Inferring Informative Grasp Parameters from Trained Tactile Data Models

Max Pflueger and Gaurav Sukhatme

***Abstract*—When choosing parameters for a tactile action, it is useful to be able to choose parameters that will optimize the discriminating power of the tactile data acquired during the action. We propose a way to extract this information about informative grasp parameters from a deep network trained to perform a classification task on the tactile data. We do this by defining a new cost function on the network that quantifies the certainty of the result, and optimizing the input on that cost function. Though our work on this technique is still in progress, we present some early experiments using this approach on the BiGS tactile grasping dataset, and propose experiments for further investigation of the technique.**

I. Introduction

New tactile sensors such as the SynTouch Biotac offer a rich and useful new stream of data for manipulation tasks. When using senors such as this it may be useful to choose grasp parameters (such as hand pose, contact points, grip pressure, etc.) that will maximize the usefulness or discriminating power of the data collected from the tactile sensors during a task. We propose that when using a deep network architecture to perform a task with tactile data (such as classifying grasp success or modeling object forces), it is possible to extract information about informative grasp parameters from implicitly learned patterns in the network.

This idea is motivated by work done with networks trained for image recognition tasks. In these networks interesting information can be extracted from the network about what it has learned and what it is looking for by optimizing the input image for a cost function elsewhere in the network. Simonyan et al. [5] showed how to generate images optimizing a particular label out of a deep CNN. We propose a similar approach using networks trained on tactile data and grasp parameters.

II. Methodology

Assume that we start with a network trained for a particular task. We can define a new loss function that quantifies the certainty of the network in its result. For example, in a classification task we can use the entropy of the category distribution. Normally in training a network we use back propagation to find the gradient in the parameter space, but we can also use the same technique to find the gradient on the input space. By stepping along that gradient, we can find inputs (including grasp parameters as they were described to the network) that are expected to produce high certainty results. With other cost functions we could produce other inputs, such as maximizing the probability of a certain result. This may seem like a more applicable cost function for a success/failure data set like BiGS, however in the the future we would like to evaluate this technique on regression datasets where the objective is clearly one of estimation accuracy.

Fig. 1. Cylinder grasping task used from the BiGS dataset.

III. Implementation and Training

A. BiGS Dataset Classifier

The BiGS dataset [2] is a collection of tactile and pose data traces for grasping attempts on 3 different objects. In our experiment we only use data from the cylinder object. The BiGS dataset provides 1000 grasp instances on the cylinder object, of which 54% are successful grasps. We further reduced the data by only training on hand pose, electrode, and AC pressure data. Electrode and AC pressure data was truncated to 200 ms immediately following the beginning of the lifting phase of the grasp. For hand pose we used a single pose value from the start of our selected time window.

We implemented our system in the TensorFlow framework [1] created by Google. We feed the electrode and pressure data into two different networks that each start with 2 layers of convolution and max pooling (convolution and pooling happen along the time dimension). The activations from those networks are then concatenated with the pose value and fed into two fully connected layers, the last of which has two nodes and is used to predict grasp success. We trained with a softmax cross entropy loss function using the TensorFlow Adam optimizer. For training we partitioned the dataset with 90% of examples for training and 10% for testing. This model achieved approximately 90% accuracy on the test set, although absolute classification accuracy is not the primary objective of this work. Accuracy on the training set is significantly higher (97-100%), so in further experiments we may test the degree to which this model is over fitting and see if that is impacting our use of the network.

B. Optimizing Grasp Pose for BiGS

To optimize the grasp pose we define a loss function as the entropy of the final softmax layer of the network:

$$\text{loss} = \sum_i y_i \log(y_i) \tag{1}$$

where y_i is the predicted probability of category i. We fixed all the model parameters at their trained values, and initialized the input values with one our testing examples. We then calculate the gradient of the classification entropy with respect to the input values, in particular the hand pose, and do a simple gradient descent along that. We are currently working to develop good evaluations for these updated hand poses to determine if and how much of an improvement they represent in the informativeness of the hand pose.

C. Future Regression Dataset

We are planning to collect a new dataset to explore the usefulness of this approach on a tactile regression problem. We believe this approach may be useful in finding optimal levels of grip strength to get useful data from the sensors. An example task here would be gripping a tool and estimating friction between the tool and a surface. Using this technique we might use the friction estimation network to choose a grip strength that will provide the best data for the estimation.

IV. Related Work

Though the BiGS dataset is a new release, there has been some work on force estimation and slip detection using a very similar experimental setup by Su et al. [6].

The technique of optimizing inputs through a pre-trained network has also been used by researchers at Google [3] to process images maximizing activations at various levels of the network, significantly for artistic effect. There is also some work to suggest that this approach may yield unpredictable and undesirable results, as shown by Nguyen et al. [4], so cautious evaluation is recommended.

This work is also related to the field of active sensing in robotics where the objective is to choose actions or sensor positions to maximize the usefulness of the information that will be gained.

V. Conclusion and Future Work

This project is still a work in progress and we are planning to further explore the usefulness of the technique in a couple of ways. Some easy network modifications, such as adding dropout during training, may improve our results by creating a more robust network.

The big next step for us will be collecting a new data set for a tactile regression/estimation problem. We believe this will provide a much more compelling application of this technique than a success/failure dataset such as BiGS. It may also be possible to use this technique in other estimation applications, such as adjusting parameters on visual, laser, or RF sensors.

Additionally, some open questions remain around how best to use this technique. While taking gradient steps, how should we treat the uncontrolled sensor data values (pressure and electrode values) as we step through the controlled parameter space (hand pose)? What is the best way to initialize the input values? This might be random values, but we could also use synthetic values, or a real data point. We will be continuing to explore these questions in our future work.

References

[1] Martın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv preprint arXiv:1603.04467*, 2016. URL http://arxiv.org/abs/1603.04467.

[2] Y. Chebotar, K. Hausman, Z. Su, A. Molchanov, O. Kroemer, G.S. Sukhatme, and S. Schaal. Bigs: Biotac grasp stability dataset. In *ICRA 2016 Workshop on Grasping and Manipulation Datasets*, Stockholm, Sweden, May 2016.

[3] Alexander Mordvintsev, Christopher Olah, and Mike Tyka. Inceptionism: Going Deeper into Neural Networks. http://googleresearch.blogspot.ch/2015/06/inceptionism-going-deeper-into-neural.html.

[4] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pages 427–436. IEEE, 2015.

[5] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

[6] Zhe Su, Karol Hausman, Yevgen Chebotar, Artem Molchanov, Gerald E Loeb, Gaurav S Sukhatme, and Stefan Schaal. Force Estimation and Slip Detection/Classification for Grip Control using a Biomimetic Tactile Sensor. In *IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, 2015.