Predictive Mapping of Spectral Signatures from RGB Imagery for Off-Road Terrain Analysis

Sarvesh Prajapati[1], Ananya Trivedi[1], Bruce Maxwell[2], Taşkın Padır[1]

Abstract—Accurate identification of complex terrain characteristics, such as soil composition and coefficient of friction, is essential for model-based planning and control of mobile robots in off-road environments. Spectral signatures leverage distinct patterns of light absorption and reflection to identify various materials, enabling precise characterization of their inherent properties. Recent research in robotics has explored the adoption of spectroscopy to enhance perception and interaction with environments. However, the significant cost and elaborate setup required for mounting these sensors present formidable barriers to widespread adoption. In this study, we introduce RS-Net (RGB to Spectral Network), a deep neural network architecture designed to map RGB images to corresponding spectral signatures. We illustrate how RS-Net can be synergistically combined with Co-Learning techniques for terrain property estimation. Initial results demonstrate the effectiveness of this approach in characterizing spectral signatures across an extensive off-road real-world dataset. These findings highlight the feasibility of terrain property estimation using only RGB cameras.

Fig. 1: Proposed RS-Net takes in RGB image and accurately predicts spectral value of a material patch.

I. INTRODUCTION

Understanding terrain properties is vital in off-road robotics, as they directly shape the robot's performance and abilities [1]. These properties encompass factors such as soil composition, friction coefficients, obstacle density, among other variables, all significantly influencing navigation, stability, and obstacle interaction. By comprehensively understanding and adapting to these terrain properties, robots can ensure safe and efficient traversal through diverse and challenging outdoor environments [2], [3].

Recent trends in off-road autonomy involve estimating terrain properties and incorporating this information into a motion planning algorithm [1], [4]–[6]. In a basic approach, terrain classification is first conducted using an off-the-shelf classifier, followed by using a lookup table to estimate the properties associated with each class [7]. However, this method loses generality when it comes to distinguishing between similar-looking materials or dealing with previously unseen objects. Another trend involves the use of specialized sensors such as haptics [8], lidars [9] and thermal sensors [10], where physical interaction with materials is utilized to estimate their properties. While these approaches provide valuable insights, they require dedicated sensors and involve physically touching objects to estimate their properties.
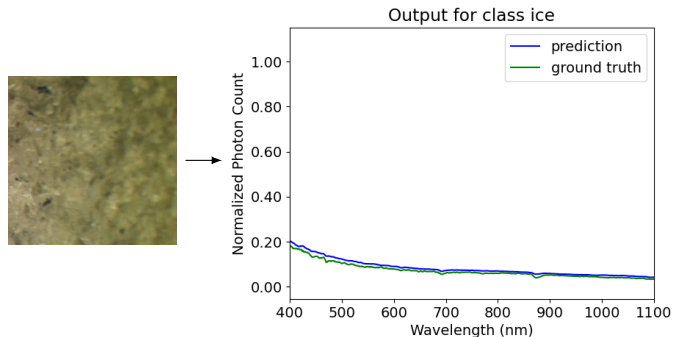
Spectroscopy is a scientific technique for examining how light interacts with materials [11]. It measures the spectrum of light that materials emit, absorb, or scatter. This approach enables learning about the material's composition, structure, and chemical properties. Each type of material has a unique pattern of wavelengths, known as a spectral signature, which is revealed when light interacts with it. By studying these signatures, spectroscopy provides a powerful tool for understanding the properties of different substances [12], [13]. Additionally, spectroscopy enables the observation of temporal changes in object properties, with alterations in electromagnetic radiation absorption or emission serving as indicators of ongoing compositional or physical transformations.

In our work, we propose utilizing spectroscopy to detect terrain properties based on their unique spectral signatures. These sensors, specifically designed for estimating physical properties, offer high accuracy. We demonstrate how a supervised deep neural network architecture can accurately predict spectral signatures from RGB data. Additionally, we illustrate how these spectral signatures can be leveraged to estimate various terrain properties, such as surface friction. This will eliminate the requirement for custom sensor assemblies currently needed in robotics spectroscopy applications [3], [14]. Our approach aims to establish a framework where offline-collected spectral data and a trained network can be used to estimate any terrain physical property, offering a more generalized solution compared to the custom solutions commonly employed in current terrain property estimation algorithms.

