EF-Net: End-to-End Friction Estimation Using Intrinsic Imaging and Deep Networks

Sarvesh Prajapati,[1] Abhinav Kumar[1] and Rupesh Pathak[1]

***Abstract*—Material recognition and friction estimation although overlooked is an important aspect of computer vision especially for robotic applications. Domains like grasping, motion planning, footstep planning, etc., are some of the few applications where it is important for the agent/robot to be aware of the material or friction coefficient associated with the material. This paper goes over why friction estimation is important for legged robots, we go over the current state-of-the-art material recognition and friction estimation methods existing datasets and address the problems by formulating an end-to-end friction estimation network. In the end, we present a novel architecture that relates material patches to their respective spectral values. All the methods replicated here surpass the results presented by the authors and our new model can learn physics-based reflectance of unseen classes as well.**

I. Introduction

With the Fourth Industrial Revolution, technology is advancing rapidly and the rise of robots in workspace and industry is increasing. To address topics like Human-Robot Interaction either for industry or household purposes, it becomes necessary for robots to be aware of their surrounding and know what materials they are interacting with. Fundamental challenges in robotics include grasping, path-planning, and day-to-day interaction with humans. For us humans, it comes naturally how much force to apply when holding a plastic cup as opposed to a ceramic cup, but for a robot, differentiating between a paper cup and a ceramic cup can be challenging, and excess force may deform the paper cup. For legged robots, it is important to know the friction values of terrain, such that the constraints of a convex optimization problem are respected and the system remains stable. Our work is motivated by quadruped locomotion problem, existing work in quadruped doesn't account for friction values and assume that the knowledge of world is given beforehand. Figure 1 shows the importance of friction values for a quadruped locomotion and [1] formulations of MPC makes it clear that friction estimation is an important aspect for quadruped locomotion.

II. Related Work

In our literature review, we explore methodologies centered on harnessing vision to predict the friction coefficient of materials, with a primary emphasis on vision-based approaches. Additionally, we investigate applications of tactile perception for the same purpose.

[1]Institute for Experiential Robotics, Northeastern University, Boston, Massachusetts, USA. {prajapati.s, kumar.abhina, pathal.r}@northeastern.edu
Code available at - https://github.com/prajapatisarvesh/friction-estimation

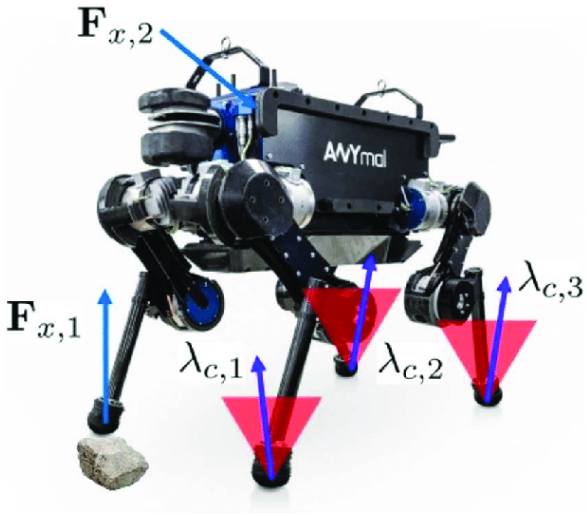

Fig. 1. Friction Cone constraints of a quadruped [2]

The relevant literature in this field primarily centers around material recognition tasks that employ image segmentation methods to estimate friction on various material surfaces. Our proposed approach transcends conventional methods by integrating the classification and recognition of materials into distinct classes as done in the paper . This involves the meticulous labeling of each pixel value with corresponding friction values. Our exploration encompasses diverse techniques discussed in various papers, including a method proposed by Kick et al. [1] for image segmentation and the utilization of a trained ResNet50 [3] model by Smith et al. [2], trained on Imagenet [4]. Furthermore, our study delves into the effectiveness of incorporating this network with transfer learning, specifically focusing on the reflectance cone of the materials.

