

# A Deep Learning Approach for Pixel-level Material Classification via Hyperspectral Imaging

Savvas Sifnaios, George Arvanitakis, Fotios K. Konstantinidis, Georgios Tsimiklis, Angelos Amditis, Panayiotis Frangos

**Abstract**—Recent advancements in computer vision, particularly in detection, segmentation, and classification, have significantly impacted various domains. However, these advancements are tied to RGB-based systems, which are insufficient for applications in industries like waste sorting, pharmaceuticals, and defense, where advanced object characterization beyond shape or color is necessary. Hyperspectral (HS) imaging, capturing both spectral and spatial information, addresses these limitations and offers advantages over conventional technologies such as X-ray fluorescence and Raman spectroscopy, particularly in terms of speed, cost, and safety.

This study evaluates the potential of combining HS imaging with deep learning for material characterization. The research involves: i) designing an experimental setup with HS camera, conveyor, and controlled lighting; ii) generating a multi-object dataset of various plastics (HDPE, PET, PP, PS) with semi-automated mask generation and Raman spectroscopy-based labeling; and iii) developing a deep learning model trained on HS images for pixel-level material classification. The model achieved 99.94% classification accuracy, demonstrating robustness in color, size, and shape invariance, and effectively handling material overlap. Limitations, such as challenges with black objects, are also discussed. Extending computer vision beyond RGB to HS imaging proves feasible, overcoming major limitations of traditional methods and showing strong potential for future applications.

**Index Terms**—Hyperspectral Imaging, Deep Learning, Material Classification, Pixel-level Classification, Real-time Object Detection

## I. INTRODUCTION

THE field of material classification has evolved significantly over the past few decades, transitioning from traditional techniques to the application of deep learning methodologies. Traditional methods, such as thresholding, edge detection, and classical machine learning algorithms (e.g., k-nearest neighbors, support vector machines), rely heavily on manually crafted features and heuristic rules. These methods are often effective for simple tasks but struggle with complex and fine-grained material differentiation due to their inherent limitations in capturing shape and colour variations.

S. Sifnaios, G. Arvanitakis, F. K. Konstantinidis, G. Tsimiklis and A. Amditis are with Institute of Communication and Computer Systems, National Technical University of Athens, 9 Iroon. Polytechniou Str., Zografou Athens GR-157 73, Greece.

S. Sifnaios and P. Frangos are with School of Electrical and Computing Engineering, National Technical University of Athens, 9 Iroon. Polytechniou Str., Zografou Athens GR-157 73, Greece.

G. Arvanitakis is with Technology Innovation Institute, Yas Island, Abu Dhabi, United Arab Emirates

S. Sifnaios is the corresponding author (savvas.sifnaios@iccs.gr)

Manuscript received November , 2024; revised XX, 202X.

Image classification and segmentation have undergone a paradigm shift with the introduction of deep learning, especially convolutional neural networks (CNNs). Material classification tasks have improved in accuracy and robustness through deep learning models that automatically learn hierarchical features from raw data. However, despite these advancements, challenges remain when using conventional RGB images, especially in scenarios where the color or the shape of the object is not that informative regarding its material.

RGB images solely rely on spatial features—such as edges, textures, and colours visible to the human eye. This reliance often leads to misclassification when objects of different classes have similar appearances, as shown in Figure 1. For instance, in the sorting industry, accurately distinguishing between materials with subtle colour differences or complex shapes is crucial for efficiency and quality control. In the pharmaceutical industry, ensuring the purity and correct identification of compounds necessitates fine segmentation capabilities [1]. Similarly, in agriculture, RGB imaging often falls short in detecting diseases in crops that present with subtle spectral variations not visible in the RGB spectrum [2]. In defense applications, accurate material classification can aid in the detection and identification of hazardous substances, ensuring safety and operational efficiency.

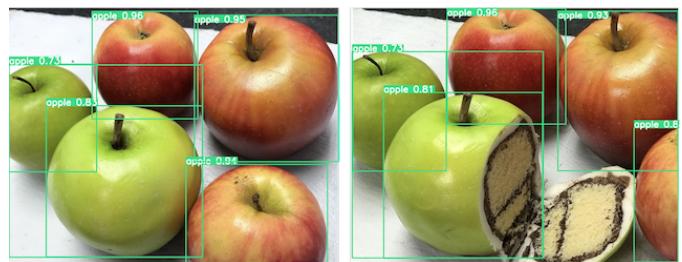


Fig. 1. Example of RGB-based object misclassification. This image depicts pastry cakes that reassemble apples leading the RGB-based model to mistakenly classify the samples as apples.

