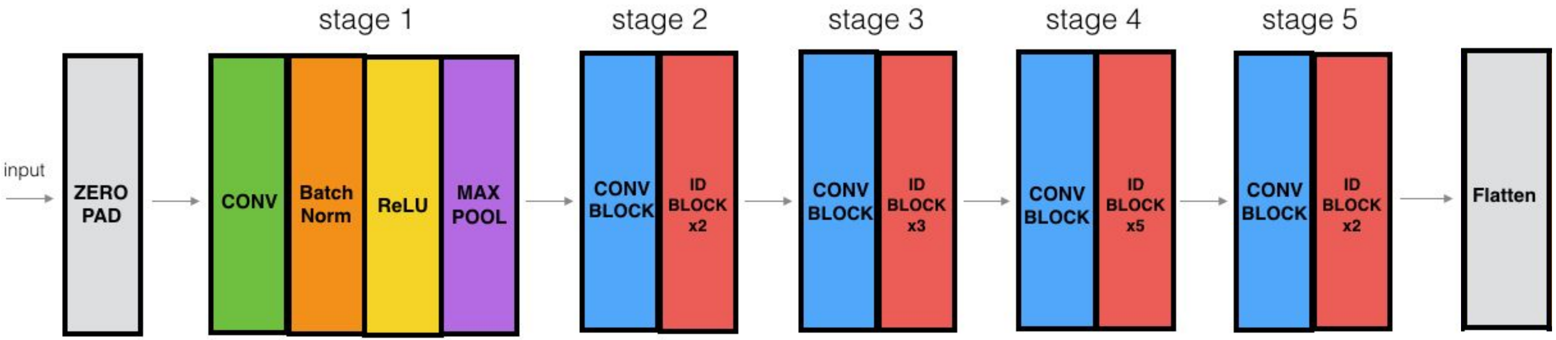

Deep Residual Learning for Image Recognition, He et al. , 2015
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Pose Matching


It could never be that easy.

The designers this put in an AveragePool!

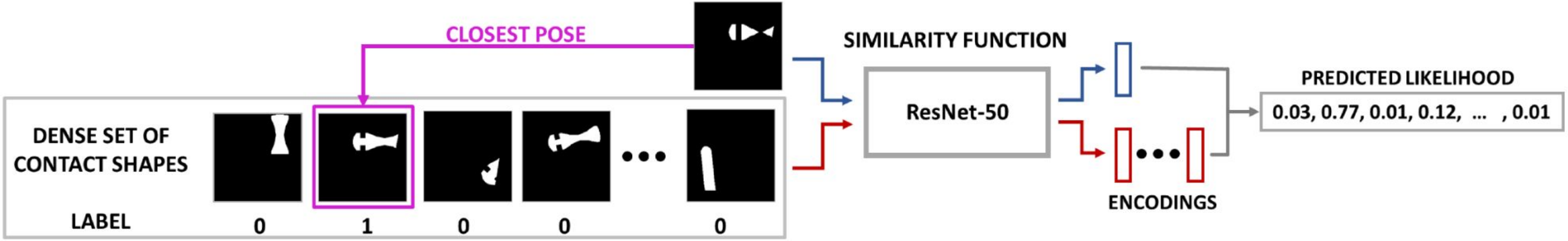
The AveragePool is removing critical spatial information.




Much Better

Deep Residual Learning for Image Recognition, He et al. , 2015
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Pose Matching




Take Largest Probability from SoftMax to get the inferred pose!

Number of Contact Shapes: N
Encoding Size: S
Encoded GelSight mask: 1xS
Encoded simulation mask: NxS
Dot product distances: 1xN

Probability = SoftMax(Dot product distances)

