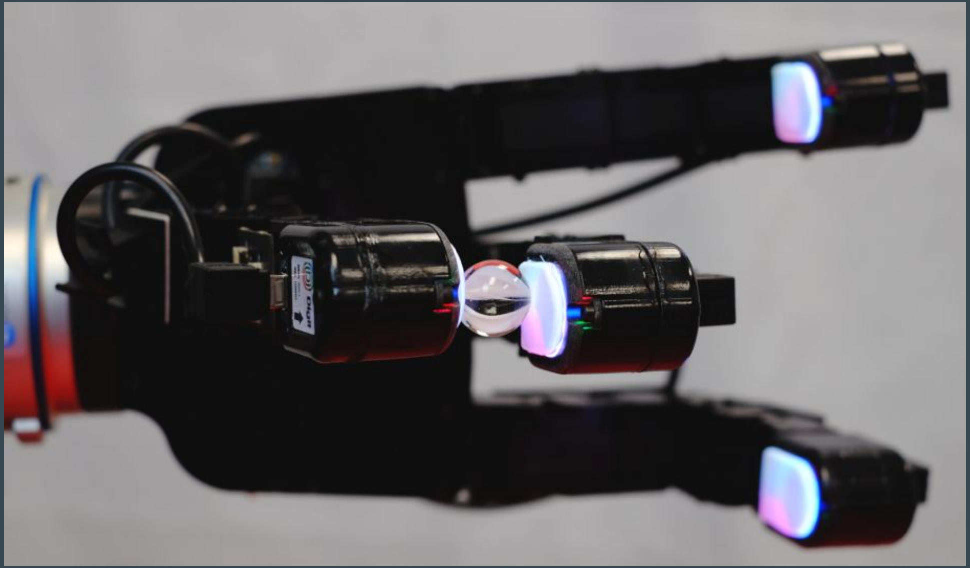
Depth and Contact Prediction

Varun Lagadapati, Leo Li, Eric Liau, Chibueze Nwade, Vaibhav Sanjay

Introduction

In this project, we use deep neural networks to predict 3D contact geometry from monocular images of a GelSight tactile sensor, which is also called vision-based tactile sensor. Specifically, we aim to acquire the inverse sensor model to reconstruct local 3D geometry from a tactile image.



