CNN based Plastic Detection using Hyper-spectral Images by Spectral Band Reflectance Computation

Deep Learning Theory Course Project

Link to Github Repository

Chirag Gupta(21BAI1079)
Srinitish Srinivasan(21BAI1394)
Yadu Krishnan(21BAI1210)

Dr. Harini S
25 April 2024

Abstract

Plastic waste emerged from build-up of plastics negatively effects the environment. As a result, many research studies have proposed plastic detection in order to tackle this problem. Hyper-spectral imaging has gained popularity over the years with the advancement of technology required to leverage their potential. Several studies show that hyper-spectral imaging is effective in detecting plastic waste and microplastics as they are able to capture the plastic reflectance spectral by using a near-infrared sensor. It has aided researchers in the detection of cancer, agricultural and military equipment and can be seen as a potential solution to the environmental problems caused by plastic. However so far, there have only been a handful of studies conducted on the detection of plastic using HSI technologies and machine learning. Both the emergence of HSI technologies and their implementation are still in their infancy at this point. Therefore, we propose a state of the art method involving data collection and image classification that leverages the potential of hyper-spectral imaging and deep learning to effectively detect plastics on the surface. This paves the way for automated picking of garbage, especially small and thin plastics that require high precision.

Introduction

Climate change presents a significant threat to our planet, affecting ecosystems, weather patterns, and human well-being. The emission of greenhouse gases, primarily from human activities such as burning fossil fuels and deforestation, contributes to global warming. This warming leads to rising sea levels, more frequent and severe weather events, shifts in precipitation patterns, and disruptions to ecosystems. Moreover, climate change exacerbates existing environmental challenges, including the proliferation of plastic pollution.(Shenming et al.)

Plastic pollution poses a substantial environmental problem, with detrimental effects on marine life, ecosystems, and human health. Single-use plastics, such as bottles, bags, and packaging, persist in the environment for extended periods, breaking down into microplastics that contaminate waterways and ecosystems worldwide. These microplastics can accumulate in the food chain, posing risks to marine organisms and potentially entering human diets. Additionally, the production and incineration of plastics contribute to greenhouse gas emissions, further intensifying climate change. Addressing plastic pollution requires concerted efforts to reduce plastic consumption, improve waste management systems, and promote sustainable alternatives.(Shenming et al.)

Hyper-spectral images (HSIs) is a three-dimensional data cube with special high dimensionality, strong correlation between adjacent bands and highly nonlinear data structure. Hyper-spectral images include images captured in different bands of the spectrum. Hyper-spectral images include more channels than regular RGB images which gives us more data to work with. It has better robustness and discriminant, and can improve the classification accuracy of the model. HSI's limitation is that the number of samples can be limited and deep learning requires large number of data. (Vasarhelyi)

Detecting plastics using hyper-spectral cameras can be challenging. Plastic can be of many types, transparent, coloured, thin or even thick. Detecting plastic using hyper-spectral cameras can be advantageous over detecting it using regular cameras, as we have much more details than regular image. Current techniques like detecting plastic with YOLOv5 considers only RGB bands. We predict their image of reflectance over 5 bands and detect plastic through their reflectance thresholds. These values can then be used to train the model to predict similar plastic using hyper-spectral images. Different types of plastic reflect light at different speeds and wavelengths. This reflectance value can also be calculated to segregate plastic into different categories.

Related Work

Reference	Description
(Sivaram et.al)	When designing a CNN model for plastic type classification encompassing seven categories (HDPE, PET, LDPE, PVC, PP, PS, Other), the process involves collecting and renaming appropriate images, resizing them to balance computational load and identification accuracy (64x64 or 120x120 pixels), and determining the CNN layers. Various structures such as LeNet-5 and AlexNet are explored, leading to a simplified model. This model incorporates convolutional, max-pooling, activation, dropout, and SoftMax layers for feature extraction and classification. System analysis aims to comprehend the problem, identify key variables, and devise optimal solutions.
(Wen et al.)	The system utilizes an Intel RealSense D435 depth camera for information acquisition, coordinating with a robotic arm controlled by a computer for sorting actions. Equipped with an AUBO-i5 robotic arm and integrated vacuum sponge sucker, objects are picked up efficiently. Communication between modules occurs via USB and Ethernet. The conveyor module features a belt with an aluminum bracket for camera adjustment. For plastic classification, a dataset of 6247 images is collected, divided into categories such as wash supplies bottle, beverage bottle, Tetra Pak, express package, and tableware box. Object detection employs YOLOX, leveraging CSPDarknet as backbone and SimOTA for efficient assignment. YOLOX-m is chosen for its optimized speed and accuracy. Data augmentation techniques including mosaic and mixup strategies enhance training sample diversity, improving detection performance.