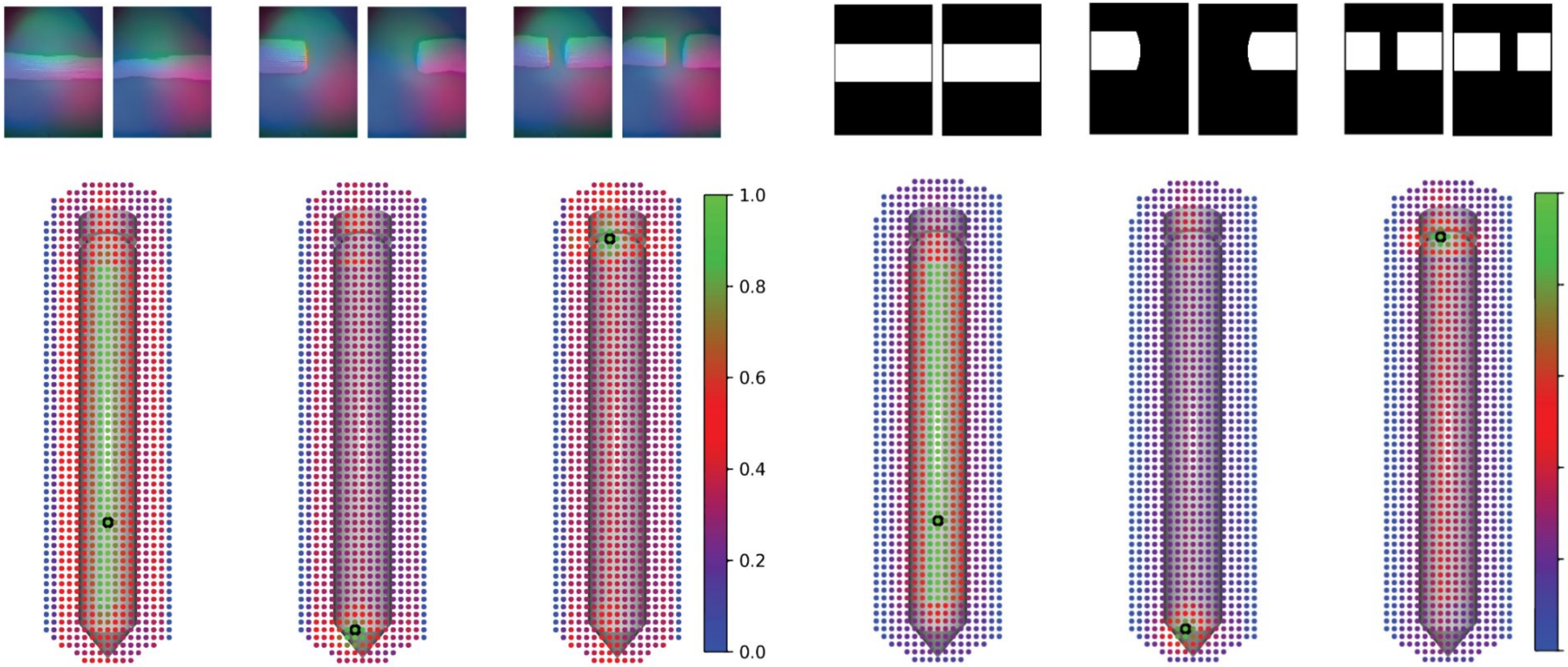
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Look! A pencil, how Fun!

Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

It is the best image I could find, give me some slack.

# Fun is not allowed

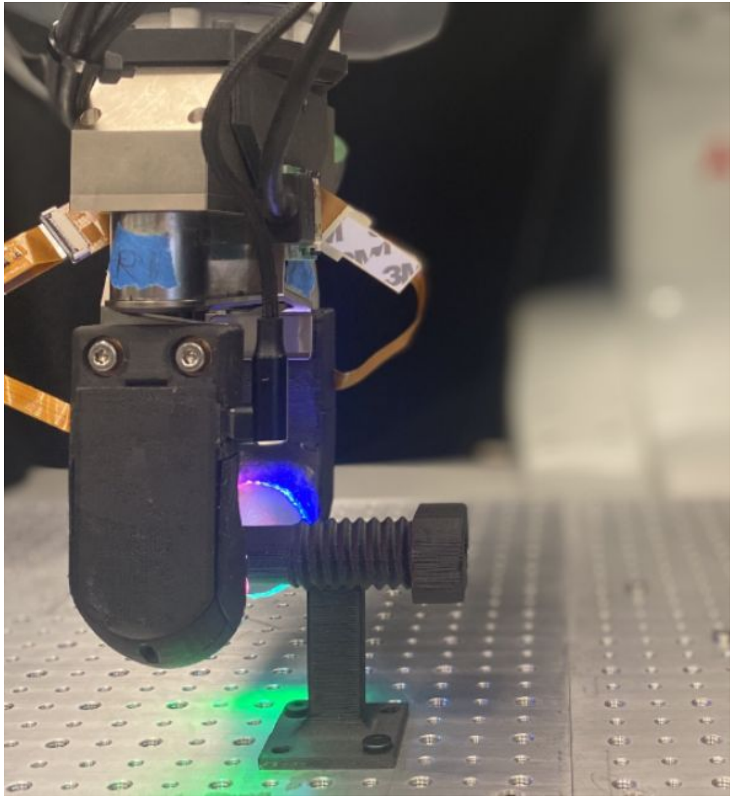



Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Pose Distributions


These sensors can be small!
Only a few $cm^2$ of perception per sensor.

- Unique contact patterns are needed to disambiguate non-unique contact mappings.
- Better to know that you don't know, than to know nothing at all.
    - Regrasp on ambiguous pose distributions.
    - Combine with other modes of sensing (Eg, Visual, Sound, Smell?).



Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Comparisons to Other Methods

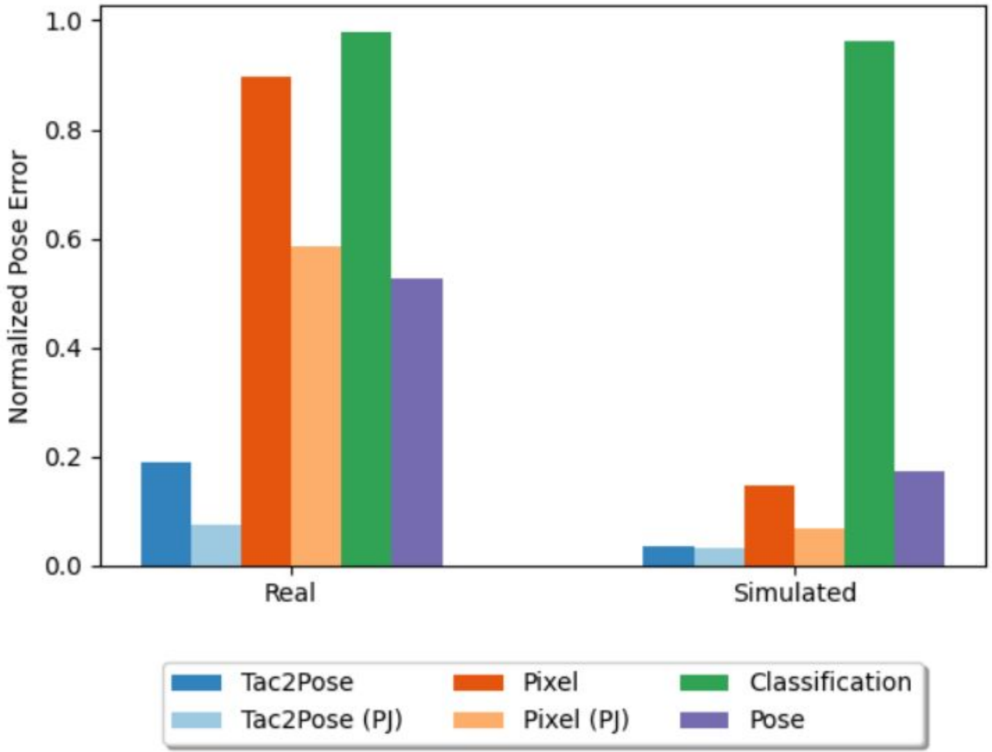
**Tac2Pose**: Method Described above.

**Pixel**: No encoder, direct pixel to pixel matching between input contact mask and all simulated contact masks.

**Classification**: Resnet-50, trained to One-Hot classify each discrete pose.

**Pose**: Resnet-50, trained to regress each pose.

***(PJ)**: Parallel Jaws, two images.*



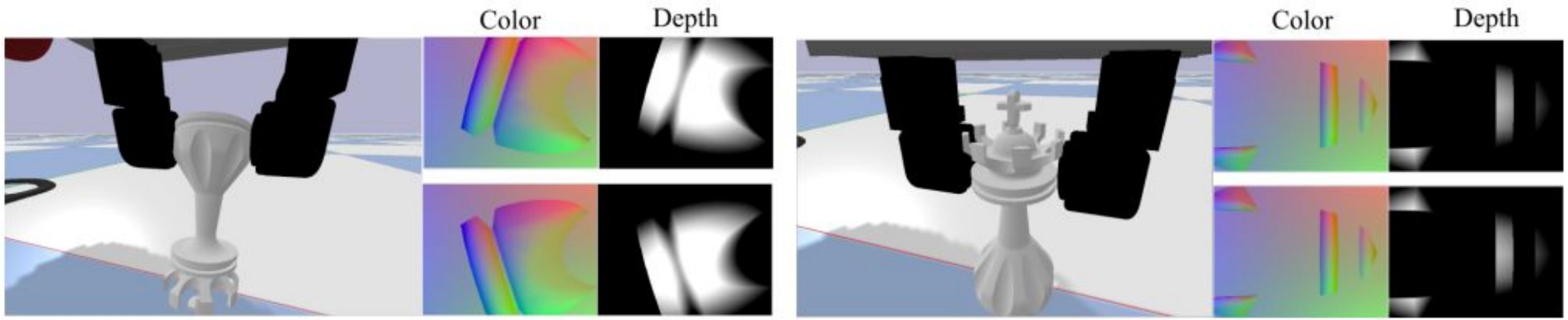

Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# Comparisons to Other Methods

| | | Tac2Pose SC mm (norm) | Tac2Pose PJ mm (norm) | Pixel SC mm (norm) | Pixel PJ mm (norm) | Classification SC mm (norm) | Pose SC mm (norm) |
|---|---|---|---|---|---|---|---|
| Long Grease |  | 26.6 (0.76) | 3.3 (0.09) | 32.8 (0.93) | 6.0 (0.17) | 33.3 (0.95) | 25.3 (0.72) |
| Snap Ring |  | 1.5 (0.10) | 1.4 (0.10) | 5.6 (0.39) | 2.2 (0.15) | 6.0 (0.42) | 5.9 (0.41) |
| Big Head |  | 7.8 (0.20) | 6.1 (0.16) | 27.6 (0.70) | 11.7 (0.30) | 35.0 (0.89) | 33.8 (0.86) |
| Cotter |  | 19.0 (0.49) | 19.6 (0.51) | 31.5 (0.81) | 36.7 (0.95) | 35.8 (0.93) | 38.1 (0.99) |
| Hanger |  | 6.6 (0.19) | 2.6 (0.07) | 31.3 (0.90) | 20.5 (0.59) | 34.2 (0.98) | 18.3 (0.53) |