Hyperspectral imaging has emerged as a promising solution to these challenges. Unlike traditional RGB imaging, hyperspectral cameras capture a wide spectrum of light beyond the visible range, providing detailed spectral information for each pixel in an image. This richness in spectral data allows for more precise material classification and segmentation, overcoming the limitations of traditional methods.

Despite its potential and the direction of mobile manufacturers' to add more spectrum bands on their phone's cameras [3]–[5], the research on computer vision beyond

RGB imaging is very limited mainly because of the **lack of publicly available datasets** and the need for **specialized hardware** that up to this day is expensive. In this paper, we propose P1CH (Pixel-wise 1D Convolutional Hyperspectral) Classifier, a lightweight convolutional neural network designed to perform image segmentation and classification using hyperspectral images. The spectral information captured by the employed hyperspectral camera ranges from 900nm up to 1700nm. The proposed approach leverages the spectral richness of hyperspectral data to enhance the accuracy of material classification, addressing the limitations of traditional methods. This advancement opens new possibilities for applications in various industries, providing a practical solution to the longstanding problem of precise material classification and segmentation. The contributions of this work are summarized as follows:

- **Dataset:** The generation and release of a dataset, containing hyperspectral images and their corresponding labeled masks, of plastic samples (HDPE, PET, PP, PS), with the aim to avail research community to investigate other models, applications, insights in the field of computer vision via hyperspectral imaging.
- **P1CH:** The design and implementation of a lightweight yet accurate (99.94% accuracy) deep learning approach, for pixel-level material classification.
- **Insights:** Extensive performance analysis highlighting the capacity (accuracy, color invariance, shape invariance) as well as the limitations (e.g. black plastics) of the computer vision with hyperspectral imaging.
- **Cost:** Taking advantage of the deep learning capabilities, cost-efficient alternatives for calibration and normalization of Hyperspectral images are proposed tailored for deep learning models.

To this end, Section II describes the related work, while Section III provides a high-level description of the system setup and the hardware equipment used in this work. The creation of hyperpectral dataset, the spectral preprocessing methodologies along with the semi-automated AI-assisted mask generation are presented in Section IV. Section V introduces the architecture of the proposed model, the preparation of the training instances. The training process process, along with the selected hyperparameters and the initial results are also discussed in that section. In addition, the experimental results on various scenarios, i.e. *mixed materials*, *shredded samples* and *overlapping objects*, and the performance of the proposed system are discussed in Section VI. This section, also, explores the limitations of the hyperspectral model in the case of dark-coloured or black samples. Finally, in Section VII a discussion about the main conclusions of this study and the future steps are provided.

## II. RELATED WORK

Material classification has historically relied on several approaches that cut across many domains, including electro-mechanical and chemical analysis techniques. Raman and Near-Infrared spectroscopy, coupled with a multivariate analysis, have widely been used to identify the materials' composition [6]–[8]. In another work, the utilisation of X-Ray

Diffraction (XRD), Energy Dispersive X-ray Spectroscopy (EDS) and Atomic Absorption Spectroscopy (AAS) techniques have been proposed, because their high sensitivity and accuracy in detecting microstructural features and hence identifying the element composition of a sample [9]. Zhang et al. [10] highlighted the use of Differential Scanning Calorimetry (DSC) and Thermogravimetric Analysis (TGA) in accurately determining the thermal properties of various materials . Zhang and Shao [11] emphasized the role of optical microscopy in material.

In an attempt to further increase the performance and robustness of material classification pipelines, various studies explored the potential of deep learning. In [12] a deep convolutional neural network (CNN) was introduced for classifying 60-GHz radar data with an accuracy of 97%. Deep neural networks have, also, been utilised to analyse surface haptic data [13], achieving a precision of 94% and recall of 92%. Moreover, CNNs have been proposed for the case of visual RGB images [14]–[17] in various use cases, e.g. commercial waste, steel products, etc. In another work, Konstantinidis et al. [18] proposed a multi-modal deep classifier based on ResNet-18 that jointly extracts information from RGB and multispectral cameras to classify plastic polymers, as well as wood by products with an object accuracy of 96%.

