CS230

Sequential Deep Learning for Terrain Classification from Raw Tactile Data of Small-Legged Robots

Hojung Choi
Department of Mechanical Engineering
Stanford University
hjchoi92@stanford.edu

Abstract

The ground contact interaction can have a significant effect on the efficiency of locomotion for small legged robots. Adapting the robot's gait parameters appropriately to the terrain can substantially improve performance. In this study, we present sequential deep neural network classifiers to distinguish terrain from data collected by tactile sensor arrays attached to the legs of a mobile robot, SAIL-R. The sequential network extracts spatio-temporal information from the tactile data. SVM is used as a baseline to investigate the performance of three different sequential models: 1D-Convolutional network, LSTM, and a combination of both. Results show that the use of sequential models achieves a significantly higher accuracy compared to SVM and the combination of CNN and LSTM gives the best accuracy due to the consideration of long time dependencies on high level features.

1 Introduction

For small ground robots and animals, the interaction between the ground and leg can have substantial effects on the efficiency of robustness of mobility during locomotion. For example, the cost of transportation is increased when running quickly on sand as the deformation of sand results in dissipation of energy. In contrast, when running on slippery tile, slippage becomes the issue. Thus, for achieving efficient and robust locomotion, employing the right gait strategy is critical.

Using numerous mechanoreceptors on their limbs, small animals and insects sense the type of terrain they are moving on and adjust their gait strategy to optimize locomotion, such as the shovel nosed lizard running on sand. Gait adaptation of ground robots using tactile sensing can be instrumental in minimizing cost of movement and enhancing maneuverability.

In this study, we investigate the benefits of using sequential deep learning models to classify terrain and terrain type based on information regarding ground-robot interaction and robot state. In particular, we use data collected from the SAIL-R robot (Fig. 1 (a)) built at the Biomimetics and Dexterous Manipulation Laboratory (BDML) at Stanford University[8][9].

2 Related work

Ranging from humanoids to bio-inspired robots, the integration of tactile sensors for robotic locomotion and contact perception has been explored extensively over the years[6, 5, 4, 2]. For better understanding the temporal implications of tactile data, recursive neural networks on data from a force torque sensor was investigated[1], but the data lacked spatial representation of the contact

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

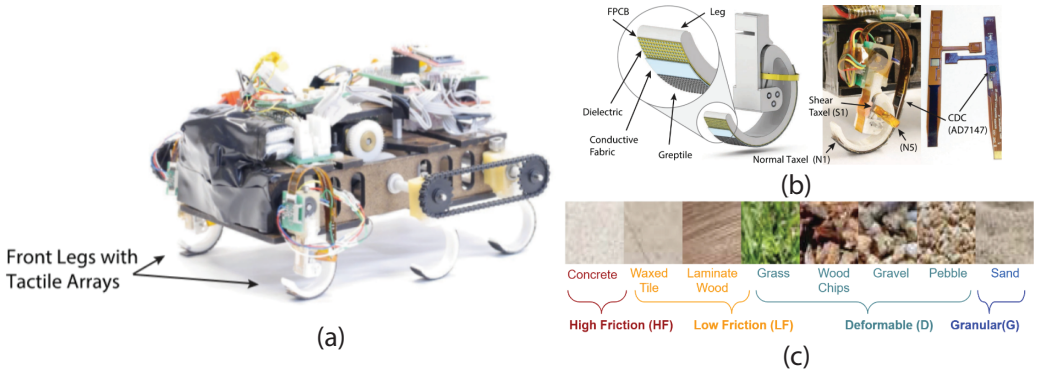


Figure 1: (a) SAIL-R robot. (b) Tactile sensor design. (c) Terrain and high-level classes

area. The application of deep learning for extracting spatio-temporal information from the fingertip mounted GelSight tactile sensor[10] was investigated in [11].

