End-to-end or not? Evaluation of the GelSight 3D reconstruction model pipelines based on end-to-end and non-end-to-end designs

Fengdi Zhang[1] , Siyue Zheng[2] , Qihui Ye[2]

[1]Tsinghua-Berkeley Shenzhen Institute, SIGS, Tsinghua University
[2]Institute of Biopharmaceutical and Health Engineering, SIGS, Tsinghua University

Learning From Data Final Project Poster, 2022 Fall

Abstract

- Four different model pipelines with end-to-end or non-end-to-end designs for the GelSight 3D reconstruction task were designed and compared to explore their performance and details of their differences.
- The non-end-to-end model pipeline that *appropriately* incorporate prior knowledge has better performance in terms of reconstruction accuracy, training speed, and the number of training samples required in the experiment.
- Inappropriate or even mutually exclusive prior knowledge can also lead to negative effects.

Motivation

GelSight sensor uses a camera and illumination sources from different directions to capture an image, which contains the 3D gradient information of the target surface.

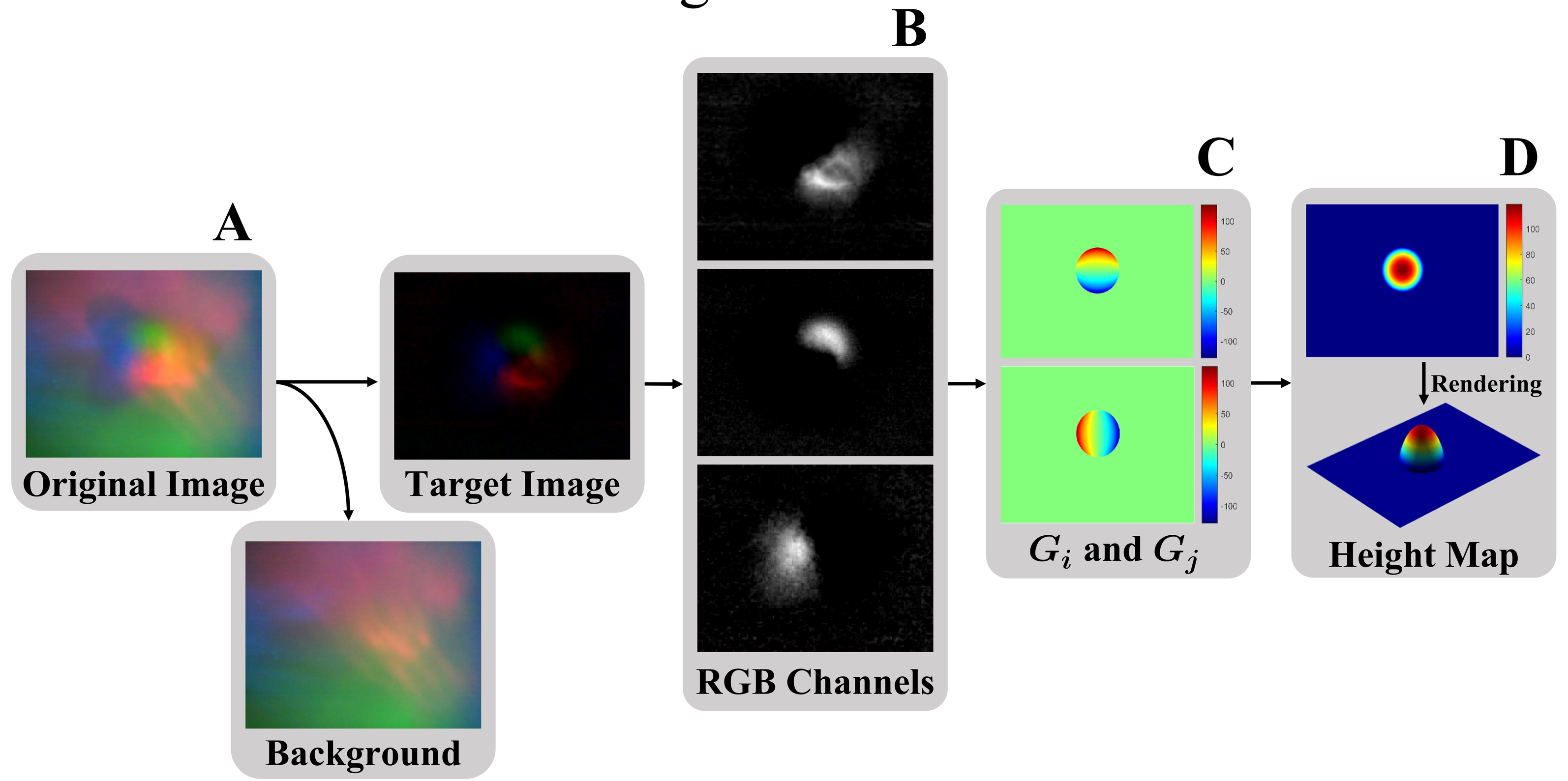

Fig. 1 The workflow of 3D reconstruction based on GelSight sensors.