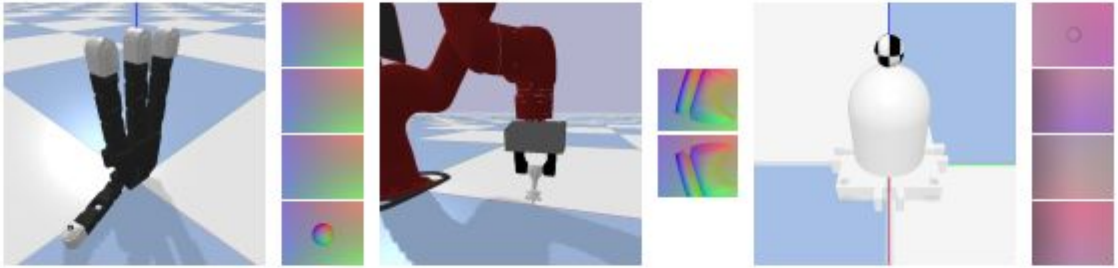
Tac2Pose: Tactile Object Pose Estimation from the First Touch, Bauza et al. , 2022

# TACTO: A Fast, Flexible, and Open-source Simulator for High-Resolution Vision-based Tactile Sensors



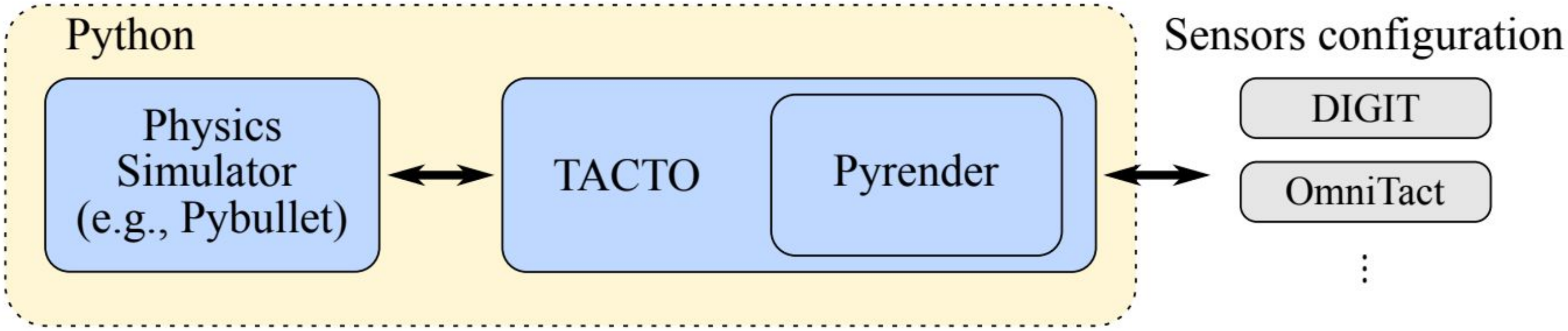

Wang, Shaoxiong, Mike Lambeta, Po-Wei Chou, and Roberto Calandra. *IEEE Robotics and Automation Letters* 7, 2022

# Motivation



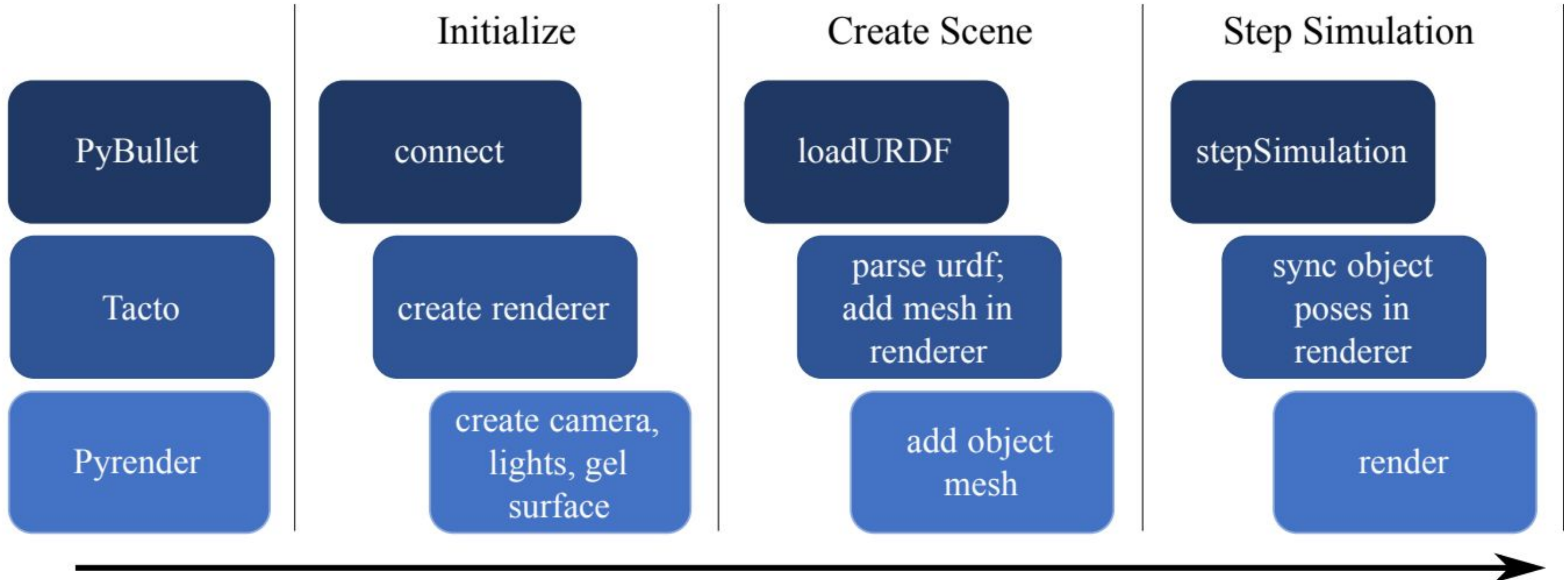
- Simulator for vision based tactile sensors
- Small Sim2Real gap
- Implements OmniTact and DIGIT sensors
- Value for different communities:
  - Hardware designers
  - Robotics
  - Machine learning