(Mhadlekar et al.)	In addressing challenges of data availability, 720 photos of local plastics were gathered from diverse sources, categorized into four classes for subsequent analysis and classification. Following collection, advanced processing steps were employed to clean and prepare the data, including various pre-processing techniques to mitigate noise and handle the unstructured nature of real-world data. Model training and evaluation ensued, where pre-trained network convolution layers were utilized for feature extraction from new samples, followed by the addition of fully connected layers and subsequent training on the dataset. Further refinement was achieved through fine-tuning, where custom layers were added for classification. The prepared networks underwent training and evaluation using varying proportions of training and testing data, employing the VGG-16 model, renowned for its deep CNN architecture and effectiveness in image recognition tasks.
(Bhanumathi et al.)	This paper explores the application of YOLOv4 and YOLOv5 algorithms for identifying ocean plastics in epipelagic layers. YOLO algorithms are prized for their speed, high accuracy, and learning capabilities. YOLOv4, an extension of YOLOv3, introduces significant enhancements, including BoF (bag of freebies) and BoS (bag of specials) techniques to boost accuracy without sacrificing inference time. Utilizing the CSPDarknet-53 backbone structure, YOLOv4 accommodates input images of any size and leverages GPU and CUDA library for enhanced computational power during training. YOLOv5, based on PyTorch, offers a smaller, faster, and lightweight alternative, although studies indicate trade-offs between accuracy and speed compared to YOLOv4 and YOLOv3. While YOLOv5 excels in real-time object detection, YOLOv4 may be preferable for custom configurations.

(Salim et al.)	The system comprises a Logitech C615 webcam and a 22-inch display monitor connected to a Jetson Nano Development Kit. The choice of Jetson Nano is due to its comprehensive features and affordability, making it suitable for deploying deep learning models widely. Mounted on an aluminum profile behind a trash can, the system detects individuals approaching with reusable bottles, displaying a bounding box around the bottle on the monitor along with persuasive text tailored to the context. The setup encourages plastic reuse and can display information in multiple languages.
(Roslan et al.)	The paper employs the YOLOv5 model for real-time detection and recognition of plastic surface defects, utilizing its backbone, neck, and head components to optimize gradient flow and information hierarchy. Specifically, the CSP-Darknet backbone and PANet neck enhance feature extraction and information propagation, while the Yolo Layer outputs relevant vectors. YOLOv5s is chosen for its compatibility with time and hardware constraints. Performance evaluation involves precision, recall rate, and mean average precision (mAP) metrics, utilizing a custom dataset augmented for robustness. The YOLOv5 model is constructed using Python and PyTorch library, with subsequent evaluation based on precision, accuracy, F-measures, and recall post-training.

(Tamim et al.)	The camera setup comprises an iPhone 12 and a Mapir Survey 3W camera positioned on separate heavy-duty tripods to capture RGB and RGNIR images of plastic waste from various distances and angles. Three public locations in Kota Kinabalu, Sabah, Malaysia are chosen for data collection, resulting in 405 images pre-processed to remove non-plastic backgrounds and resized to 416×416. Annotations are added using LabelImg, and the dataset is split into training, validation, and testing sets. The YOLOv5 object detection model is utilized, featuring CSPDarknet53 as the backbone, PANet as the neck, and three YOLO heads, with YOLOv5m selected for its balance of complexity and architecture. K-fold cross-validation with a value of 10 ensures unbiased performance evaluation, while mean average precision (mAP) metrics assess the models effectives in detecting plastics.
(Zhou et al.)	In preprocessing, PRISMA L2C images undergo radiometric correction but not ortho-rectification. Ortho-rectification is achieved using ground control points (GCPs) from Sentinel 2 images, ensuring an RMSE of less than 6 meters. GCPs are then applied to polygons generated in chapter 2.2, up-sampling the shapefile from 30 to 10 meters. The image resolution is unified, revealing a wide range of plastic product sizes. To meet the fixed input size requirement for deep learning, a strategy is employed: a moving window of 5x5 pixels determines the smallest polygon unit, with larger polygons cropped into smaller patches. Centroids of these subsets are used to generate bounding boxes of 64x64 pixels, forming the final input image size. The dataset comprises 12,468 annotated plastic images across five functional classes. Training and validation datasets are split at an 8:2 ratio. Five state-of-the-art deep learning networks—ResNet-18, ResNet-50, VGG-11, VGG-19, and DenseNet—are evaluated on the dataset using PyTorch 1.12.1 with Python 3.8, trained on three NVIDIA Quadro P4000 GPUs.

(Tamil, Owen et al.)	Data augmentation techniques involve geometric transformations to enhance dataset size and diversity. These transformations, such as flipping, rotation, and random cropping, maintain pixel values to preserve image features. For instance, flipping creates mirror images by rotating objects 180 degrees, while color space transformations modify color distributions for varied lighting. Random cropping enlarges image sections proportionally, boosting data efficiently. Noise injection, like salt and pepper noise, prevents overfitting by dotting images with white and black dots. Manual acquisition of plastic waste datasets with diverse viewpoints is crucial for training deep learning models effectively. Enhancing the quality of plastic waste datasets involves capturing images in real-scene backgrounds like rivers and landfills, challenging current object detectors. Annotated datasets with complete or scattered plastic waste objects in real scenes improve model training and assessment. Utilizing near-infrared (NIR) spectral information, ranging from 780 to 2500 nm, offers valuable insights beyond visible light. NIR reflectivity varies among plastic types, with spectra around 1084 nm to 1212 nm showing notable features. Integrating NIR images into existing datasets enhances deep learning model training and detection performance.