[1]Institute for Experiential Robotics, Northeastern University, Boston, Massachusetts, USA. {prajapati.s, trivedi.ana, t.padir}@northeastern.edu
[2]Khoury College of Computer Sciences, Northeastern University, Boston, Massachusetts, USA. b.maxwell@northeastern.edu
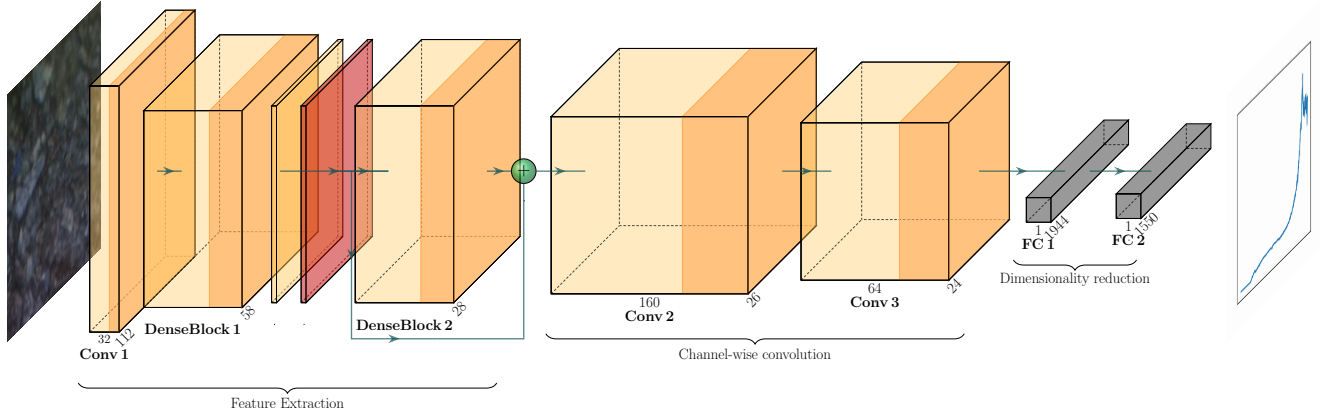
Fig. 2: RS-Net: End-to-End model for mapping RGB images to spectral profile. The input RGB image is passed through DenseNet's first and second dense block, output from first transitional layer is concatenated with output of second dense block, the feature maps are concatenated, passed through convolution and fully-connected layers for spectral mapping.

II. NETWORK ARCHITECTURES

In this section, we introduce RS-Net, a deep neural network architecture to convert RGB images into a spectral profile. Furthermore, we put forward a potential methodology for expanding this architecture to facilitate the estimation of terrain properties.

A. RS-Net Architecture

Figure 2 illustrates the RS-Net architecture. Figure 4 shows a typical spectral profile that indicates how the intensity of light varies across the spectrum, providing information about the wavelengths of light that are present and their relative intensities. RS-Net takes an input RGB image x and predicts a spectral profile x'_s. In the following sections, we break down the various components integral to our neural network design.

DenseNet: To extract high-level global features, the input RGB image undergoes multi-stage processing. We utilize a pretrained DenseNet169 [15] for this purpose, with the flexibility to substitute other DenseNet models, as the initial two blocks remain consistent across variants. In DenseNet, the image is initially processed through a convolution and pooling layer. We represent this as a composition of operations $G(\cdot)$.

$$x_1 = G(x) \tag{1}$$

The input image is then processed through dense block 1 and 2 of the DenseNet. This sequential passage enables the network to learn fine-grained details in dense block 1 and high-level abstract representations in dense block 2. Features extracted from these layers contribute to capturing the texture of the surface, thereby enhancing our ability to achieve better spectral correspondence and differentiating between materials of similar appearance. The first dense block has 6 layers of convolution while the second has 12 layers of convolutions. After first dense block, we get x_7, which is passed through transition layer that performs operation $G(\cdot)$. Finally this output from transitional layer is passed through second

dense block, outputting x_{21}. As per DenseNet, the general representation of a dense block for l convolutions is:

$$x_l = H_l([x_0, x_1, ..., x_{l-1}]) \tag{2}$$

Which for our network cam be written as:

$$x_7 = H_7([x_1, x_2, ..., x_6]) \tag{3}$$
$$x_8 = G(x_7) \tag{4}$$
$$x_{21} = H_{21}([x_8, x_9, ..., x_{20}]) \tag{5}$$

Fused Representation: Features from max-pooling of first transition layer and second dense block are concatenated for creating a fused representation of the global features. This operation outputs x_f which is the concatenation of x_8 and x_{21} and represented as:

$$x_f = x_8 \oplus x_{21} \tag{6}$$

CNN Layers: The fused representation is then fed through the first CNN layer, reducing input channels from 160 to 64, which in-turn helps in reducing the spatial dimension.