Another noteworthy method we encountered in the paper by Hanson et al. [5] as utilizing image spectrum and IMU data for material recognition. In this modular approach, spectral data is generated when the material's surface is illuminated with IR spectrum light, and the reflected light is observed within the spectrometer. The network is then trained to correlate values from the spectral, image, and IMU data to different material classes.

In this paper, we extend existing work into a End-to-End friction estimation network using U-Net [6] and eventually propose EF-Net (End-to-End friction estimation network) that takes in an RGB image and outputs the spectral value associated with the image, which can be further used for estimating

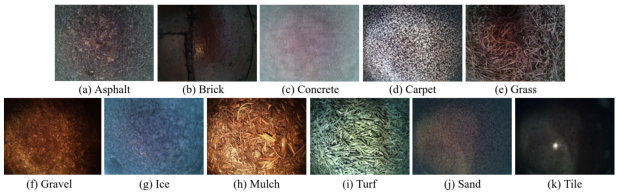

Fig. 2. Vast Dataset by Hanson et al. [5]

Fig. 3. Apple dms segmentation dataset [11]

the friction values.

III. Dataset

There aren't a lot of datasets available out there. Some of the widely used datasets are Material Recognition in the Wild [7], Flickr Material Database [8], etc. Another limitations of these datasets are either they are not labelled properly or are not suitable for real world implementation. Another problem in friction estimation is there's no dataset available that maps pixel values to friction coefficient, some papers [9] did create a dataset, but haven't oponsourced it yet. We used two datasets for our approach, Apple DMS Dataset [10], one of the challenge was downloading this dataset as the dataset only contains the labels and images needs to be downloaded seperately, and there's no proper information available for the same. We also used VAST dataset [5] for our proposed EF-Net that maps RGB values into corresponding spectral data.

A. Apple DMS Dataset

The Dense Material Segmentation Dataset (DMS) consists of 3 million polygon labels of material categories (metal, wood, glass, etc) for 44 thousand RGB images. It consist 3.2 million dense segments on 44,560 indoor and outdoor images. Figure 3 shows the dataset.

B. VAST Dataset

The extensive dataset for material recognition is modular, comprising three distinct types of information: images, spectral data, and inertial measurement unit (IMU) data. The author utilized material recognition using multi-modal approach, we just use the RGB images and estimate the spectral values. Figure 2 shows the dataset.

IV. Methodology

Our methodology unfolds in a systematic progression, commencing with the reproduction of established results in the realm of segmentation, as detailed in the corresponding subsection. This foundational step serves as a benchmark for our subsequent endeavors. Building upon this, our strategy evolves to address the challenge through a regression lens, delineated in the Regression subsection.

A pivotal contribution emerges in the form of EF-Net, a novel architecture meticulously designed to cater specifically to our unique problem domain. The intricacies of EF-Net are expounded upon in the dedicated subsection, shedding light on the innovative features and tailored design choices that distinguish it in the landscape of neural network architectures. This multi-stage approach underscores our commitment to a comprehensive exploration of the problem space, from replication to adaptation and ultimately innovation.

A. Segmentation

In our pursuit of a comprehensive solution, we embarked on a meticulous recreation of the U-Net architecture, leveraging the powerful Apple DMS Dataset. Our focus was on segmentation, resulting in a finely tuned model that produced segmented images across 57 distinct classes. These classes encompass a spectrum of materials such as wood, paper, asphalt, stone, and more. Each of the 57 segmented classes is methodically mapped to specific friction values. This mapping is informed by insights gleaned from a diverse set of sources, including the Material Recognition CNNs and hierarchical planning for biped robot locomotion on slippery terrain paper. By bridging the gap between segmentation and friction values through a carefully curated dataset and informed mapping, our approach advances beyond conventional segmentation tasks.

B. Regression

In the pursuit of refining our regression methodology, we undertook a strategic modification of several established architectures. These adaptations were specifically geared towards transforming these architectures into potent tools for regression tasks. The foundational step involved acquiring ground truth masks from the Apple DMS Dataset, and subsequently utilizing the earlier-established mapping to generate masks with friction values assigned to every pixel across the entire image.