Methodology

Data Acquisition

We used a DJI Phantom P4 Multispectral camera drone for this purpose. The camera can capture 5 different images, namely, Red, Blue, Green, NIR and Infrared. We first set plastic on the ground in random positions. The

Figure 1.1

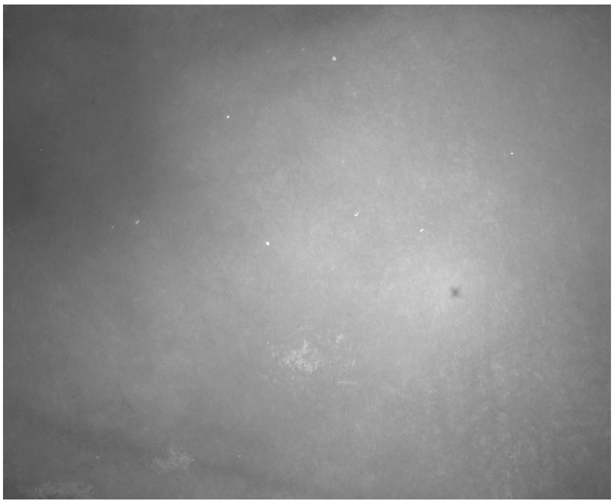

Figure 1.2

position and type of plastic placed was completely random. We then flew the drone right above it. The drone was set to a height of 70 feet from the ground. It was set to capture images with an interval of every 3 seconds. In this interval period, the drone was either moved horizontally in x and z axes, or rotated in y axis to capture different images. This process made sure that no two images were the same. We captured a total of 660 images inclusive of all the bands.

Data Processing

After obtaining the images from the drone, they were exported as TIFF files for further pre-processing. Each drone snapshot includes RGB and Blue, Green, Red , Red Edge and NIR Bands. For each band we create its reflectance image to effectively detect plastics on the surface. We make use of a Geographic transform library named Rasterio to carry out our data preparation tasks. Firstly each spectral band along with the RGB image is s

Table 1

Band	Value
Blue	0.67
Green	0.69
Red	0.68
Red Edge	0.67
NIR	0.61

Panel Calibration Values

tacked to create a 6-dimensional vector of images. For each image, we get their bands using Ratserio and Gdal. Since images need to be represented as RGB Images on 2-Dimensional plots, we consider only one raster band since all 3 bands represent the same digital values.

Now we convert each band into its corresponding reflectance image. Table 1 represents the panel calibration of the drone considered. We first define the panel region and find the radiance image. The following equations describe the computation of the radiance image.

$$R = D\lambda + \theta \quad (1)$$

In (1), R represents the radiance image, D represents the digital value or pixel for the corresponding band . λ represents the gain and θ represents the offset. Both gain and offset were computed by us based on the surface on which the plastic was placed and the band. Table 2 describes the calibration of both gain and offset based on the band used by us . We use these gain and offset values to obtain the radiance image. We use the radiance image to get the panel region. The panel values are based on the drone calibrations.

Table 2

Band	Gain	Offset
Blue	8.000	1.309057
Green	8.000	0.885130
Red	4.500	0.748694
Red Edge	4.500	0.827855
NIR	5.000	0.833199

Gain and Offset Values

The coordinates of the panel region is considered as $[799{,}647{,}840{,}670]$ each corresponding to $[x_1, y_1, x_2, y_2]$. We compute the mean values of the panel region which represents the mean radiance. Equation (2) describes the computation of mean radiance.

$$\mu(I_b) = \frac{\sum_{i=1}^{i=N} P_i}{N} \quad (2)$$

Here, I_b represents the image of a particular band, P_i represents the digital value within the panel region and N represents the total number of values within the panel region. This is followed by the computation of the radiance to reflectance ratio. Equation (3) describes the computation of the same.

$$\delta = \frac{\rho}{L_e} \quad (3)$$

In (3), ρ represents the panel reflectance and L_e represents the mean radiance computed as per (2). Finally, the conversion factor δ is multiplied by the radiance image to find the reflectance image. Equation (4) describes this computation.

$$\rho_b = R_b * \delta \quad (4)$$

Data Preparation

After computing the reflectance image of each band, we set up a mathematical mapping. Equation (5) represents a transformation.

$$F(I) = \rho \quad (5)$$

In (5) I represents the RGB Image and ρ represents the corresponding reflectance image. We aim to effectively perform this mapping to detect regions of plastic and non-plastic based on reflectance values. Further this framework reduces the computational steps needed to compute the reflectance image based on surface and spectral characteristics.

We do not perform any spectral data augmentation techniques on the RGB image owing to loss of spectral features on gaussian blurring, random hue transforms. Certain techniques such as random flipping and rotation to a noise factor of 0.7 were performed considering there were only 660 samples and 110 after stacking the bands on the channel dimension.

A 80-20 random train-test split was performed on the samples for better generalisation for testing and inference.

