Terrain Classification for Small-Legged Robots using Deep Learning on Tactile Data

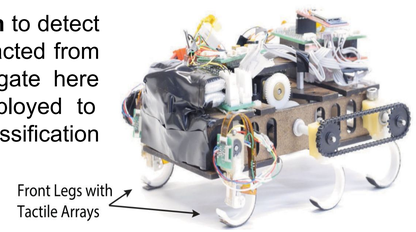
Hojung Choi and Rachel Thomasson
{hjchoi92, rthom}@Stanford.edu
Department of Mechanical Engineering, Stanford University

Motivation

For ground **mobile robots**, the interaction between the robot and its environment can have a profound impact on the efficiency and robustness of locomotion. Adjusting the robot's gait according to the terrain is crucial, especially for small and lightweight robots.

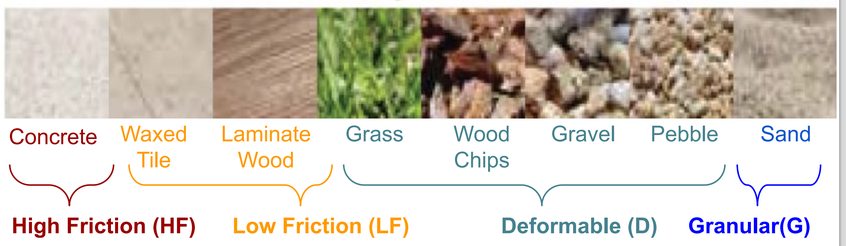
SAIL-R is a small six-legged robot built at Stanford University. While it does not have an onboard vision system to use for **terrain classification**, it is equipped with **capacitive tactile sensors** on it's legs.

Previous work used an **SVM approach** to detect terrain class using features hand extracted from these tactile sensors [1]. We investigate here whether **deep learning** can be employed to improve accuracy on the terrain classification task.



Dataset and Features

Raw Sensor Data:
Time-series data from tactile sensors on one of SAIL-R's legs
→ 5 Normal Taxels + 1 Shear Taxel
Each sample corresponds to one step of the robot on a specific terrain
Labels:
8 Terrains are further divided into 4 higher-level classes:



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

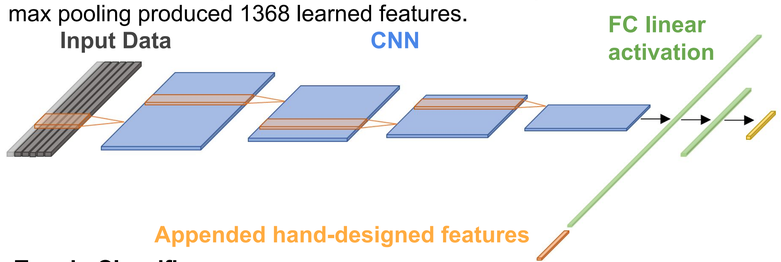
Hand-Designed Features:
33 hand-designed features from [1] were used as inputs to an SVM baseline.

The use of a neural-network architecture that combined these features with learned features was explored, and the impact of this addition on performance was examined.

Methods

1D Convolutional Neural Network Feature Extractor:
The time-dimension of sensor data taken during a dynamic robot running encodes important information about the properties of the terrain. A 1D CNN [2] was used to make use of temporal information for feature extraction from the raw tactile sensor data. Each sample was transformed into a 6x1x69 [CxHxW] tensor, where the 6 sensor taxels comprise the 6 channels.

Four convolutional layers with ReLU activations, dropout with rate=0.5, and max pooling produced 1368 learned features.



Terrain Classifier:
Two fully connected layers, using ReLU activation, transform the learned features into a terrain prediction. When examining the effect of added hand-designed features, they were concatenated with the learned features before being input to the first fully connected layer.