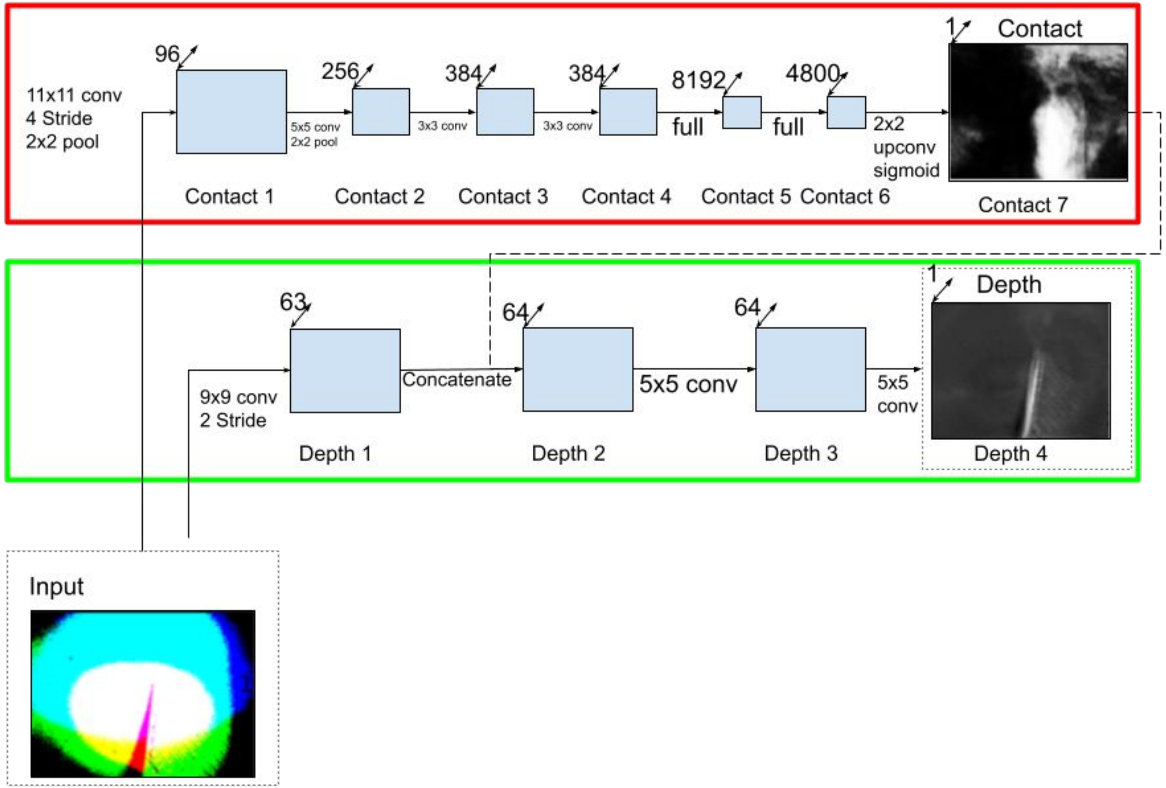
Literature Review

- Eigen et al. handle depth estimation by using two deep network stacks: Coarse and Precise [1]
- Laina et al. utilize CNNs to create depth mapping [3]
- Ma et al.'s model is a single deep regression network that learns directly from the RGB-D raw data [4]
- Godard et al. utilize a binocular stereo footage to train their singular convolutional neural network based on the constraints of epipolar geometry [2]
- Wang et al. find the depth of a scene, by first calculating the mapping of the image from color (RGB) to horizontal and vertical surface gradients and then applying a fast Poisson solver to integrate gradients and get the depth [5]

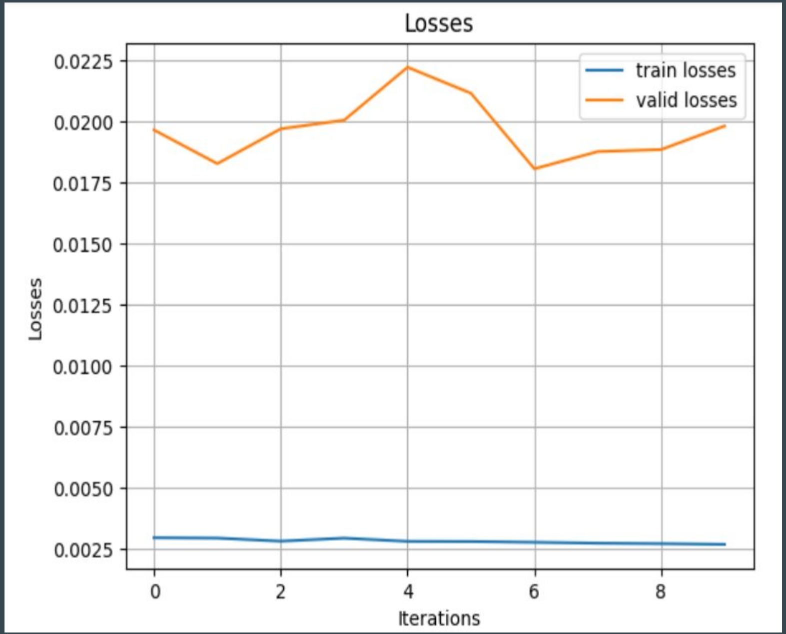
Experimental Setup

- 1000 training images
  - Tactile Image and Ground Truth Depth Image
  - Split 80/20 train/val
- Transforms:
  - Setting size to 240 x 320
  - Turning to PyTorch Tensor
  - Normalize Tactile Images
- Augmentations
  - We randomly rotated the images between -5 to 5 degrees
  - Random Horizontal Flip: Randomly flip the images horizontally with probability 50%