Current GelSight-based 3D reconstruction model pipelines can be broadly divided into two categories: end-to-end model pipelines and non-end-to-end one. Both of them enrolls prior knowledge problems which set some questions:

- Which model pipeline performs better for GelSight-based 3D reconstruction tasks, the end-to-end or the non-end-to-end one?
- Is it always better and in which aspects?

Method

➢ EXPERIMENT DESIGN:

We conducted the following four experiments according to workflow map in **Fig. 1** to test the effect of different prior knowledge mapped with deep learning model from:

- A→D (no prior knowledge)
- B→D (A→B as known)
- A→C (C→D as known)
- B→C (A→B&C→D as known)

The height maps are reconstructed by minimizing an error function E using the 2D fast Poisson solver[1].

$$E = \sum_{i,j}\left[\left(\frac{\partial h}{\partial i} + G_i\right)^2 + \left(\frac{\partial h}{\partial j} + G_j\right)^2\right]$$

where G_i is the vertical gradient, G_j is the horizontal direction, and h is the height map to be reconstructed.

➢ DATASET:

The dataset was collected by a real GelSight sensor containing 143 samples.

➢ EXPERIMENTAL SETUP:

Deep learning models used in the experiments are the same U-Net-like structures to avoid additional variables.

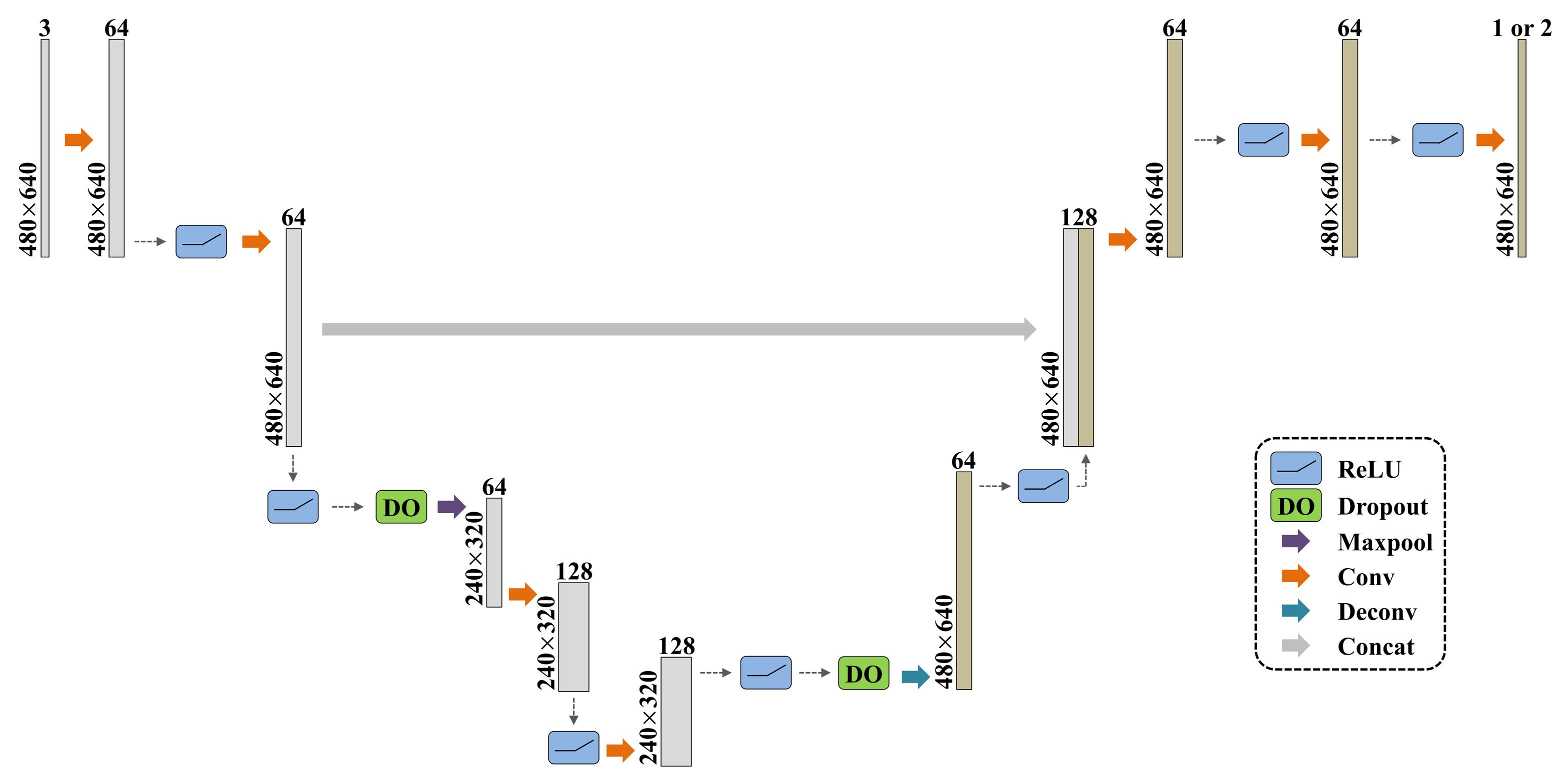

Fig. 2 The U-Net-like structure of the models used in experiments.

$$\mathcal{L} = \left\|I_{\text{out}} - I_{\text{target}}\right\|_2$$

The loss function used is an L2 regression loss function where I_{out} is the image output and I_{target} is the target image.

Numerical Results

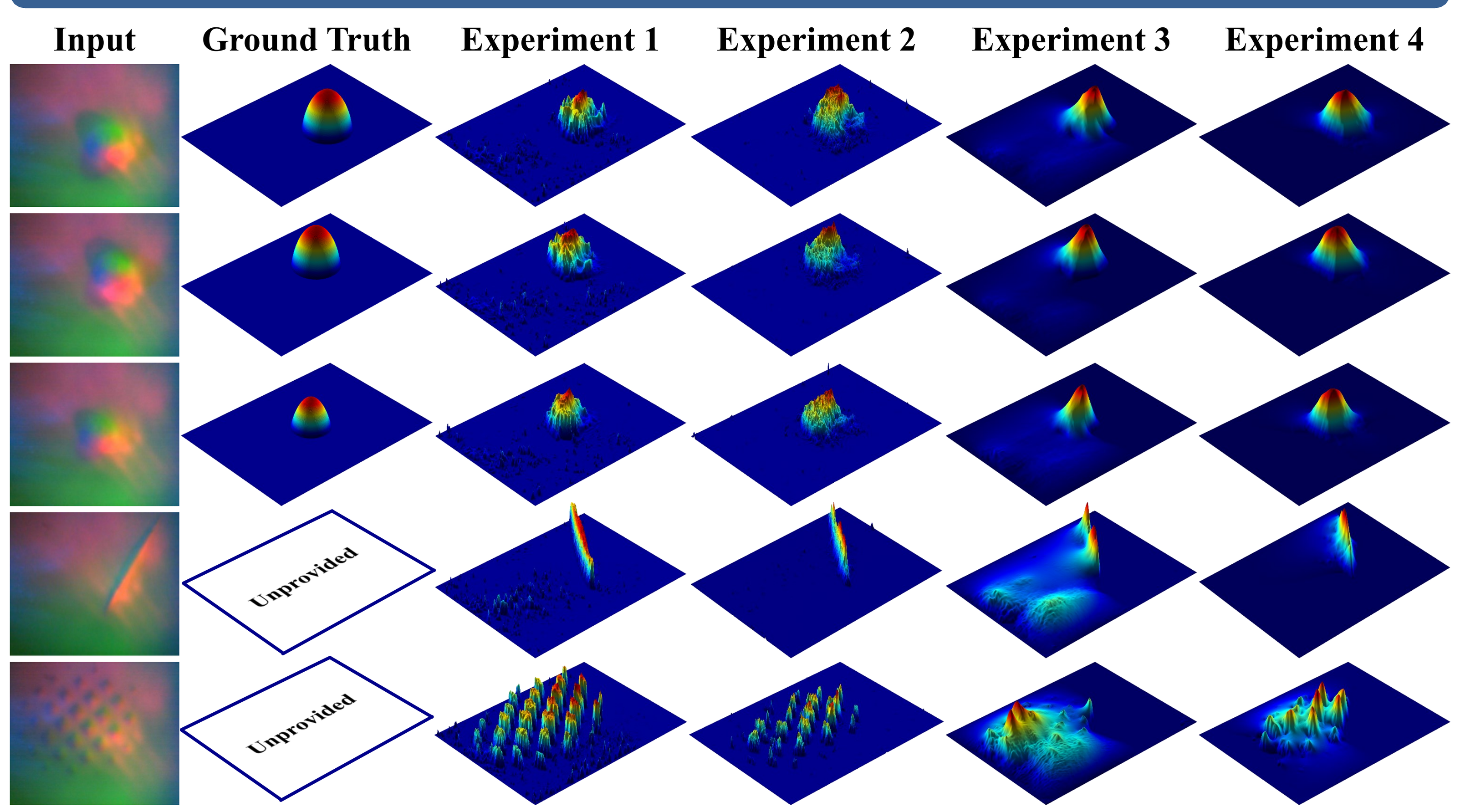

Fig. 3 Results of the representative samples in validation and test set.