Subsequently, the output from the first CNN layer is processed by a second CNN layer, with 64 input channels and 9 output channels. This step is done for regularization, and reducing the dimensionality such that when the network is passed through fully-connected layers, it is more memory efficient and learns the spectral representation efficiently.

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. The above can be represented as a series composite function, $S(\cdot)$, and written as:

$$x'_s = S(x_f) \tag{7}$$

This is how our network maps RGB images x to their corresponding spectral profile x'_s. We evaluate the performance

of this network on a real-world off-road terrain dataset in section III.

B. Spectral Profiles to Terrain Property Estimation

We propose an extension of the network to predict various physical properties associated with the terrain based on the given spectral profile x'_s. Traditional linking of physical properties with RGB images faces challenges due to the under-constrained nature of the problem [16], [17]. Research, such as that by Hanson et al. [18], highlights the benefits of utilizing different modalities for estimating diverse physical properties crucial for off-road terrain analysis. Instead of treating the task as a straightforward pattern-matching problem, we employ the concept of Collaborative-Learning (Co-Learning) [19]. This approach enables learning representations from multiple modalities, whether they have strong or weak relations, facilitating knowledge transfer between them. By efficiently using RGB to predict spectral data and leveraging the learned parameters to infer a third modality, such as friction, our Co-Learning architecture facilitates the estimation of various physical properties of different off-road materials. Typically, Co-Learning architectures incorporate a split between modalities to predict different representations, facilitating seamless knowledge transfer and enhancing the understanding of underlying physical properties. Figure 3 shows a sample Co-learning architecture for terrain property analysis. The image x is passed through DenseNet, S_F and S_G are composite functions, they can be convolution, max-pooling, batch normalization and even fully-connected layers.
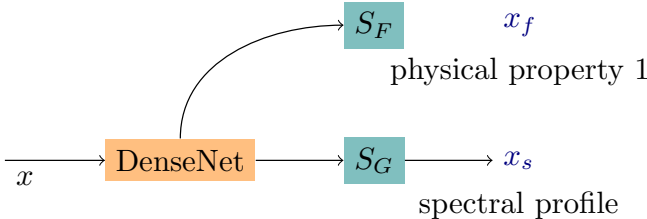

Fig. 3: Example of Co-Learning for simultaneous spectral profile and terrain property estimation

III. EXPERIMENTS

A. Evaluating RS-Net on trained labels and untrained labels

For evaluating our method, we utilize the open sourced dataset from [18] which includes high-resolution macro images of surface texture, spectral reflectance curves, RGB images, and localization data from a 9-DOF IMU, obtained across 11 varied terrains and different lighting conditions. We trained RS-Net on 6 out of the 11 classes in the VAST dataset. These classes include asphalt, brick, grass, ice, sand, and tile. However, the model was not trained on carpet, concrete, gravel, mulch and turf. The model was trained for 50 epochs, on an Adam optimizer [20] with a learning rate of $1e^{-3}$. Upon evaluation on the testing data, it achieved an MSE loss of 0.0015 Normalized Photon Count. Figure 4 displays the output
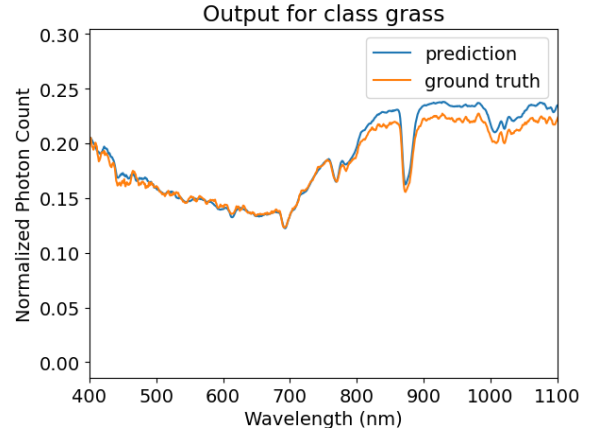

Fig. 4: Performance of RS-Net on seen class

of RS-Net on grass, which constitutes one of the classes the network was trained for. The model demonstrates a remarkable ability to predict the spectral profile with high fidelity.