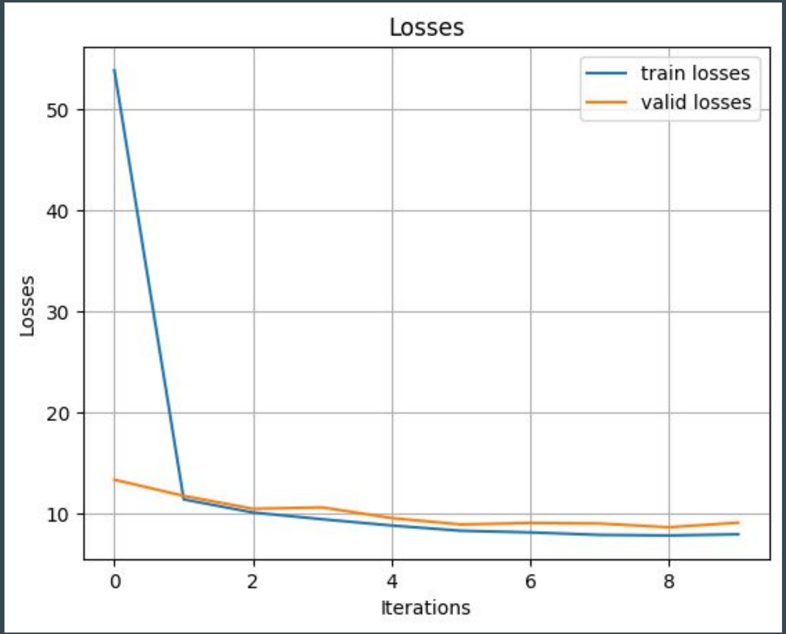
Methodology



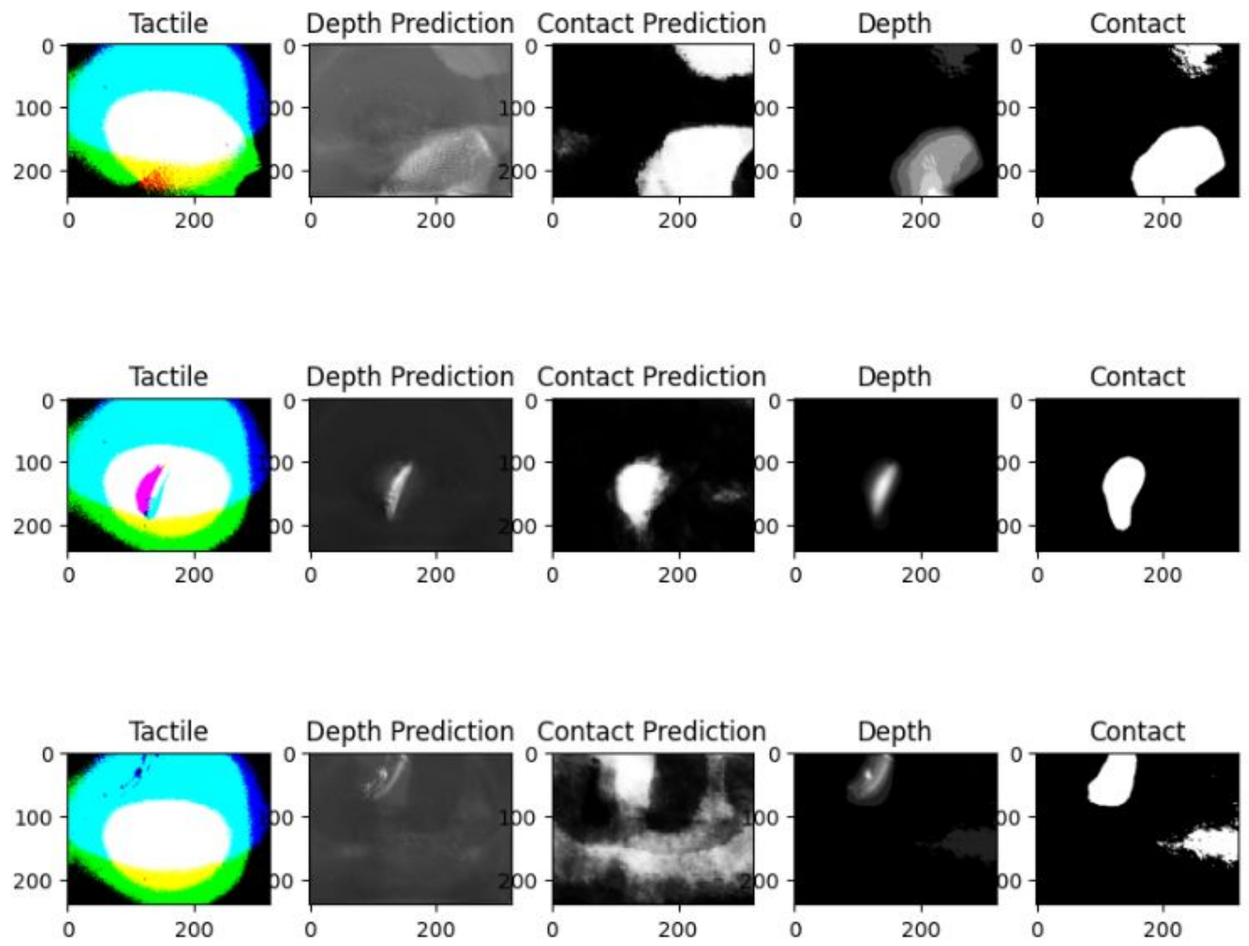
Results and Discussion



Loss Graph for ContactNet Model

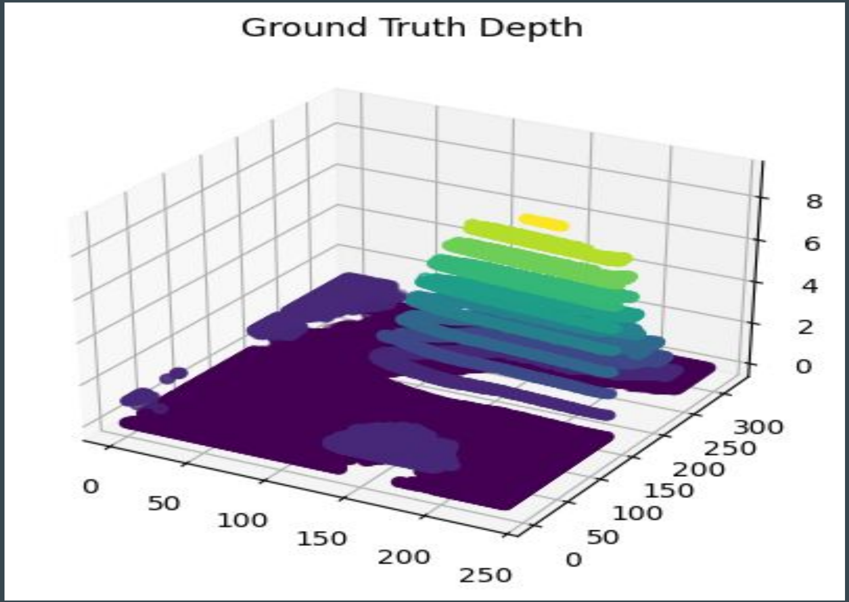


Loss Graph for TactileDepthNet Model

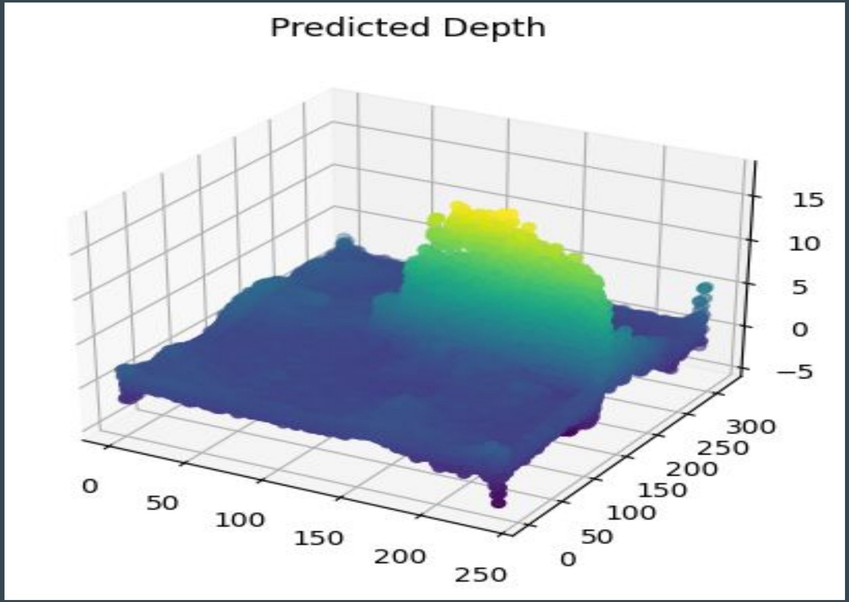


Tactile
Depth Prediction
Contact Prediction
Depth
Contact
0
100
200
0
200

Results and Discussion (Contd.)



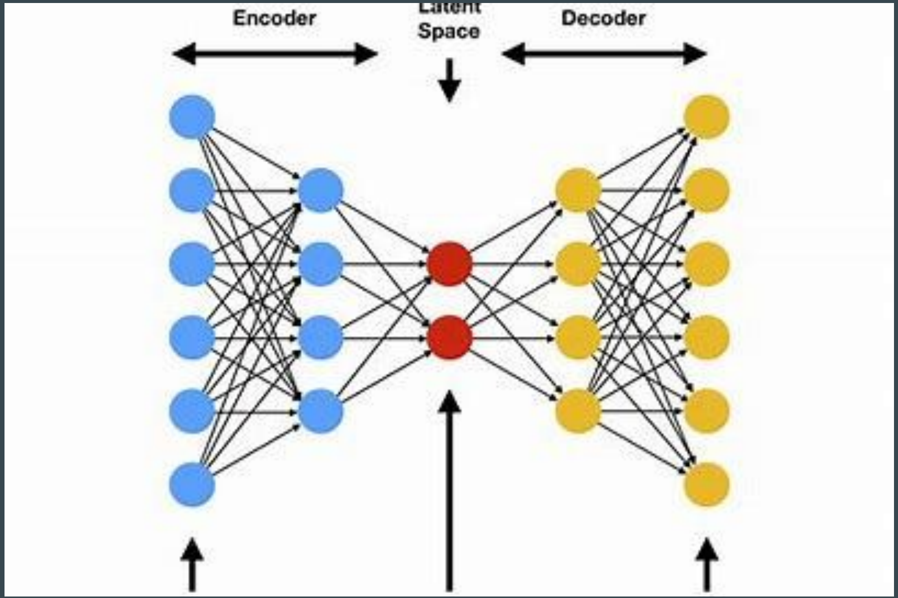
Ground truth depth corresponding to a tactile image



Predicted depth corresponding to a tactile image

Future Work

- Use different layer combinations.