While the cited work above demonstrates impressive classification performance, these methods are often slow and computationally intensive. Hyperspectral imaging has become a powerful means for in situ material classification, since they offer the advantage of real-time and efficient analysis. Shaikh and Thörnberg [19] investigated the impact of water vapor on polymer identification by means of short-wave infrared hyperspectral imaging yielding an accuracy rate around 88%. According to Shaban [20], hyperspectral imaging has been used to determine different characteristics pertaining to concrete without disturbing its structure at sites, where it would not be possible to take samples back to laboratories for analysis; the typical accuracies are above 90%. With about 93% accuracy Capobianco et al. [21] have characterized ancient roman wall paintings using Hyperspectral Imaging as well as aid in artwork authentication (Polak et al. [22]), achieving a precision of 95%.

The combination of deep learning and hyperspectral imaging has shown great promise in improving the material classification capability in Earth Observation applications. Notable examples include models, e.g. Xception-based, CNN-based systems, and R-VCANet achieving accuracies up to 99%, with precision and recall values around 94% and 93%, and F1 score reaching 91% [23]–[27]. Venkatesan et al. [28] applied deep recurrent neural networks in medical hyperspectral images to achieve feature recognition with a precision rate of about 96%. Xiong et al. [29] developed material tracking methods for hyperspectral videos using deep learning, achieving an F1 score of 94%. Medus et al. [30] applied CNNs to classify hyperspectral images in industrial food packaging, reporting an accuracy of 99%. Okada et al. followed a patch-wise approach for the identification of 5 different mineral types, using a VGG16-based CNN to classify the acquired HS images they achieved high accuracy over 90% [31]. Extreme Learning

combined with Stacked Autoencoders for feature extraction has also been applied on HS images, in [32], for the detection of plastic films within cotton feed stock, creating a pixel-level classification map with accuracy up to 95%. Zhu et al. in [33], also attempted to classify different cotton seeds. Their work encompasses a multivariate analysis for manual extraction of the 10 most informative features, which were subsequently fed to CNN for the classification task, resulting in a prediction accuracy of 88%. Moreover, Artzai et al. worked with HS images with 76 bands of non ferrous metals, and utilised them to train a CNN-based U-Net like network [34]. This approach analysed the HS images as a whole and created a classification map with an accuracy equal to 95%. Finally, Roy et al. [35] in their work proposed a methodology that incorporates PCA as a preprocessing step with a deep convolutional network, i.e. HybridSN, for the pixel-wise classification of Hyperspectral Images, achieving an accuracy score of 99%.

The aforementioned studies with hyperspectral classification using deep learning have yielded remarkable results. However, it is important to note that most of the already limited approaches that exist in the field commonly either deal with HS images in batches or utilise 2D Convolutional Neural Networks (CNNs). In order to achieve this, they use 2D convolutional kernels that inevitably lead to loss of accuracy at object borders. Notably, for 2D CNNs convolutions are designed to capture spatial context within each individual frame but this may blur boundaries or misclassify pixels on the boundary between neighboring objects that have different spectral signatures. This means that spectral continuity across edges is not incorporated by the 2D convolution operation, causing edge artifacts and reducing classification accuracy here. These limitations highlight the need for more refined techniques that analyse HS images on a pixel level and can preserve boundary integrity while leveraging the rich spectral data provided by hyperspectral imaging.

### III. HYPERSPECTRAL IMAGING SETUP

This section provides a detailed description of the physical infrastructure of the cyber-physical system where the dataset was created using materials that were conveyed under the vision and hyperspectral sensors.

#### A. Imaging Spectroscopy

Hyperspectral imaging combines conventional imaging and spectroscopic methodologies with the goal to simultaneously obtain spatial and spectral information from various wavelengths of the electromagnetic spectrum for every individual pixel in an image of a scene, with the objective of locating objects, classifying materials, or detecting processes [36]. Pushbroom sensors capture spectral information across a swath as the sensor moves, line scan sensors capture data one line at a time, while whiskbroom sensors, scan point-by-point to build an image, and snapshot sensors capture the entire spectral image in a single exposure.