This meticulously crafted dataset, with pixel-wise friction values, served as the invaluable ground truth for training our regression models. The architectures, fine-tuned for this purpose, underwent a rigorous training process to learn and predict friction values at the individual pixel level. Our innovative approach encompasses the seamless integration of dataset creation, mapping, and architecture refinement, yielding a regression methodology poised to yield precise and meaningful pixel-wise friction estimations for diverse real-world scenarios.

1) SRCNN: Building on insights gained from a previous mini project and inspired by findings in the literature, we strategically incorporated the SRCNN model into our regression framework. The decision was rooted in the recognition, as per the referenced paper, that SRCNN exhibits promise in handling regression tasks. Drawing from our prior experience with SRCNN in a different context, we harnessed its capabilities for the nuanced challenges posed by our formulated problem.

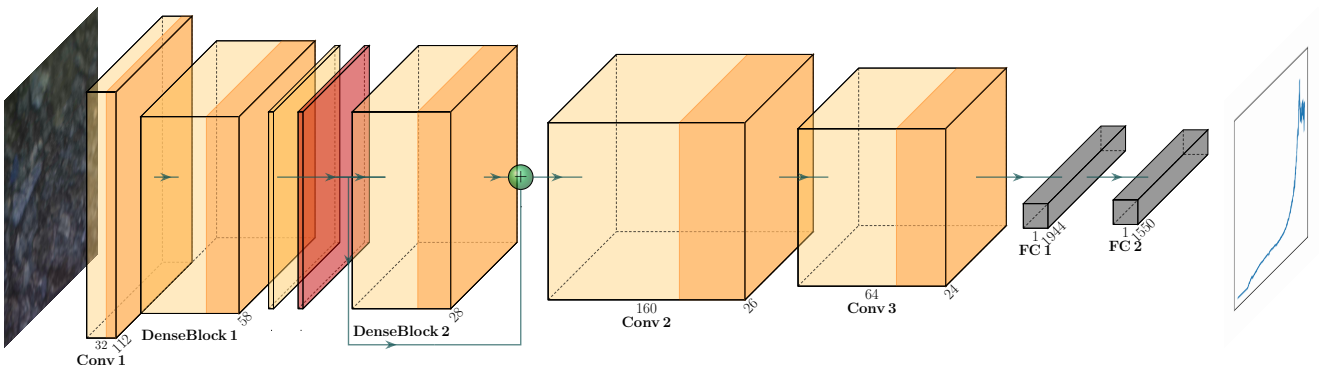

Fig. 4. Our proposed EF-Net

In this adaptation, SRCNN was employed to train on our dataset, meticulously crafted with ground truth masks containing pixel-wise friction values. The utilization of SRCNN represents a deliberate choice, informed by both its versatility and the prior success observed in our earlier mini-project. This integration not only aligns with established methodologies but also underscores our commitment to leveraging proven approaches to address the unique complexities inherent in our regression-based problem formulation.

2) U-Net: In a strategic evolution of our U-Net model, initially crafted for segmentation tasks, we implemented key modifications to tailor its functionality for regression purposes. The transformation involved a meticulous adjustment of various components:

Output Channel Configuration:

- **Previous Setting:** 60
- **New Setting:** 1

By transitioning from 60 output channels to a singular channel, our U-Net model was reconfigured to generate a single-channel output, aligning with the requirements of a regression task.

Loss Function Update:

- **Previous:** Cross Entropy Loss
- **Updated:** Mean Squared Error (MSE) Loss

Recognizing the distinct demands of a regression task, we replaced the Cross-Entropy Loss with the more fitting Mean Squared Error (MSE) Loss. This adjustment fine-tuned the training process to better align with the regression-based objective.

Training Data Ground Truth:

Utilized ground truth masks from the Apple Dataset. Ground truth masks from the Apple Dataset served as the foundational training data, facilitating the model to learn and predict pixel-wise friction values for the entire image.