Previous work on terrain classification and gait adjustment with SAIL-R[9] chose an SVM with 39 hand-designed features derived from tactile data, control input, robot state. This resulted in average accuracy of 82.6% on terrain class but showed low accuracy($< 70\%$) on high-friction and low-friction terrains as some of the hand-designed features were not useful for classification. In [3], the application of 1D convolutional network on SAIL-R tactile data has been explored but the model was innately incapable of considering long time dependencies. We propose to employ sequential deep learning models capable of accounting for long time dependencies such as LSTM to learn useful spatio-temporal features from raw tactile data to perform terrain classification with higher accuracy.

3 Data

3.1 Data Collection

The two tactile sensors made of flexible PCB in each of the front leg of the SAIL-R robot has been used to collect data. (Fig. 1(b)) The sensor consists of an array of 6 capacitive taxels ($5 \times 5mm$) where 5 is installed along the C-shaped leg for normal force measurement and another is located at the center of the arc for shear force measurement.(Fig. 1(b)) This design allows the decoupled measurement of normal force and shear force. The sensor array uses a 16 bit CDC microcontroller (Analog Devices AD7147) to read sensor array output at 217Hz.

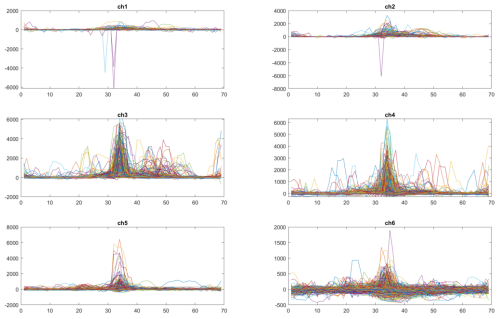
The terrains for locomotion experiments were selected to span a variety of physical properties such as surface friction, stiffness, and dissipation as shown in (Fig. 1(c)). They were grouped into 4 different classes based on their similarity of properties: high friction and high stiffness(HF), low friction and low stiffness(LF), deformable(D), and granular(G).

3.2 Data Preprocessing

The sensor readings from 6 channels(Fig. 2(a)) were segmented and transformed to a $6 \times 1 \times 69$ $[C \times H \times W]$ tensor to be used as input, representing one step that is 32ms long. The 5981 examples containing an equal proportion of each terrains were divided to 4189, 896, 896 subsets for training, validation, test sets, respectively. From data collected during locomotion, 33 hand designed features were derived as shown in (Fig. 2(b)) for SVM which is the baseline. Some of the features such as 'Individual normal taxel force peak amplitude' were inferable from the raw data while others such as 'Motor RPM' or 'average input current' were not as they are control inputs and robot states.

4 Methods and Experiments

In order to extract spatio-temporal information from the raw sensor data and use it for classification, we have implemented three different sequential neural network structures. Each network takes all 6



(a)

Index	Feature Name
1	Peak amplitude of sum
2	Area under curve of sum
3	Gait parameter – Motor RPM
4	Gait parameter – Ratio T_s to T_c
5	Average amplitude of sum
6	Gait parameter - ω_{slow}
7-11	Individual normal taxel force peak amplitude
12-16	Ratio of individual taxel force peak to peak sum
17-19	Min, max, average shear taxel force
20	Motor RPM at peak sum
21	Input current at peak sum
22-23	Input current average, range
24-28	Average normal force on individual taxels
29-33	Motor RPM at peak force for individual taxels

(b)

Figure 2: (a) Overlapped segmented data. CH6 is the shear taxel. (b) 33 hand designed features

channels from the $6 \times 1 \times 69$ tensor to consider the spatial correlation of the tactile data at a given instance and sequentially steps through the 69 readings per channel to extract temporal correlation. Unlike [9] which performed classification only on class using SVM, we also classify terrain to investigate whether the neural network can distinguish terrains in the same high-level class that are physically similar. For each network structure, the 12 non-inferable features on control input and robot state are concatenated to the learned features as different inputs on the same terrain can result in disparate sensor readings, confusing the model when these features are absent. Over 50 tests were done searching the best network structures with varying hyperparameters such as learning rate, weight decay rate, dropout rate, number of CNN layers, number of LSTM layers, and number of fully-connected layers. The presented structures below has shown the best results so far.