# Methods




Wang, Shaoxiong, et al. *TACTO: A Fast, Flexible, and Open-Source Simulator for High-Resolution Vision-Based Tactile Sensors*. 2020.

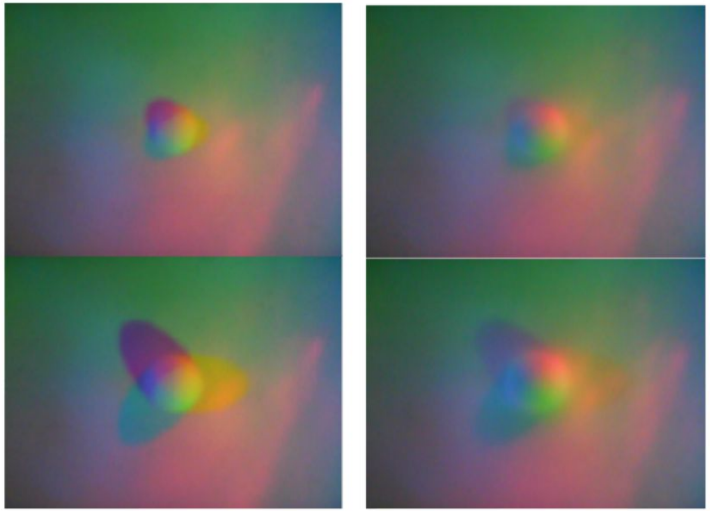
# Software Architecture



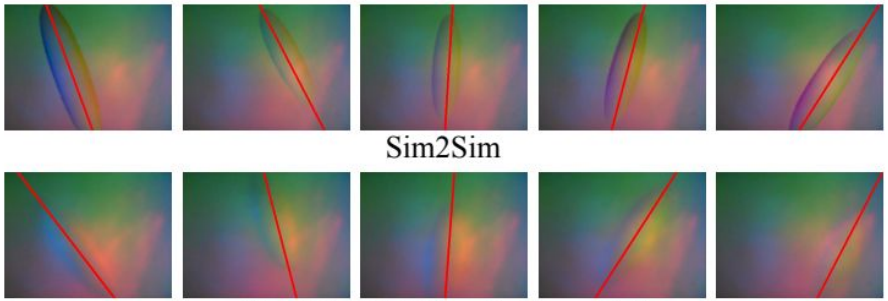

Wang, Shaoxiong, et al. *TACTO: A Fast, Flexible, and Open-Source Simulator for High-Resolution Vision-Based Tactile Sensors*. 2020.