Convolutional Neural Network Model

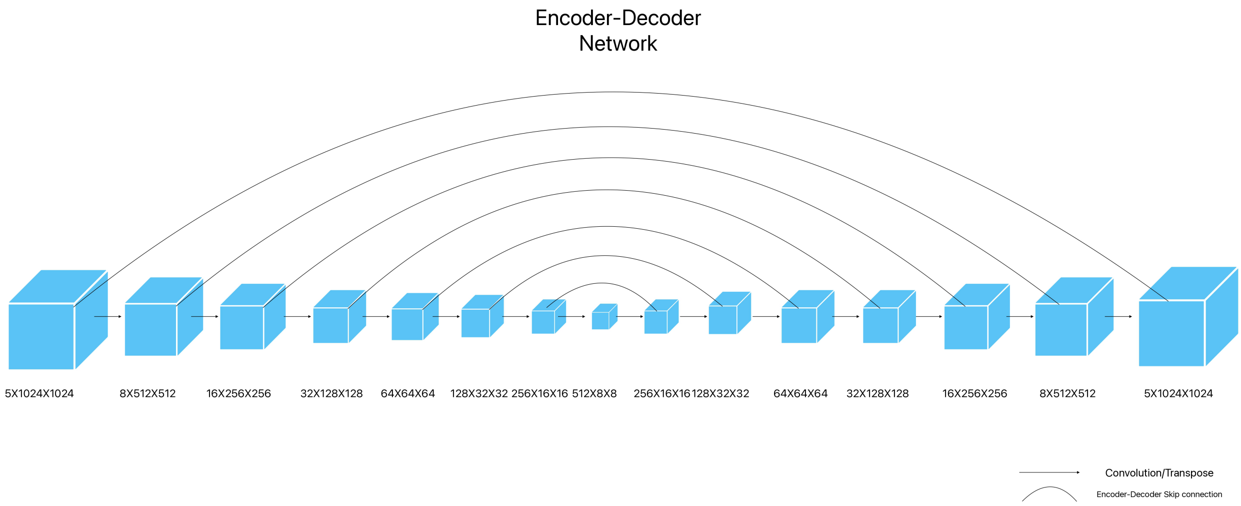

Figure 2: Architecture of Encoder-Decoder framework

After preparing the dataset both X and Y, we now define a deep learning framework to learn an effective mapping. The aim of the deep learning framework is to predict the reflectance images for each band in order to detect plastic and non-plastic regions.

To accomplish this task, we make use of a popular Convolution Neural Network architecture known as SegNet. Figure 2 represents the architecture of the CNN framework we used. The architectural network consists of 7 convolutional blocks in the encoder and the same number of blocks in th e decoder. Each layer in the encoder comprises of 3 components (i)Convolutional layer, (ii)Batch Normalisation (iii) Activation. Double stride convolutional layers square kernels of dimension 3 have been used in every layer to downsample the datapoints. The input for the network is an image of shape (B, 5,1024,1024)where B refers to batch size, 5 is the number of bands 1024X1024 is the input image resolution. The architecture reduces the image to a feature embedding of dimension . This embedding at the latent space is capable of extracting low-level characteristics of all 5 bands which represent the region s of plastic and non-plastic. As seen in figure 2, skip connections are included by adding connecting feature maps from encoder to decoder layer. This enables the network to retain information from the encoder required for decoding the image and is important since there is no visible feature difference between crop and weed, therefore low-level features learnt are retained . For reconstruction, the architecture consist of fractional stride convolutions of factor 0.5, it is to note that this operation is usually mistaken for de-convolutions. The number of kernels is tuned as an exponential growth for the encoder network which learns ,where stands for number of layers, feature mappings at the latent space, and the inverse mapping is used for tuning the number of kernels for the decoder network. The usage of Batch normalisation is prominent in all layers of the network since it enables faster learning by reducing internal covariate shift and has a slight L2 regularisation effect. To prevent overfitting, the proposed architecture makes use of 2D-Dropout with a noise parameter of 0.8 owing to very few samples.

Hardware and Software Setup

The experiments were performed on a 2022 model Macbook Pro, with Apple Silicon M2 chip. The configurations of the system are inclusive of a 8GB unified memory with 256GB Hard disk storage. The system has 8 Core CPU, which has a split-up of 4 performance cores and 4 efficiency cores, and a 10 core GPU. In addition to this, the system comes with a 16 core neural engine. The models used for the experiment were trained on PyTorch 2.0.1 with the help of helper image transforms package Torchvision version 0.15.2 . It is to note that the Beta version of Torchvision was used for image transformations. PyTorch makes use of the Metal Performance Shaders backend for accelerated GPU training.