This comprehensive reconfiguration underscores our commitment to methodical optimization, ensuring that our U-Net model is aptly equipped to excel in the nuanced realm of regression, where precision and fine-grained predictions are paramount.

C. EF-Net

In the development of our EF-Net architecture (Figure 4), a dedicated model crafted for spectral-to-friction mapping, we executed a meticulous design process leveraging the VAST dataset. Here is a detailed breakdown of the architecture:

Network Design:

Input Processing: The input image undergoes processing through DenseNet169, specifically passing through denseblock1 and denseblock2.

Integration Layer: The outputs from the max pool layer after denseblock1 and denseblock2 are concatenated, creating a fused representation.

CNN Layers:

- The fused representation is then fed through the first CNN layer, reducing input channels from 160 to 64.
- Subsequently, the output from the first CNN layer is further processed by a second CNN layer, with input channels as 64 and output channels as 9.

Fully Connected Layers:

- The output from the second CNN layer is flattened and passed through a fully connected layer with input channels as 1944 and output channels as 1550.
- This is followed by another fully connected layer, maintaining input and output channels at 1550.

This intricately designed architecture reflects our commitment to precision in spectral-to-friction mapping, utilizing a fusion of DenseNet169 and customized CNN layers to capture and process the nuanced variations present in the spectral data.

V. Experiments

A. Implementation

1) U-Net Segmentation: In the training of our U-Net model tailored for segmentation tasks on the Apple DMS dataset, a carefully chosen configuration was employed to optimize performance. The training settings are detailed below:

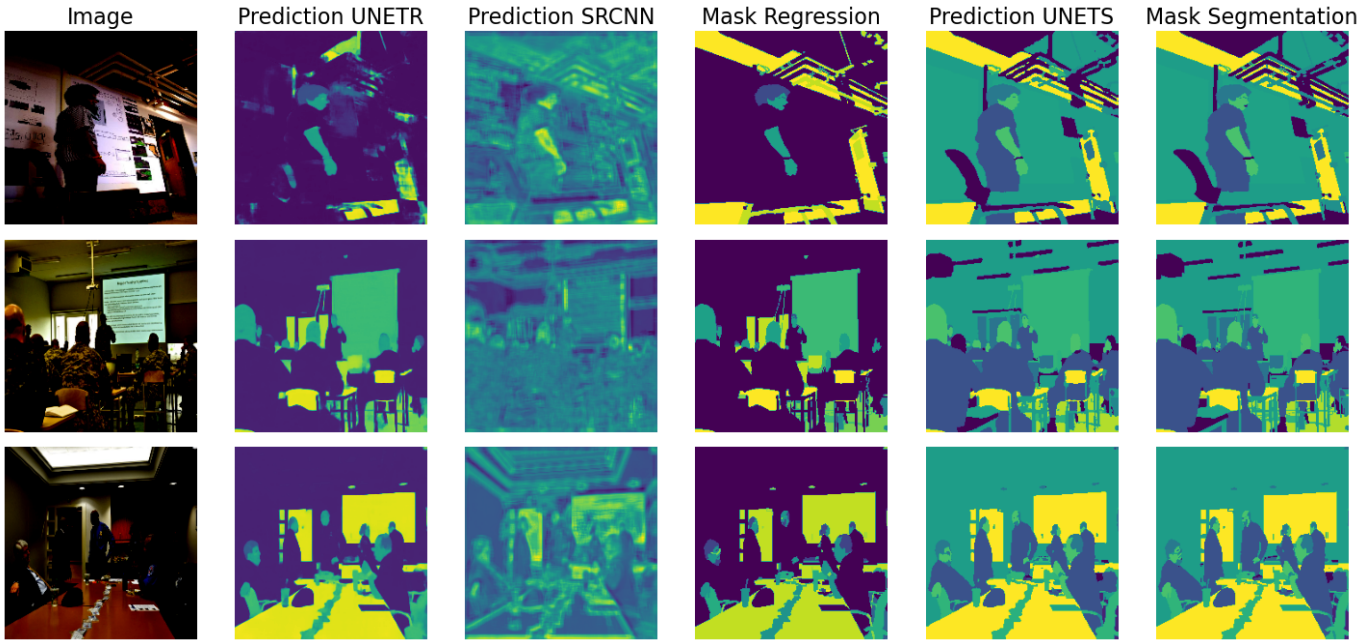

Fig. 5. Comparison between different results

- **Batch Size:** 32
- **Learning Rate:** 3e-4
- **Number of Epochs:** 300
- **Loss Function:** Cross Entropy Loss
- **Optimizer:** Adam Optimizer