# Sim vs Real

Sim Real



Pose estimation results

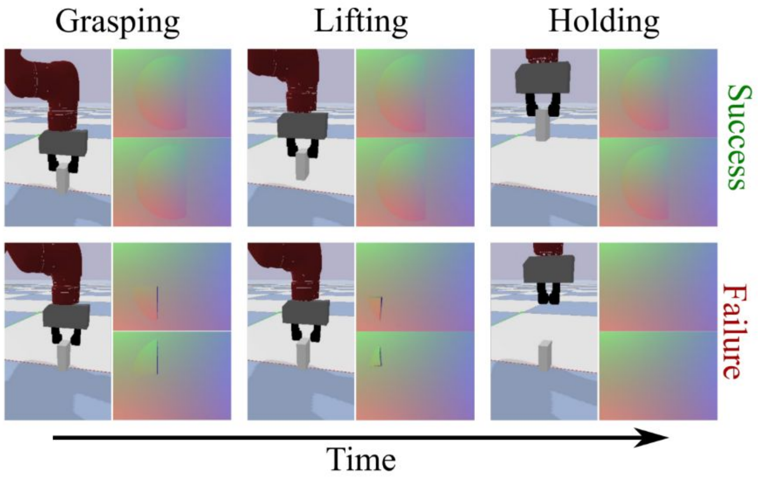


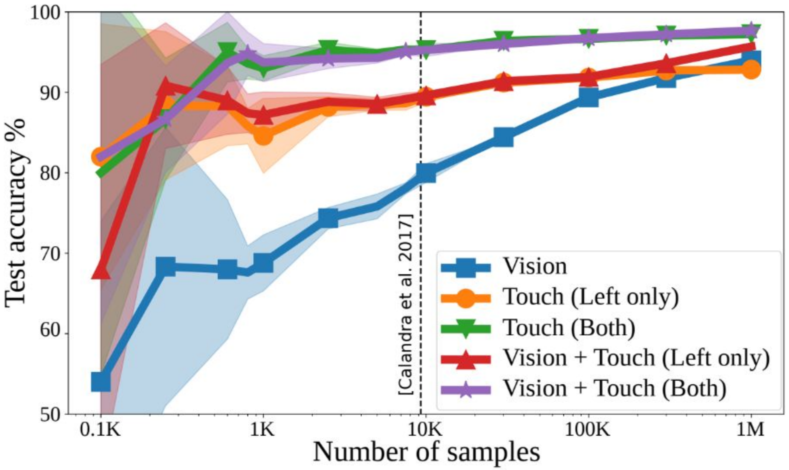
Sim2Sim

Sim + Real (128 real datapoints)

Wang, Shaoxiong, et al. *TACTO: A Fast, Flexible, and Open-Source Simulator for High-Resolution Vision-Based Tactile Sensors*. 2020.

# Results



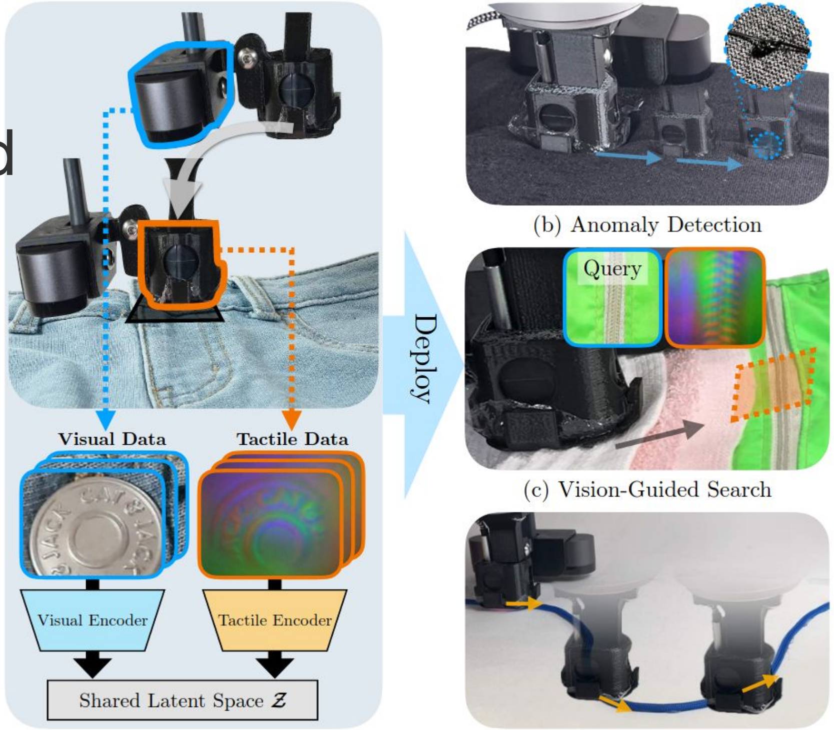




Wang, Shaoxiong, et al. *TACTO: A Fast, Flexible, and Open-Source Simulator for High-Resolution Vision-Based Tactile Sensors*. 2020.

Questions?

# Learning Self-Supervised Representations from Vision and Touch for Active Sliding Perception of Deformable Surfaces

- Align visual and tactile data
  - Train encoders to embed into a shared latent feature space.
  - Uses cross-modal contrastive loss
  - Object agnostic representation.




(a) Self-Supervised Learning

(b) Anomaly Detection

(c) Vision-Guided Search

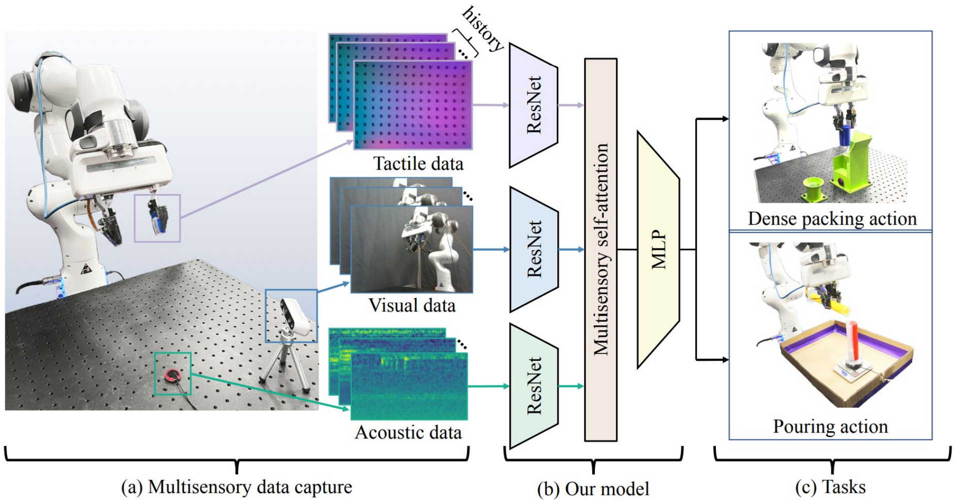
(d) Tactile Servoing

Kerr et al. (*ICRA 2023*)