Quantitative Results

Training Loss Progression

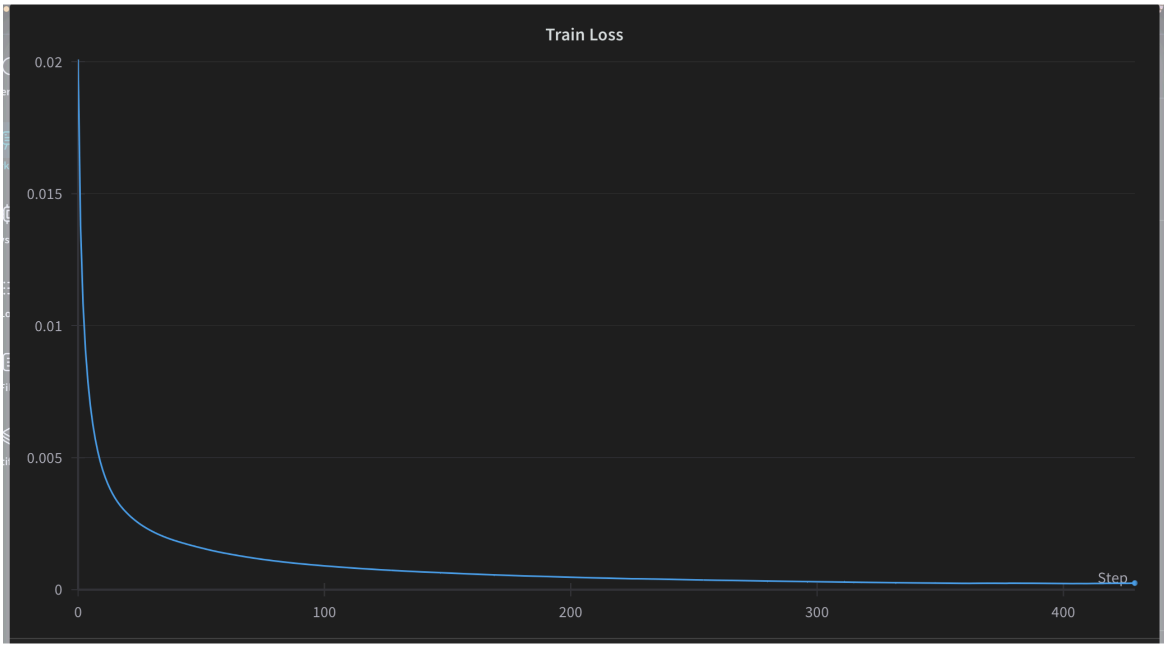

Figure 3

Training Stability using Mean Square Error Loss

As discussed above, we trained the model using the mean square error loss in order to re-construct the reflectance image accurately. The following figures represent the mean-square error progression with time while training the model. As observed in figure 3, the training progression is smooth and stable**. The overall reconstruction loss after 500 epochs is approximately 0.001**. This shows that the model captures all the underlying features and constructs a high quality reflectance image.

Mean Predicted Reflectance

We consider the distance of separation between the mean predicted reflectance of each band and the actual reflectance. Using the mean

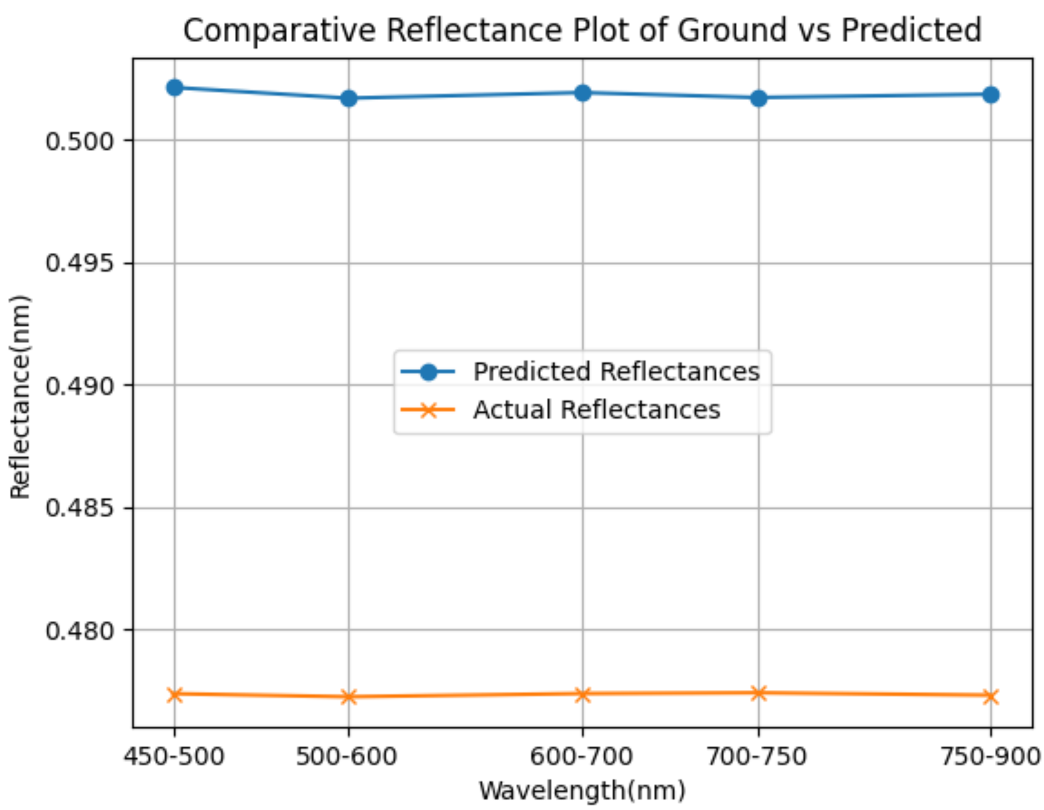

Figure 4

predicted reflectance, we threshold the reflectances of each band and classify them as plastic or non plastic regions. Therefore, accurate prediction of reflectances for each band is a necessity which may not be captured by the overall mean square error. Figures 4 and 5 represents the predicted vs actual reflectances for each band. As observed in these figures, the difference between the ground truth and predicted reflectances for each