2) SRCNN Regression: In the training of our SRCNN model, adapted for regression tasks on the tweaked Apple DMS dataset, a carefully chosen configuration was implemented to optimize model performance. The training settings are outlined below:

- **Batch Size:** 100
- **Learning Rate:** 3e-4
- **Number of Epochs:** 300
- **Loss Function:** Mean Squared Error (MSE) Loss
- **Optimizer:** Adam Optimizer

3) U-Net Regression: In the training of our U-Net model, specifically adjusted for regression tasks on the tweaked Apple DMS dataset, a meticulous configuration was applied to optimize model performance. The training settings are detailed below:

- **Batch Size:** 100
- **Learning Rate:** 3e-4
- **Number of Epochs:** 300
- **Loss Function:** Mean Squared Error (MSE) Loss
- **Optimizer:** Adam Optimizer

4) EF-Net: In the training of our proposed EF-Net, tailored for spectral-to-friction mapping using the VAST dataset, a carefully configured set of parameters was employed to optimize model performance. The training settings are as follows:

- **Number of Epochs (NUM EPOCHS):** 100
- **Batch Size:** 100
- **Learning Rate (lr):** 0.001
- **Loss Function:** Mean Squared Error (MSE) Loss
- **Optimizer:** Adam Optimizer

These parameters were thoughtfully selected to strike a balance between computational efficiency and the model's capacity to capture complex relationships within the output.

B. Evaluation

The assessment metric for the prediction model in segmentation relies on the validation data and is compared against the original mask. In the context of segmentation, the loss is calculated using the Dice score loss. Conversely, for other methods, we employ Mean Squared Error (MSE) loss for error evaluation.

MODEL	EVALUATION	SCORE
SRCNN	MSE	0.10±0.01
UNET Regression	MSE	0.03±0.01
UNET Segmentation	DICE	0.9402±0.01
EF-NET	MSE	0.002±0.001

TABLE I
RESULTS OF THE IMPLEMENTATION

VI. DISCUSSION AND RESULTS

Our experimentation and evaluation yielded noteworthy outcomes across various models, as depicted in the results (see Figure 5). A comprehensive overview of the findings is provided below:

Segmentation Model: Achieved commendable results in accurately delineating materials in the dataset.