4.1 1D Convolutional Network

The 1D convolutional network consists of 4 convolutional layers where the first layer expands the channels from 6 to 24 and the channel size is kept constant as shown in (Fig. 3(a)) [3]. A kernel size of 1 by 5 is used to capture the temporal information from the 69 readings per channel. A stride of 1 with zero padding was applied to keep the width constant. Each layer was passed through a ReLU activation function and a max pooling reducing the width by half. A dropout at a rate of 0.5 was carried out for regularization during training. The resulting 1368 learned features was concatenated with 12 non-inferable features to be passed to a fully-connected layer. The two fully-connected layers used ReLU activation to convert the learned features into classification results on class or terrain. For training, the Adam optimizer with a learning rate of 1e-4 and cross entropy loss was chosen.

$$Cross\ Entropy\ Loss = -log(\frac{exp(x[class])}{\Sigma_j exp(x[j])})$$

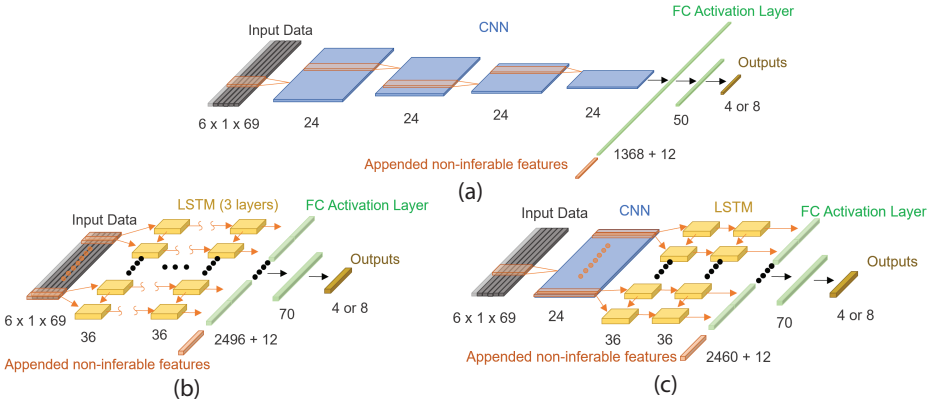


Figure 3: (a) 1D Convolutional Network. (b) LSTM. (c) 1D Convolutional Network + LSTM

4.2 LSTM

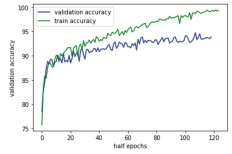
The LSTM network is composed of 3 LSTM layers(Fig. 3(b)) where each iteration takes one temporal instance of the 6 channel sensor reading and outputs a 36 dimensional hidden unit. During the learning step, Adam optimizer with a learning rate of 1e-4 was used. For regularization, a weight decay of 1e-5 was applied to the optimizer and a dropout at a rate of 0.5 was used between LSTM layers. Each of the 36 hidden unit outputs from the LSTM layer is concatenated to each other along with the non-inferable features to be passed through two fully-connected layer with a $\tanh$ activation for classification. Cross entropy was chosen as the loss function.

4.3 1D Convolutional Network with LSTM

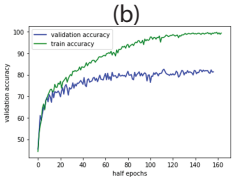
In order to investigate the temporal relations of higher level spatial features of the raw tactile data, a network combining 1D CNN and LSTM was used(Fig. 3(c)). The 1D convolutional layer is the same as the first layer of the aforementioned 4 layer 1D convolutional network including max pooling and dropout. The LSTM and fully-connected layer is a 2 layer version of the aforementioned LSTM network with other hyperparameterson being the same.