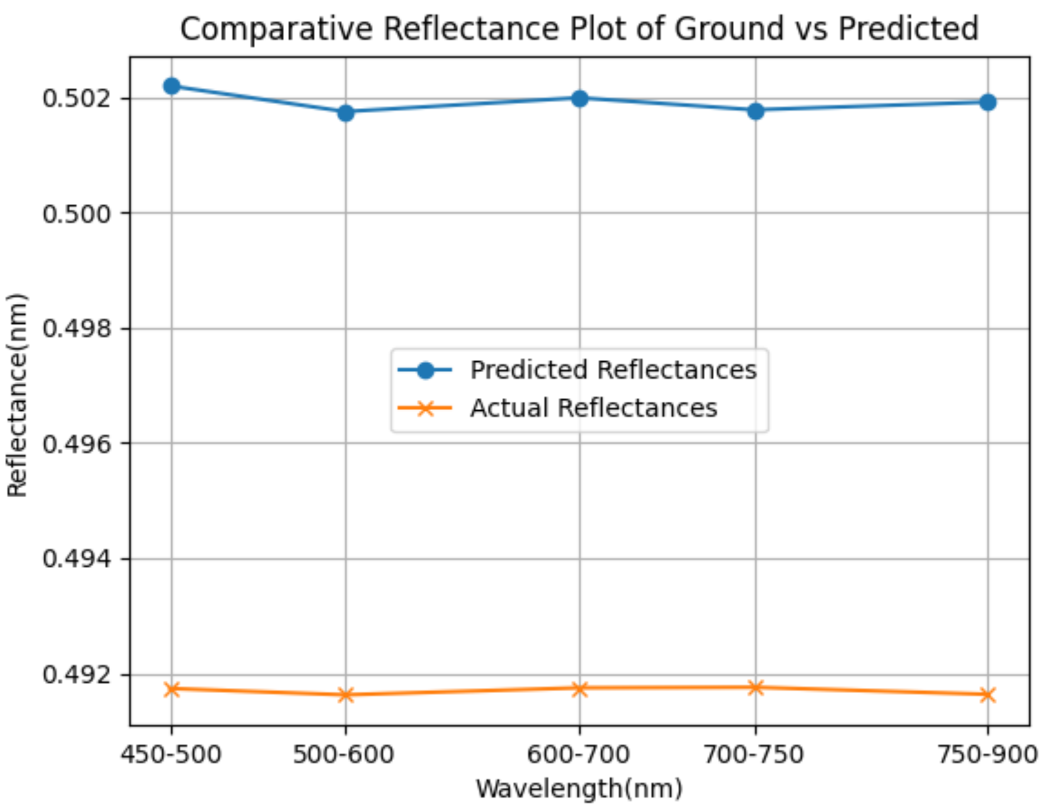

Figure 5

band is negligible. Moreover, even the reflectance trends over each band is very similar further validating our proposed model architecture and training process. This indicates that this framework can be used to detect plastic in images using hyper-spectral bands.

Qualitative Results

In this section, we show the reflectance images for each band, both predicted and actual. Figures 6-10 represents the image of reflectance for each band by comparing both predicted and actual reflectances. From these figures we can observe that, Blue has the highest reflectance for plastic and detects plastic to the highest precision while other bands have low reflectances around the same region. This is indicated by the reflectance graphs as well with the curve having the highest deviation from the blue band to the green band. We can see that the model captures most of the