Furthermore, we assess RS-Net's performance using data featuring unfamiliar labels to evaluate its capacity for generalization to novel materials. Illustrated in Figure 5, the spectral profile of turf, resembling that of grass, produces an output similar to grass, while gravel, akin to sand, and concrete, somewhat akin to asphalt, yield corresponding outputs. This suggests the network's potential to classify diverse objects based on their materials or physical properties, even with limited training. Additional outputs from the network, for both familiar and unfamiliar labels, are elaborated in the Appendix section.

B. Testing material classification

We conduct basic material classification tests using the output of RS-Net. Rather than employing Co-Learning or the concept of perceptual loss, we simply pass the spectral data through a 4-layer fully connected network and trained it for ten epochs. The confusion matrix on the test data is illustrated in Figure 6. The overall F-Score for the testing data was 0.79 with class accuracy of $0.79, 0.99, 0.99, 0.62, 0.57, 0.75$ for asphalt, grass, ice, sand, brick and tile respectively.

Overall, the network demonstrates an ability to discern patterns even in the case of unseen classes, indicating promising potential for generalization.

IV. DISCUSSION AND FUTURE WORK

In robotics, a key challenge lies in generalizing models for dynamic environments, where existing datasets often fail to capture the variability. For instance, spectral profiles and surface friction can vary based on factors like weather conditions, such as "dry asphalt," "ice on asphalt," and "water on asphalt." Figure 7 shows how grass's spectral profile shifts in the y-axis with changes in environmental conditions, reflecting corresponding changes in predictions and the network's adaptability to RGB data variations. To enhance model generalization, we
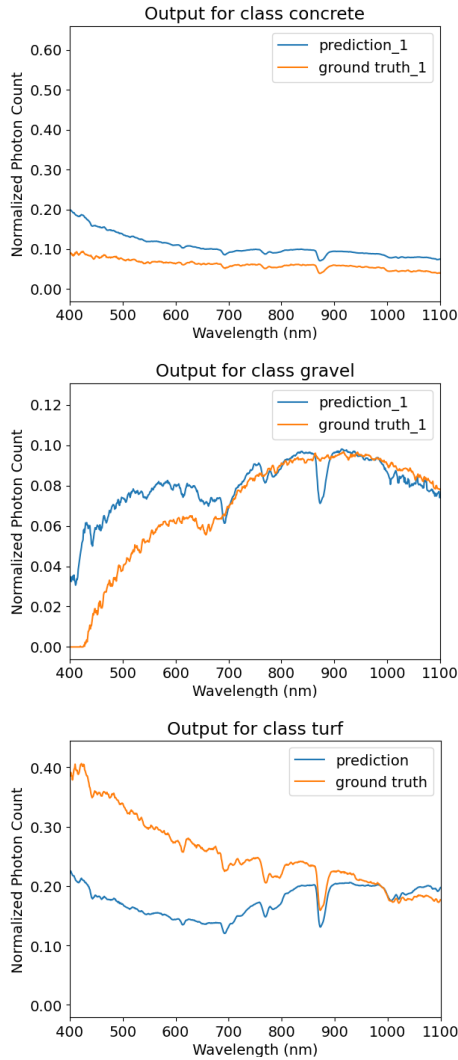
Fig. 5: Predicted spectral profile v/s ground truth data for classes not in training set
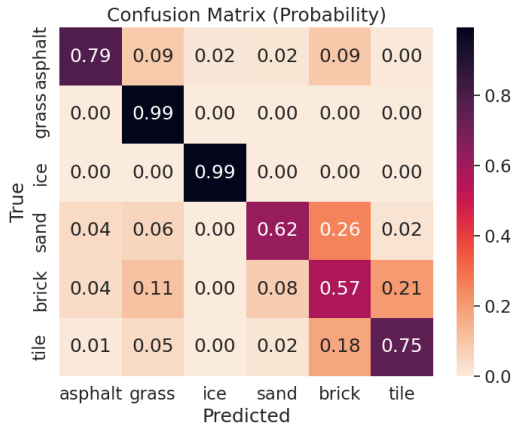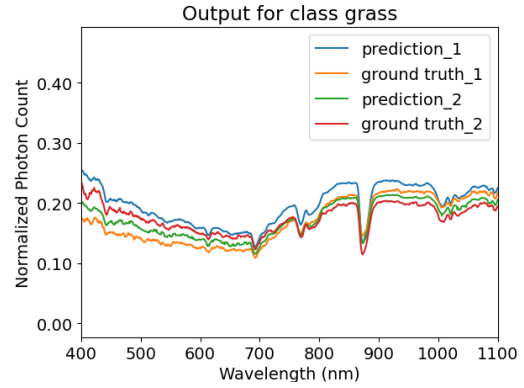