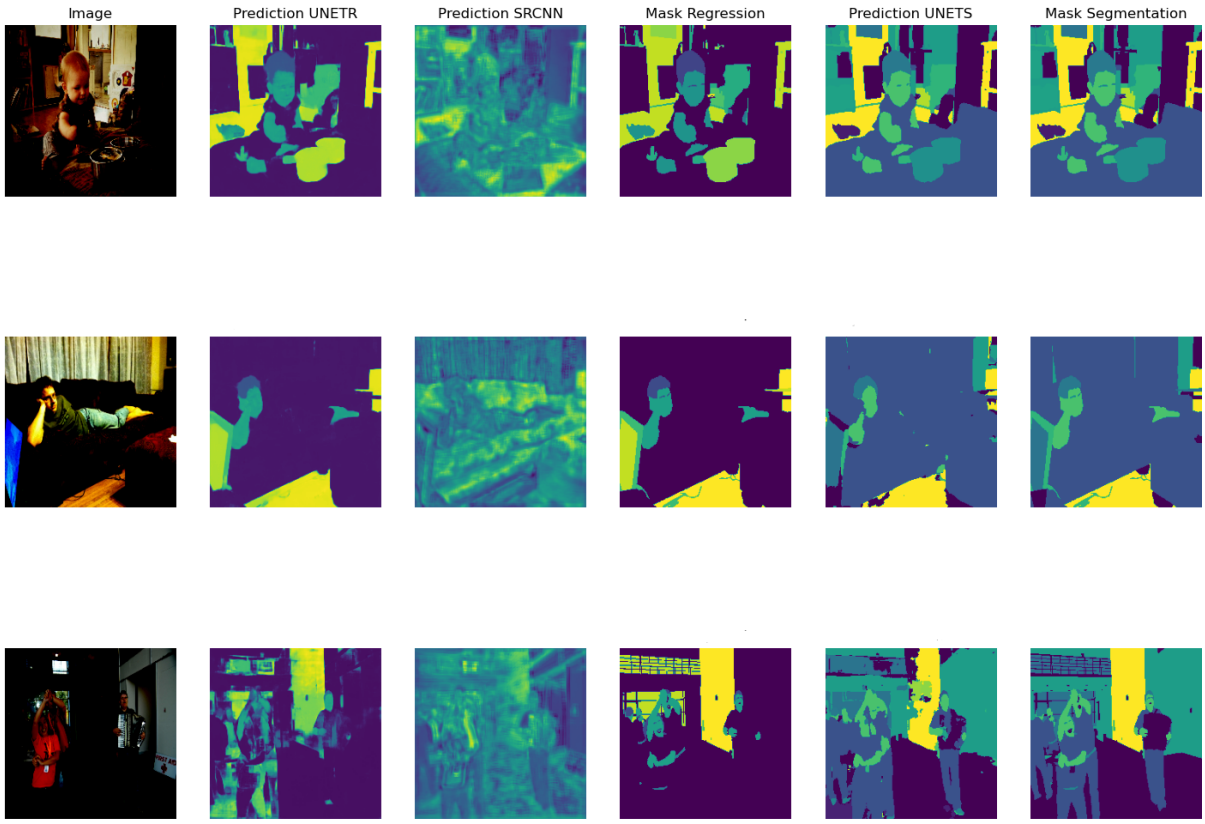

Fig. 6. Cases where Regression model performed better than segmentation data

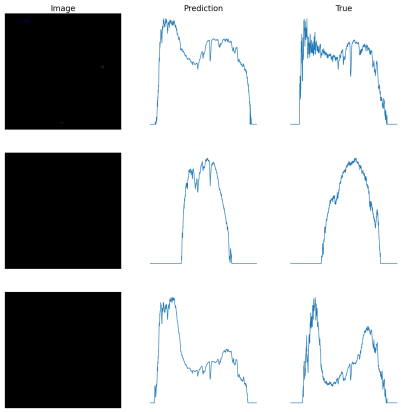

Fig. 7. Results from test set of seen data

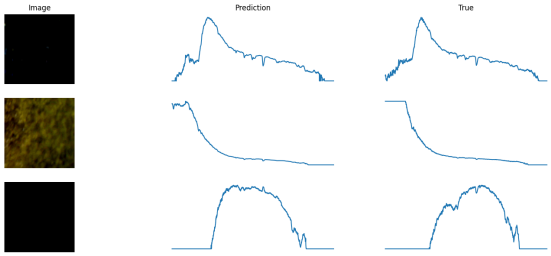

Fig. 8. Results from unseen data

U-Net Regression Model: Demonstrated superior performance in the regression task compared to the segmentation model (see Figure 6).

SRCNN Regression Model: Underperformed in the primary regression task but exhibited outstanding features for low illumination images, suggesting potential for further exploration in feature detection for such conditions (see Figure 5 and 6).

EF-Net: Displayed high accuracy for seen datasets (see Figure 7). Maintained good performance for unseen datasets, emphasizing robust generalization capabilities (see Figure 8).

These results collectively underscore the efficacy of the U-Net regression model and the EF-Net for the given problem. The SRCNN, despite its limitations in the primary task, revealed valuable characteristics that open avenues for potential applications in low illumination scenarios. This comprehensive analysis provides a foundation for future refinements and exploration, allowing for a deeper understanding of each model's strengths and areas for improvement.