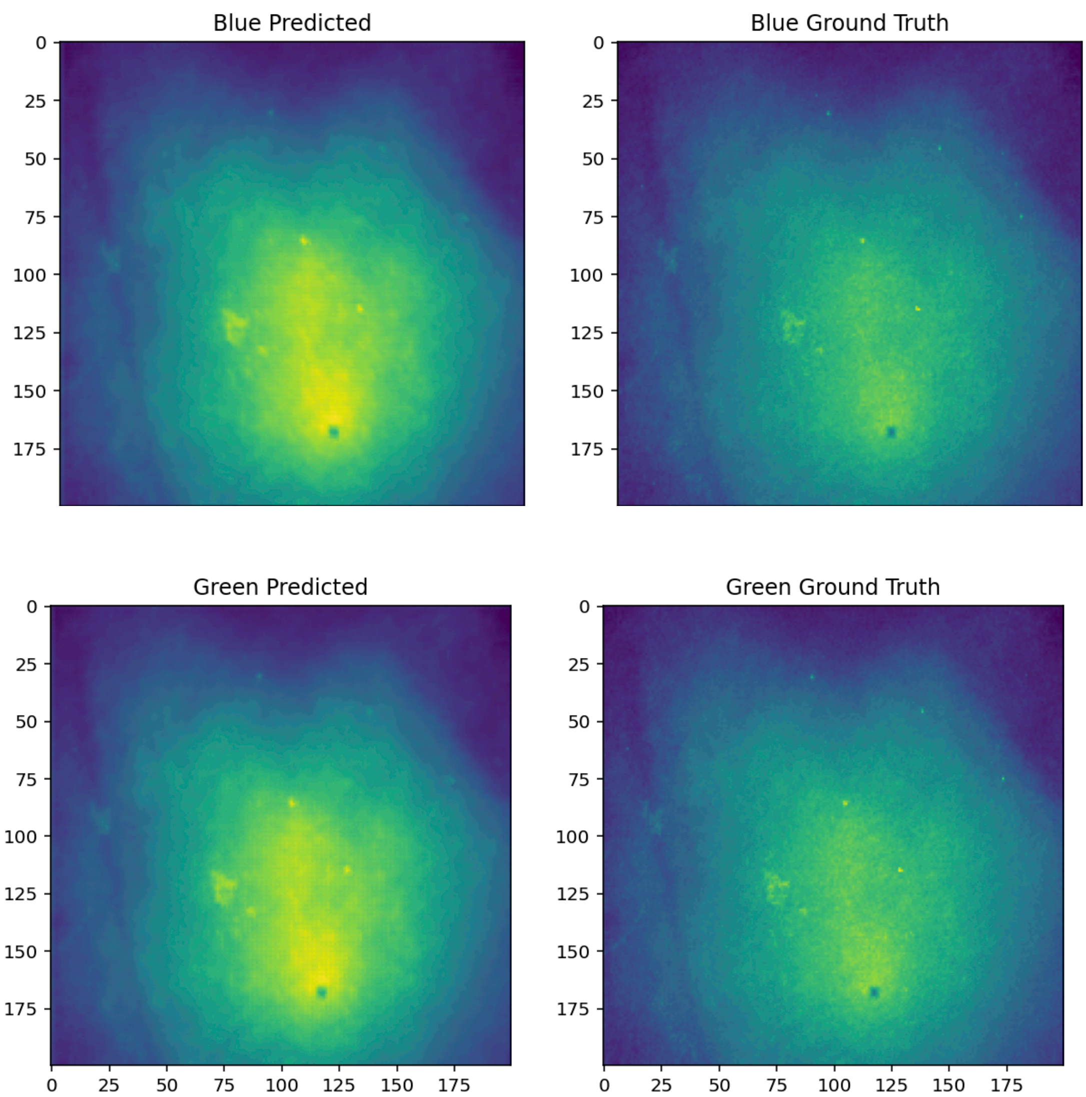

Figure 7

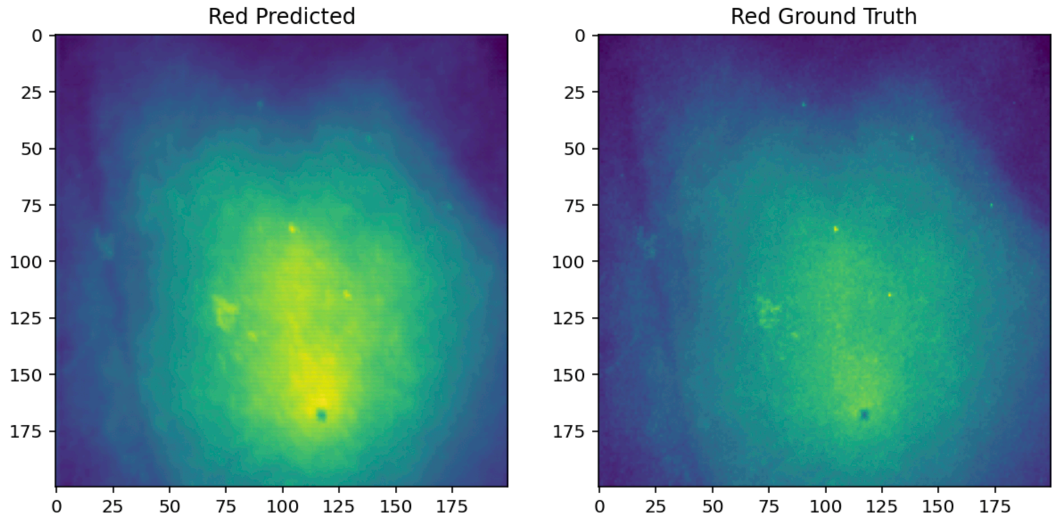

Figure 8

underlying hyper-spectral characteristics and detects the regions of plastic further validating our proposed approach.