    - Autoencoder idea
- Improve our models' robustness.
- Use different optimization techniques.



Takeaways

- Construction of Neural Networks in PyTorch
- Convolutional Neural Networks
- Effective loss functions used in depth estimation model training.
- Importance of image normalization
- The role of depth estimation in 3D Reconstruction.

References

[1] David Eigen, Christian Puhrsch, and Rob Fergus. "Depth Map Prediction from a Single Image using a Multi-Scale Deep Network". In: 2014. arXiv: 1406. 2283 [cs.CV].

[2] Cl ́ement Godard, Oisin Mac Aodha, and Gabriel J. Brostow. "Unsupervised Monocular Depth Estimation with Left-Right Consistency". In: 2017. arXiv: 1609.03677v3 [cs.CV].

[3] Iro Laina et al. "Deeper Depth Prediction with Fully Convolutional Residual Networks". In: 2016. arXiv: 1606.00373 [cs.CV].

[4] Fangchang Ma and Sertac Karaman. "Sparse-to-Dense: Depth Prediction from Sparse Depth Samples and a Single Image". In: 2018. arXiv: 1709 .07492v2 [cs.RO].

[5] Shaoxiong Wang et al. "GelSight Wedge: Measuring High-Resolution 3D Contact Geometry with a Compact Robot Finger". In: 2021.