Fig. 7: Spectral profile of grass in different environment conditions

intend to enrich the VAST dataset with such nuanced data points.

Our future work involves exploring the relationship between spectral profile shifts and material properties, and developing a Co-Learning network capable of predicting various physical properties like friction and traversability. We also aim to investigate how material deposition affects spectral profiles and, consequently, physical properties. This research direction promises to deepen our understanding of material interactions in dynamic environments and enhance robotic system capabilities.

The existing RS-Net faces deployment challenges as it relies on forwarding material patches to the network, which limits its effectiveness. This dependency on image segmentation or passing patches beneath the robot affects the network's accuracy in long-range forecasting. To overcome this limitation, we suggest incorporating hyperspectral data [21] by either transfer learning of spectral data through learning or adapting the network to directly learn features from hyperspectral data.

V. CONCLUSION

This study underscores the importance of integrating different modalities in material recognition and friction estimation tasks to advance robotic applications. The introduction of RS-Net, a novel architecture mapping RGB images to spectral profiles, presents a promising solution to this challenge. Through experimentation, we have demonstrated the network's ability to generalize to unseen classes, indicating its potential to improve terrain property estimation tasks. Furthermore, we discuss the limitations of current datasets and propose how an augmented data set can enhance the network's ability to learn detailed physical information about a material from just an RGB image.

Fig. 6: Confusion matrix for terrain classification

REFERENCES

[1] P. Ewen, A. Li, Y. Chen, S. Hong, and R. Vasudevan, "These maps are made for walking: Real-time terrain property estimation for mobile robots," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7083–7090, 2022.

[2] N. Hanson, B. Pyatski, S. Hibbard, C. DiMarzio, and T. Padır, "Hyperdrive: Visible-short wave infrared hyperspectral imaging datasets for robots in unstructured environments," *arXiv preprint arXiv:2308.08058*, 2023.

[3] N. J. Hanson, "Material informed robotics–spectral perception for object identification and parameter inference," Ph.D. dissertation, Northeastern University, 2023.

[4] G. Williams, A. Aldrich, and E. A. Theodorou, "Model predictive path integral control: From theory to parallel computation," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 2, pp. 344–357, 2017.

[5] A. Trivedi, M. Zolotas, A. Abbas, S. Prajapati, S. Bazzi, and T. Padır, "A probabilistic motion model for skid-steer wheeled mobile robot navigation on off-road terrains," *arXiv preprint arXiv:2402.18065*, 2024.

[6] ——, "A probabilistic motion model for skid-steer wheeled mobile robot navigation on off-road terrains," *arXiv preprint arXiv:2402.18065*, 2024.

[7] M. Brandão, Y. M. Shiguematsu, K. Hashimoto, and A. Takanishi, "Material recognition cnns and hierarchical planning for biped robot locomotion on slippery terrain," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, 2016, pp. 81–88.

[8] T. N. Le, F. Verdoja, F. J. Abu-Dakka, and V. Kyrki, "Probabilistic surface friction estimation based on visual and haptic measurements," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2838–2845, 2021.

[9] X. Cai, S. Ancha, L. Sharma, P. R. Osteen, B. Bucher, S. Phillips, J. Wang, M. Everett, N. Roy, and J. P. How, "Evora: Deep evidential traversability learning for risk-aware off-road autonomy," *arXiv preprint arXiv:2311.06234*, 2023.