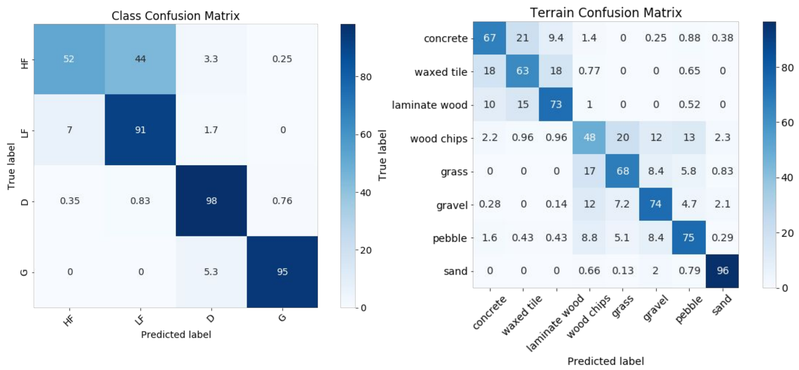
The loss function used was *Cross Entropy Loss*:

$$-\log\left(\frac{\exp(x[class])}{\sum_j \exp(x[j])}\right)$$

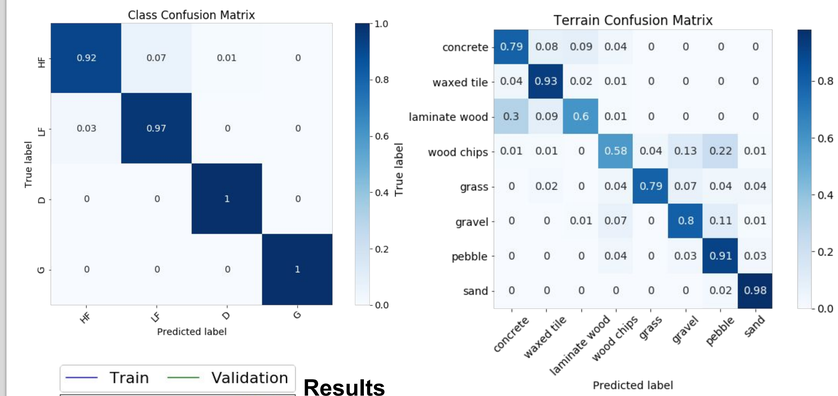
Appended Features:
Models including all 33 hand-designed features were investigated, as well as models including only a subset (i.e. control input), as many of the features can be inferred from the raw data.

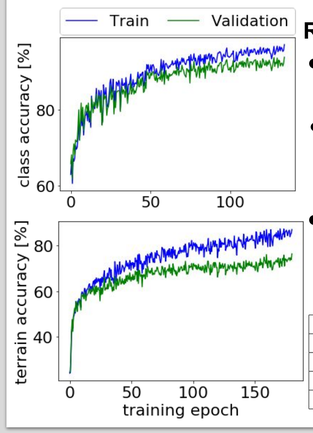
SVM Baseline

An SVM classifier with a PUK kernel (as used previously for terrain classification in [1] and [3]) served as a baseline for performance. Inputs were the 33 hand-designed features.



Results and Discussion





Results

- HF (μ=1) and LF (μ=0.5) are the most difficult to separate as their surface properties are similar.
- Further distinguishing terrain classes into individual terrains poses an additional challenge and results in lower classification accuracy.
- While including all 33 features results in the best accuracy, features that *cannot* be inferred from raw data are of particular benefit to include.

Model	Class Accuracy	Terrain Accuracy
SVM	84%	70%
CNN with raw data	87%	69%
CNN with raw data + all features	97%	80%
CNN with raw data + control input features	93%	76%

Conclusions and Future Work

Conclusions:

- Deep learning improves accuracy from SVM on the terrain classification task by over 10% both for higher-level classes and individual terrains.
- A 1D CNN utilizes temporal information and successfully extracts features relevant to terrain classification from raw sensor data.

Future Work:

- Architectures such as LSTMs, GRUs, or Transformers may be investigated for real-time, mid-step terrain classification.
- Current data is low resolution in the time dimension. Higher resolution data (especially on shear taxel) may improve classification between HF and LF.
- With better individual terrain classification, optimizing gait parameter by terrain rather than high-level terrain class may improve locomotion.
- Expanding this work to a terrain property classifier (e.g. coefficient of friction, stiffness) could lead to more generalizable gait adaptability.

References:

[1] Wu, X. Alice, et al. "Tactile Sensing and Terrain-Based Gait Control for Small Legged Robots." *IEEE Transactions on Robotics* (2019).
[2] Li, Dan, et al. "Classification of ECG signals based on 1D convolution neural network." *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 2017.
[3] Wu, X. Alice, et al. "Integrated ground reaction force sensing and terrain classification for small legged robots." *IEEE Robotics and Automation Letters* 1.2 (2016): 1125-1132.