Learning Self-Supervised Representations from Vision and Touch for Active Sliding Perception of Deformable Surfaces, Kerr et al. , 2022

# See, Hear, and Feel: Smart Sensory Fusion for Robotic Manipulation

- Explore using multi sensory data for performing Tasks
- Combines vision, tactile, and audio data

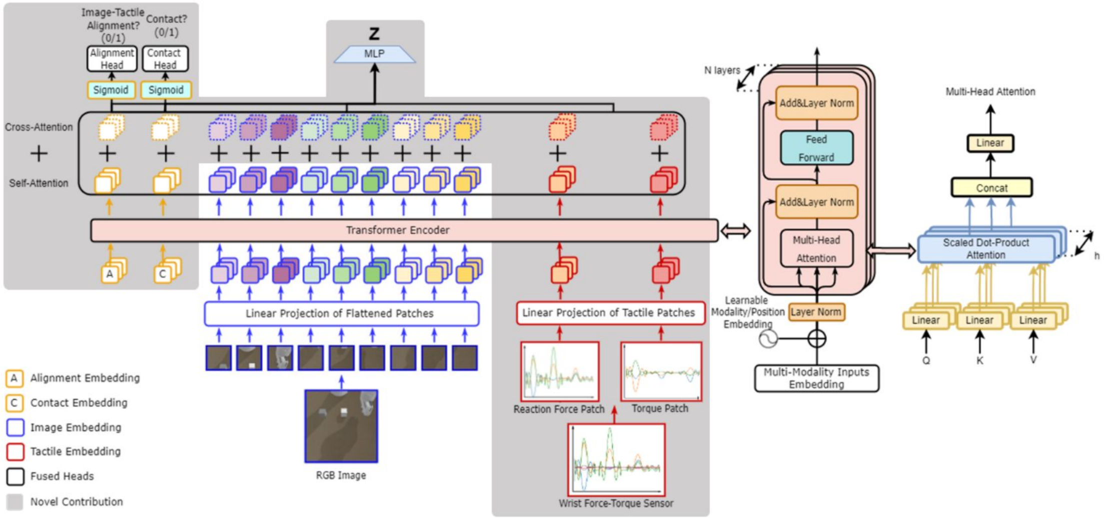



Li, Hao, et al. *See, Hear, and Feel: Smart Sensory Fusion for Robotic Manipulation*.. *DOI.org (Datacite)*, https://doi.org/10.48550/ARXIV.2212.03858.

Li et al. (*Corl 2022*)

# Visuo-Tactile Transformers for Manipulation (VTT)

- Modality Patches
- Self and Cross-Modal Attention
- Learned embeddings:
  - Contact
  - Alignment
  - Position/Modality
- Compressed Representation Head
- Combined Reinforcement Learning



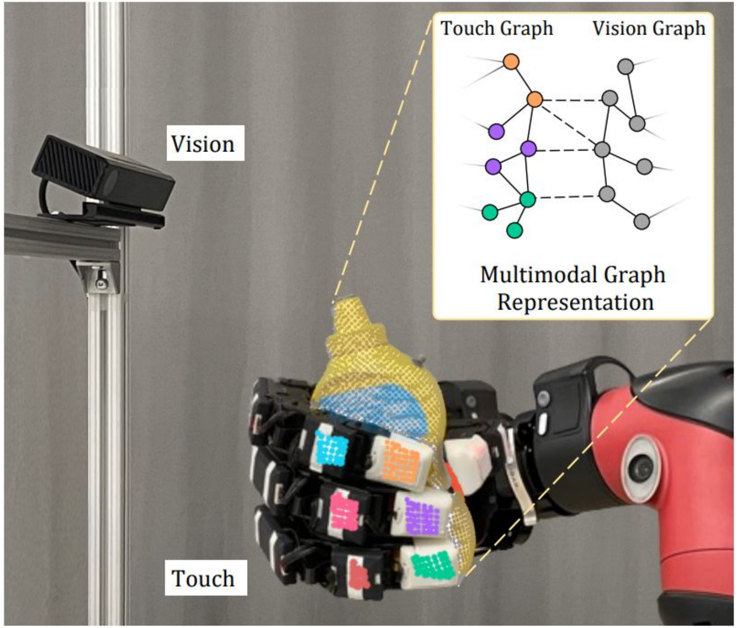

https://arxiv.org/abs/2210.00121

$$\ell_{VTT} = BCE_{logits}(MLP(Al_{head})), Al_{gt}) + BCE_{logits}(MLP(C_{head}), C_{gt})$$