[10] T. Bhattacharjee, H. M. Clever, J. Wade, and C. C. Kemp, "Material recognition via heat transfer given ambiguous initial conditions," *CoRR*, vol. abs/2012.02176, 2020. [Online]. Available: https://arxiv.org/abs/2012.02176

[11] C. Pasquini, "Near infrared spectroscopy: A mature analytical technique with new perspectives–a review," *Analytica chimica acta*, vol. 1026, pp. 8–36, 2018.

[12] Z. Yang, T. Albrow-Owen, W. Cai, and T. Hasan, "Miniaturization of optical spectrometers," *Science*, vol. 371, no. 6528, p. eabe0722, 2021.

[13] K. D. Hakkel, M. Petruzzella, F. Ou, A. van Klinken, F. Pagliano, T. Liu, R. P. van Veldhoven, and A. Fiore, "Integrated near-infrared spectral sensing," *Nature communications*, vol. 13, no. 1, p. 103, 2022.

[14] N. Hanson, W. Lewis, K. Puthuveetil, D. Furline, A. Padmanabha, T. Padir, and Z. Erickson, "Slurp! spectroscopy of liquids using robot pre-touch sensing," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 3786–3792.

[15] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," *CoRR*, vol. abs/1608.06993, 2016. [Online]. Available: http://arxiv.org/abs/1608.06993

[16] H. Barrow and J. Tenenbaum, "Recovering intrinsic scene characteristics from images," *Recovering Intrinsic Scene Characteristics from Images*, 01 1978.

[17] Y. Weiss, "Deriving intrinsic images from image sequences," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2, 2001, pp. 68–75 vol.2.

[18] 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.

[19] G. Song and W. Chai, "Collaborative learning for deep neural networks," 2018.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[21] N. Hanson, B. Pyatski, S. Hibbard, C. DiMarzio, and T. Padır, "Hyperdrive: Visible-short wave infrared hyperspectral imaging datasets for robots in unstructured environments," in *2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2023, pp. 1–5.

APPENDIX

A. Results

Figure 8 and 9 depict the predicted versus ground truth spectral profiles generated by RS-Net. It is evident that certain materials may exhibit similar profiles at specific wavelengths, suggesting the potential use of this approach to categorize on unknown materials into known categories. A shift in the y-axis can be seen in Figure 8 compared to Figure 4 suggesting that the data was collected under different environment condition. Figure 9 demonstrate RS-Net's ability to generalize on out-of-class data. Notably, as the features of asphalt and concrete somewhat overlap, the model produces somewhat noisy and/or shifted profiles for concrete as observed in Figure 9.
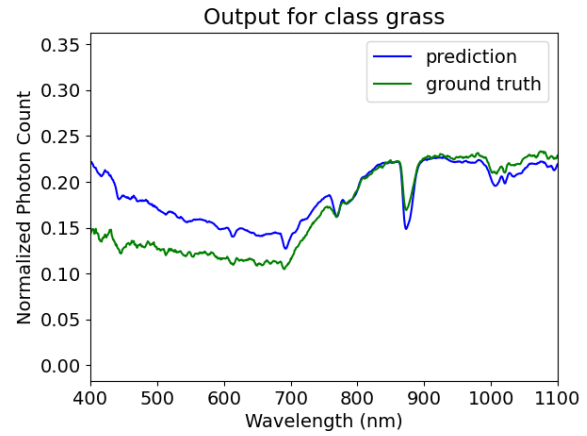

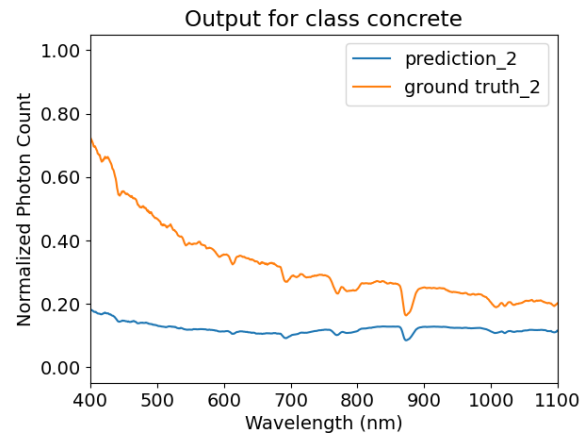Fig. 8: Predicted spectral profile v/s ground truth data for grass

Fig. 9: Predicted spectral profile v/s ground truth data for concrete