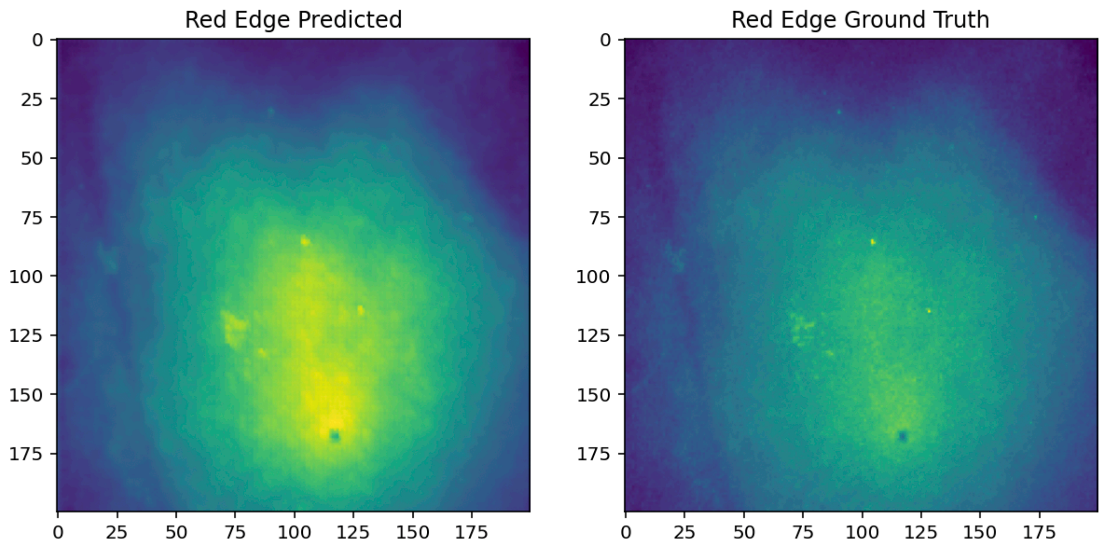

Figure 9

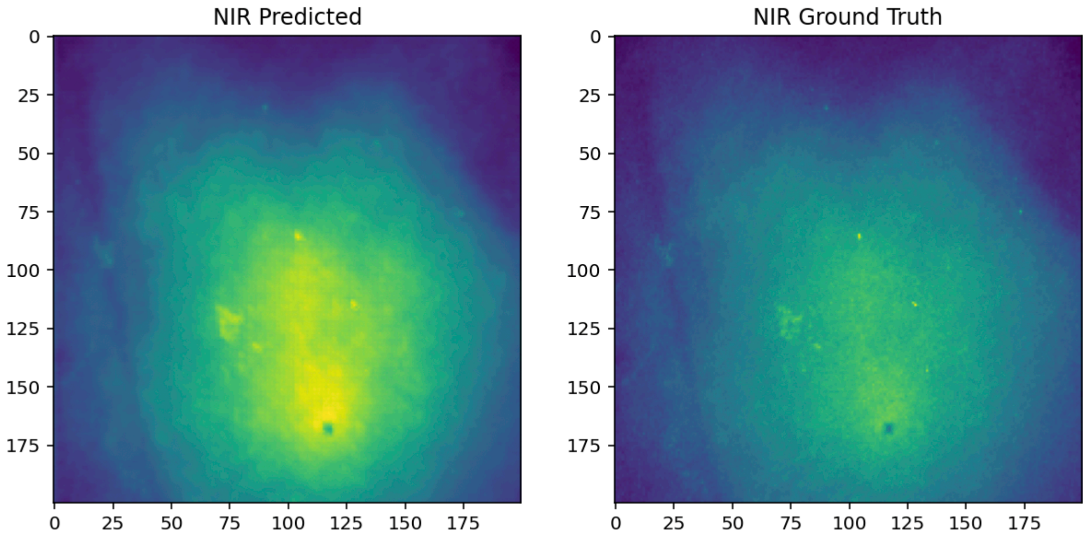

Figure 10

Conclusion and Further Work

In this study, we aim to propose a novel approach in detection of plastic. We make use of Hyper-spectral imaging by considering 5 bands namely Blue, Green, Red, Red Edge and NIR. We captured images using a drone flown at a height of 70 feet to ensure the dataset is built around plastic of different sizes such as in a real-world scenario. We make use of the reflectance values of each of the spectral bands to detect plastic and non-plastic regions by building a convolution neural network based on the architecture of the SegNet. The model inputs an RGB image and directly maps it into the reflectance image for each of the 5 bands. Based on these reflectance values, plastic regions are detected. Our model is able to capture most of the underlying characteristics and accurately detect the regions of plastic. As a future work, we aim to build a anomaly-detection based contrastive network to detect the plastic regions of the predicted reflectance images instead of using basic thresholding. Further, we aim to improve the resolution and quality of the predicted reflectance images by using complex generative architectures such as pix2pix GANs. This is a huge step in automated detection and removal of surface plastics which is a need of the hour requirement in the current global scenario.

Bibliography

- Cp, Sivaram & Sekar, Mukesh & K S, Rajkhishore. (2022). PLASTIC DETECTION AND CLASSIFICATION USING DEEP LEARNING A PROJECT REPORT (Project Work I -Phase I).
- M, Bhanumathi & DhanyaS, & R, Gugan & KirthikaKG,. (2022). Marine Plastic Detection Using Deep Learning. 10.3233/APC220057.
- Roslan, Muhammad Izzat Bin, Zaidah Ibrahim, and Zalilah Abd Aziz. "Real-time plastic surface defect detection using deep learning." *2022 IEEE 12th symposium on computer applications & industrial electronics (ISCAIE)*. IEEE, 2022.
- Mhadlekar, Siddhi S., Vedika S. Nalawade, and Pratiksha M. Patil. "PLASTIC DETECTION AND CLASSIFICATION USING DEEP LEARNING NEURAL NETWORK." (2022).
- Salim, Aria Amadeus, et al. "PLEESE-Plastic Reuse Reminder System using Computer Vision Deep Learning." *2022 1st International Conference on Technology Innovation and Its Applications (ICTIIA)*. IEEE, 2022.
- Shenming, Q., Xiang, L. and Zhihua, G. (2022). A new hyperspectral image classification method based on spatial-spectral features. *Scientific Reports*, 12(1). doi:https://doi.org/10.1038/s41598-022-05422-5.
- Tamin, Owen, et al. "A comparison of RGB and RGNIR color spaces for plastic waste detection using the YOLOv5 architecture." *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*. IEEE, 2022.
- Tamin, Owen, et al. "Machine learning for plastic waste detection: State-of-the-art, challenges, and solutions." *2022 International Conference on Communications, Information, Electronic and Energy Systems (CIEES)*. IEEE, 2022.

- Vasarhelyi, K. (2023). *The impact of plastic on climate change*. [online] Environmental Center. Available at: https://www.colorado.edu/ecenter/2023/12/15/impact-plastic-climate-change#:~:text=The%20refinement%20of%20plastics%20emits.
- Wen, S., Yuan, Y. and Chen, J. (2023). A Vision Detection Scheme Based on Deep Learning in a Waste Plastics Sorting System. *Applied sciences*, 13(7), pp.4634–4634. doi:https://doi.org/10.3390/app13074634.
- Zhou, Shanyu, et al. "Function Assignment of Plastics based on Hyperspectral Satellite Images and High-Resolution Data Using Deep Learning Algorithms." *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2023.