Model	Class	Terrain
SVM Baseline	84%	70%
1D Convolutional Network	93%	76%
LSTM	94%	79%
1D ConvNet + LSTM	94%	83%

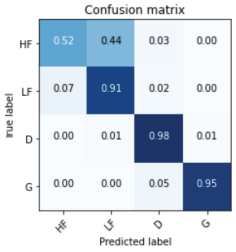
(a)



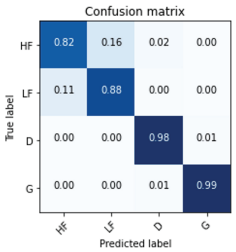
(b)



(c)



(d)



(e)

Figure 4: (a) Summary of results. (b) Class accuracy plot (c) Terrain accuracy plot (d) SVM class confusion matrix (e) 1D ConvNet + LSTM class confusion matrix

5 Results and Discussion

5.1 Results

For all classification approaches, accuracy on class is notably higher than on terrain as reported in (Fig. 4(a)). The confusion matrices for each model in (Fig. 4(d)(e)), (Fig. 5),(Fig. 6),(Fig. 7) show that the classifiers are generally confused with terrains in the same high-level class while they perform well in distinguishing between classes. One exception is the relatively lower accuracy for high-friction(HF) where the models mistake HF data as LF but not vice versa.

For both classification tasks, models applying sequential models show a significant improvement on accuracy compared to the baseline SVM approach where the model showing the highest performance is 1D ConvNet + LSTM. (Fig. 4(d)(e)) reports an improvement of 30% in classification for HF for the 1D ConvNet + LSTM model compared to the baseline SVM. While sequential models do improve in individual terrain classification as well, the degree is marginal compared to improvements seen in class prediction tasks.

Examining the accuracy plot for the highest performing network as shown in (Fig. 4(b)(c)), both train and validation accuracy shows a steep rise in earlier epochs. The train accuracy reaches above 99% for both class and terrain predictions while the validation accuracy plateus at 94% for class and 83% for terrain.

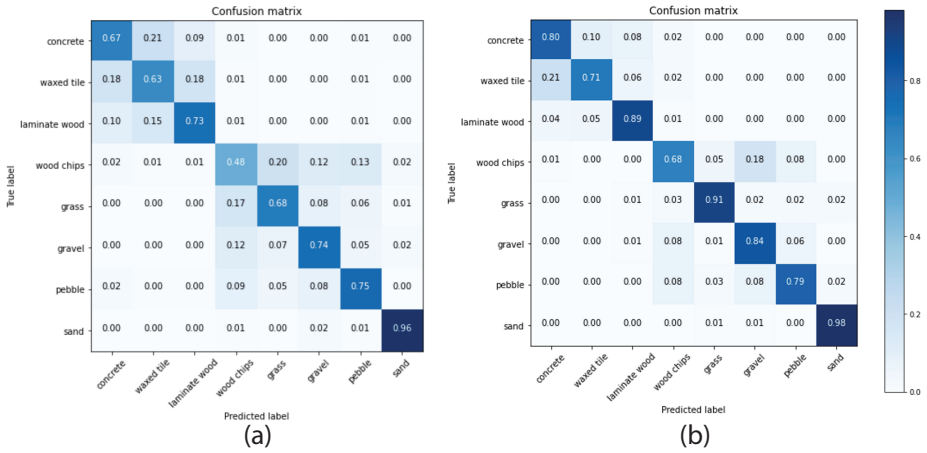


Figure 5: Terrain confusion matrix. (a) SVM results (b) 1D ConvNet + LSTM results