Chen et al., CoRL 2022

# Hierarchical Graph Neural Networks for Proprioceptive 6D Pose Estimation of In-hand objects

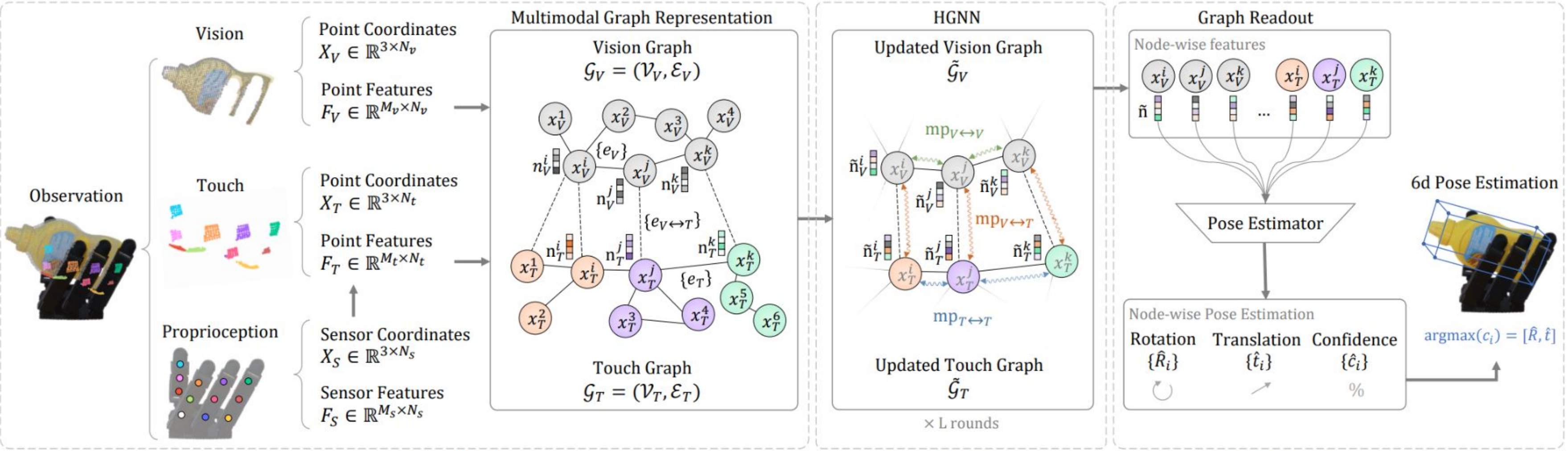
- HGNN combines vision and touch
- Geometrically informed 6D object pose estimation
- Multimodal graph message passing
- Proprioceptive information for in-hand object representation




https://drive.google.com/file/d/1UoOAbM37O4ylzQxXaQae2rK8-ijnWz4n/view

Rezazadeh et al. ICRA 2023

# Architecture




(a) Graph Encoder (b) Graph Processor (c) Pose Decoder

**Node-wise pose estimation loss**

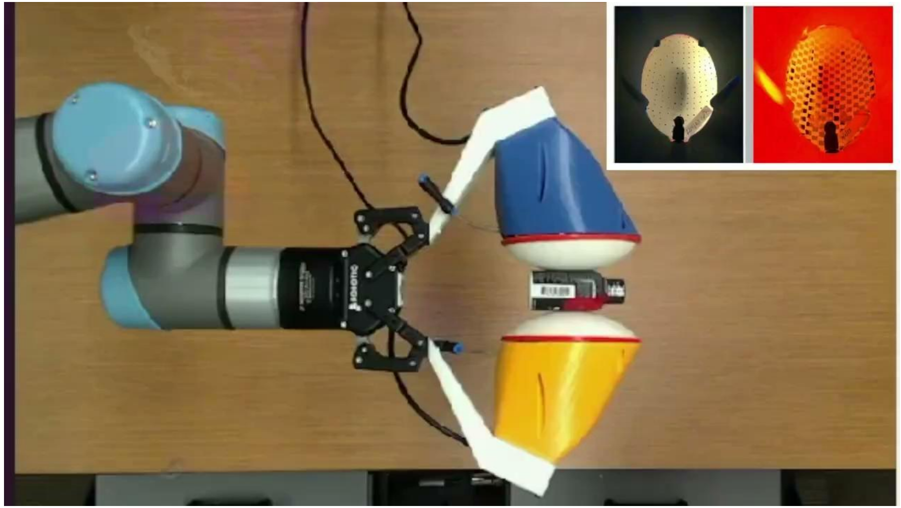
$$L_i^{\mathbf{n}} = \frac{1}{P} \sum_{p=1\ldots P} ||(R_{gt}x_p + t_{gt}) - (\hat{R}_i x_p + \hat{t}_i)||$$

**Total loss**

$$L = \frac{1}{K} \sum_{i=1\ldots K} (L_i^{\mathbf{n}} \hat{c}_i - \lambda \log(\hat{c}_i))$$

# Recap

- Tactile perception
- Signal categories
- Types of sensors
- Haptic vs Tactile sensing
- Gelsight
- Tac2Pose
- Tacto
- Tactile sensing for Deep Learning



# Questions?