VII. Conclusion

The challenge of friction estimation, often overlooked in its complexity, has been a focal point of our research. In addressing this gap, we underscore the significance of acquiring a robust dataset tailored for this specific purpose. Our devised neural network demonstrates remarkable capabilities by successfully mapping RGB images to their corresponding spectral values, showcasing its adaptability to both seen and unseen data.

Notably, the versatility of our proposed methodology extends beyond friction estimation. The same framework can be harnessed for diverse applications, including but not limited to material recognition. This broader utility highlights the potential impact of our work on advancing not only the field of friction analysis but also contributing to a broader spectrum of material characterization tasks. As we bridge the gap between spectral data and real-world friction coefficients, our findings pave the way for enhanced understanding and application in multiple domains.

VIII. Future Work

In our ongoing research endeavors, the upcoming focus revolves around the meticulous creation of a specialized dataset, specifically geared towards mapping spectral data to friction coefficients. The primary objective is to establish a robust and nuanced correlation between the intricate spectral features and their corresponding friction values. Concurrently, we plan to design a dedicated neural network architecture with the overarching goal of achieving precise end-to-friction mapping. This novel network will take spectral data as its input and produce friction coefficients as output, seamlessly integrating into our existing framework.

Furthermore, recognizing the necessity for continuous improvement, an essential component of our future work involves the fine-tuning of our End-to-Friction Network (EF-Net). This refinement process aims to optimize the network's parameters, ensuring enhanced performance and accuracy in the spectral-to-friction mapping task. The culmination of these efforts will result in a complete End-to-Friction Network capable of not only processing images but also providing detailed friction values for each pixel. This strategic integration of dataset creation, network design, and EF-Net fine-tuning aims to elevate the sophistication and efficacy of our overall network architecture.

IX. Acknowledgements

We would like to thank Prof. Bruce Maxwell for his continuous support throughout the duration of this course and his invaluable guidance for parts of this project work, without which we wouldn't have gotten desired results. We would like to thank Prof. Taskin Padir for providing us with invaluable guidance and his continuous support in the RIVeR Laboratory.

References

[1] G. Bledt, M. J. Powell, B. Katz, J. Di Carlo, P. M. Wensing, and S. Kim, "Mit cheetah 3: Design and control of a robust, dynamic quadruped robot," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 2245–2252.

[2] G. Xin, W. Wolfslag, H.-C. Lin, C. Tiseo, and M. Mistry, "An optimization-based locomotion controller for quadruped robots leveraging cartesian impedance control," *Frontiers in Robotics and AI*, vol. 7, 03 2020.

[3] P. Kicki and K. Walas, "Friction from reflectance: Transfer learning approach," in *2019 4th International Conference on Robotics and Automation Engineering (ICRAE)*, 2019, pp. 79–83.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[5] N. Hanson, M. Shaham, D. Erdoğmuş, and T. Padir, "Vast: Visual and spectral terrain classification in unstructured multi-class environments," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 3956–3963.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597

[7] S. Bell, P. Upchurch, N. Snavely, and K. Bala, "Material recognition in the wild with the materials in context database," *Computer Vision and Pattern Recognition (CVPR)*, 2015.

[8] L. Sharan, R. Rosenholtz, and E. H. Adelson, "Accuracy and speed of material categorization in real-world images," *Journal of Vision*, vol. 14, no. 10, 2014.

[9] D. Noh, H. Nam, M. S. Ahn, H. Chae, S. Lee, K. Gillespie, and D. Hong, "Surface material dataset for robotics applications (smdra): A dataset with friction coefficient and rgb-d for surface segmentation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 6275–6281.

[10] P. Upchurch* and R. Niu*, "A dense material segmentation dataset for indoor and outdoor scene parsing," in *ECCV*, 2022. [Online]. Available: https://arxiv.org/abs/2207.10614

[11] P. Upchurch and R. Niu, "A dense material segmentation dataset for indoor and outdoor scene parsing," in *European Conference on Computer Vision (ECCV)*, 2022.