5.2 Discussion

The higher accuracy for both class and terrain prediction tasks in models using sequential network can be attributed to their ability to learn useful features for classification. Only a few hand-designed features among the full feature set is reported to be useful for prediction using SVM [9] While the learned useful features may allow all three sequential models to achieve a similar accuracy of 94% for high-level class prediction, it may be the consideration of long time dependencies of data which LSTMs are inherently capable of that improves accuracy in terrain classification. The highest terrain accuracy of the '1D ConvNet + LSTM' model shown in (Fig. 4(a)) suggests that the temporal development of higher level features can be more effective than the temporal development of the raw data. It is also noteworthy that while the 'LSTM' and '1D ConvNet + LSTM' models have the same class accuracy, '1D ConvNet + LSTM' performs better of HF at the cost of lower LF accuracy. Considering how during the experiment deeper '1D ConvNet + LSTM' resulted in a similar confusion matrix as the 'LSTM' model, it might be the case that the optimizer found a different local minima. For actual robotic applications, the '1D ConvNet + LSTM' can show better gait performance as class prediction accuracy is reasonably high (>80%) among all classes. Class accuracy is higher than terrain for all models because terrains in the same class share similar physical properties and is innately more challenging to distinguish.

Some of the difficulty in achieving a higher accuracy on either tasks may stem from the quality of data. For instance, as mentioned in [9], the HF class has a coefficient of friction of $\mu = 1.0$ compared to $\mu = 0.5$ for LF, which is not a large difference. The problem of using innately less separable data is exacerbated with sensor noise and limited sampling rate where the sudden spikes in sensor reading while contacting rigid and high friction surfaces may not be clearly captured. Undesirable contact conditions such as slippage between the robot leg and terrain may also have made some of the HF data similar to other classes such as LF. Class imbalance also can create bias in the trained model as the dataset contains an equal number of each terrain but each class encompasses different number of terrains. The training results shown in (Fig. 4(b)(C)) also suggests that the dataset size is not large enough as various different regularization parameters and methods failed closing the gap between train and validation accuracy.

6 Conclusion/Future Work

The presented deep learning networks all outperform the baseline SVM, where the combination of CNN and LSTM showed the best performance. Further improvements can be made by collecting better and more data using sensors that have higher spatial-temporal resolution with low noise. The classes of terrains can be selected more carefully such that their physical properties are more separable. Also, other learning frameworks such as the Transformers[7] that has been gaining popularity recently.

7 Contributions

The author would like to express his sincerest gratitude to Rachel Thomasson whose keen insights and observations have led to better understanding the results and reasons behind it. The author would also like to thank Dr. Taemyung Huh and Dr. Alice Wu for sharing their raw data from the SAIL-R robot and help making this project a reality.