The results shows that:

- The model pipeline with the most prior knowledge in **Experiment 4** achieves the highest reconstruction accuracy.
- Removing the background with prior knowledge significantly improves the reconstruction accuracy in **Experiments 3 and 4** but leads to a decrease in **Experiments 1 and 2**.

Table 1 The lowest validation RMSE loss for the four experiments

Experiment #	Mapped with DL models	Mapped with prior knowledge	RMSE
1	A→D	-	0.0708
2	B→D	A→B	0.0763
3	A→C	C→D	0.0707
4	B→C	A→B & C→D	**0.0616**

The model converges faster and achieves lower training loss also need less samples when more prior knowledge is added appropriately.

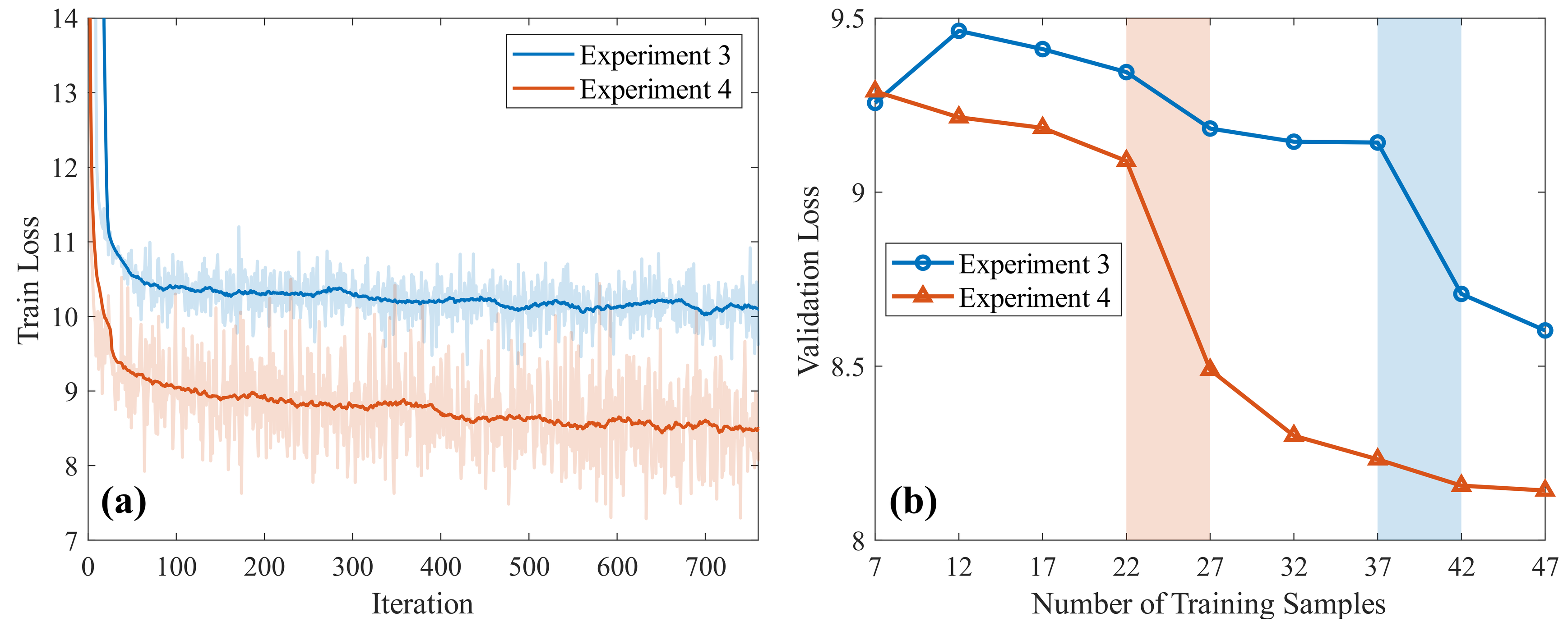

Fig. 4 (a) Training processes of the deep learning models in Experiments 3 and 4. (b) Performance using different numbers of training samples in Experiments 3 and 4.

Conclusion

- The non-end-to-end model pipelines that *appropriately* incorporate prior knowledge outperform the end-to-end models that do not incorporate prior knowledge.
- Inappropriate prior knowledge may have negative effects.
- Prior knowledge is not independent of each other but is related.
- When a set of mutually compatible and complementary prior knowledge is added to the model pipeline, the performance will be improved. Otherwise, the performance will be degraded.

Reference

[1] Li J, Dong S, Adelson E H. End-to-end pixelwise surface normal estimation with convolutional neural networks and shape reconstruction using GelSight sensor[C]. 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2018: 1292-1297.