8 Appendix

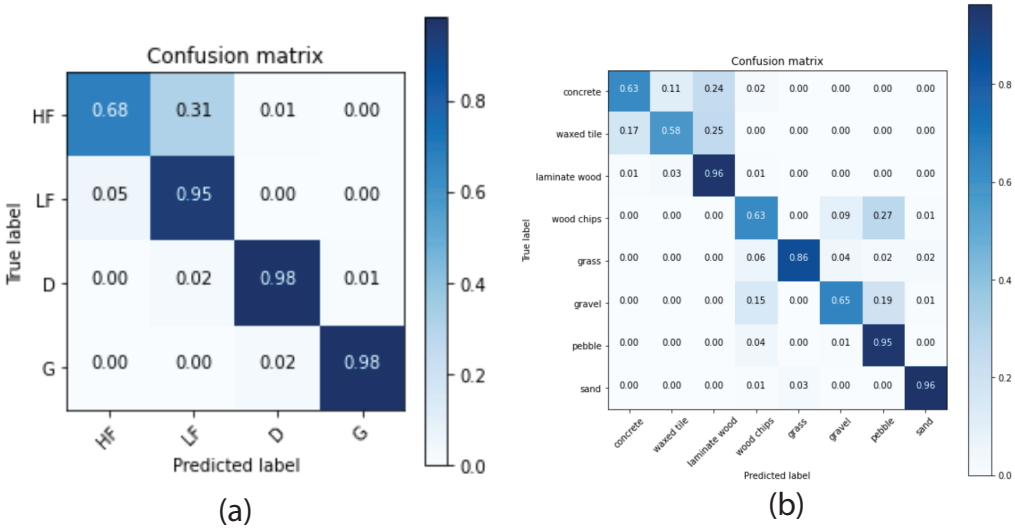


Figure 6: Results from the 1D convolutional network (a) Class (b)Terrain

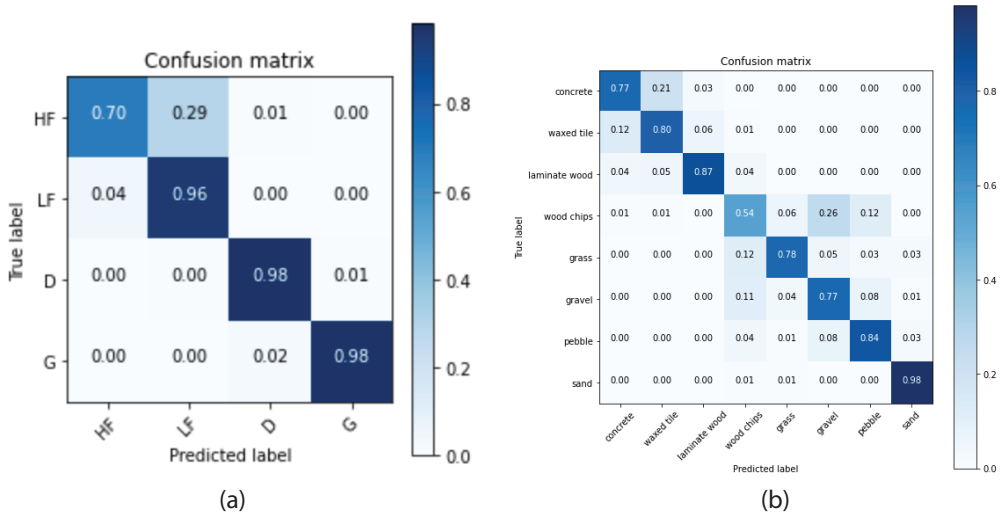


Figure 7: Results from LSTM (a) Class (b) Terrain

References

[1] Jakub Bednarek et al. “What am I touching? Learning to classify terrain via haptic sensing”. In: *ICRA* (2019), pp. 7187–7193.

[2] W. Bosworth et al. “Robot locomotion on hard and soft ground: Measuring stability and ground properties in-situ”. In: *Int. Conf. Robot. Autom.* (2016), pp. 3582–3589.

[3] H Choi and Rachel Thomasson. “Terrain Classification for Small Legged Robots Using Deep Learning on Tactile Data”. In: *Stanford CS229 Spring Project* (2020).

[4] M.Y. Chuah and S. Kim. “Enabling force sensing during ground locomotion: A bio-inspired, multi-axis, composite force sensor using discrete pressure mapping”. In: *Sensors J* 14 (2014), pp. 1693–1703.

[5] K. Hirai et al. “The development of Honda humanoid robot”. In: *Int. Conf. Robot. Autom.* 2 (1998), pp. 1321–1326.

[6] S. Hirose. “A study of design and control of a quadruped walking vehicle”. In: *Int. J. Robot. Res.* 3 (1984), pp. 113–133.

[7] Jeeheh Oh, Jiaxuan Wang, and Jenna Wiens. “Learning to exploit invariances in clinical time-series data using sequence transformer networks”. In: *arXiv preprint arXiv:1808.06725* (2018).

[8] X Alice Wu et al. “Integrated ground reaction force sensing and terrain classification for small legged robots”. In: *IEEE Robotics and Automation Letters* 1.2 (2016), pp. 1125–1132.

[9] X Alice Wu et al. “Tactile Sensing and Terrain-Based Gait Control for Small Legged Robots”. In: *IEEE Transactions on Robotics* (2019).

[10] Wenzhen Yuan, Siyuan Dong, and Edward H Adelson. “GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force”. In: *Sensors* 17 (2017).

[11] Wenzhen Yuan et al. “Active Clothing Material Perception using Tactile Sensing and Deep Learning”. In: *ICRA* (2018).