#### B. Camera

The SPECIM FX17 line scan camera, following push broom technology, utilises a matrix detector and an imaging spectrograph to capture spectral data efficiently. Light enters through high-performance optics and an entrance slit, forming a line image that the spectrograph disperses into a spectrum (900 - 1700 nm, across 224 bands). This setup allows each axis of the detector to record spatial position and spectral information simultaneously. It ensures measurement stability despite sample or camera movement, requires less illumination power while achieving higher intensity, and is significantly more efficient than filter-based cameras, yielding a purer spectrum.

1) *Conveyor*: Once the line scan camera is selected for capturing, the system comprises a conveyor belt that facilitates the horizontal movement of objects at an adjustable speed, in order to achieve synchronised sampling frequency of the hyperspectral sensor and objects' movements. The camera was placed 0.73 cm from the conveyor, while the illumination source was placed 0.5 meters vertically above the conveyor belt, at a 45-degree angle, and 0.3 meters horizontally from the center of the camera.

#### C. Illumination

The deployed custom-made LDL-222X42CIR Full-Spectrum bar light manufactured by CCS Inc., offers better performance for Imaging spectroscopy challenges in contrast with Led lights [37]. Specifically, it includes four different halogen bulbs that emit light in distinct regions of the spectrum and a power rating of 87W. This permits it to span the complete range of wavelengths from 400nm to 2400nm. This light source contains a dispersion layer that uniformly distributes the output light in multiple directions, resulting in more consistent illumination of the scene and the objects within it.

#### D. Acquired Data

The HS camera employed in this work is ample of achieving a maximum sampling rate of 400 FPS, with each scanned line being of shape  $640 \times 224$ . Moreover, camera provides the option of applying on-chip spatial and/or spectral binning, hence reducing the respective dimensions by factors of  $\times 2$ ,  $\times 4$ , or  $\times 8$ . In this work, no spectral or spatial binning was applied, thus the acquired data were of shape  $n_{rows} \times 640 \times 224$ , where  $n_{rows}$  is depended on the selected sampling rate and the acquisition duration. Finally, it is worth mentioning that the pixel size of the selected camera is 0.9375mm, thus allowing for very precise and accurate segmentation.

### IV. DATASET

Following the aforementioned hyperspectral imaging setup, this section presents the dataset used in the experimental part of this work. At first, a comprehensive description of the dataset is provided, highlighting the different material classes involved in the dataset and summarizing the number of images, objects, and pixels for each category. The following subsections address the processes involved in dataset preparation,

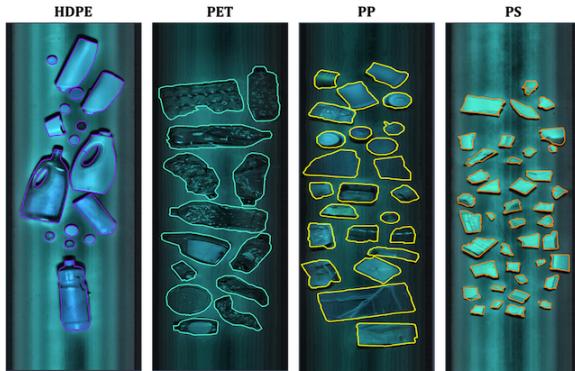


Fig. 2. The false colour, contrast stretched HS images used for the generation of the training set.

including the normalization of the images to ensure numerical consistency. The generation of ground truth masks is subsequently discussed, pointing out their necessity for the proposed supervised learning approach. Moreover, the reasoning behind the utilization of Raman spectroscopy throughout labeling process is explained. Finally, the generation of the training batches and their format is outlined.

#### A. Dataset description

For the generation of the train and test dataset an ensemble of plastic samples of different material categories were collected consisting of *HDPE*, *PET*, *PP*, and *PS* objects.

**Train Set:** In detail, 169 objects were selected to represent the classes. The physical objects were separated according to their material classes and subsequently placed randomly on the conveyor belt, described in Section III-B. For the augmentation of the training set by every sample was captured twice with slightly different light conditions. A comprehensive analysis of the dataset composition is presented in Table I.

TABLE I  
SUMMARY OF IMAGES, OBJECTS, AND PIXELS PER CATEGORY OF THE TRAINING SET.

Material	Images	Objects	Pixels
HDPE	2	40	794,142
PET	2	24	810,483
PP	2	45	831,340
PS	2	60	801,936

At this point, it is important to mention that a 5<sup>th</sup> class was introduced to the dataset. Namely, this class is *Background* and it represents the set of pixels depicting the conveyor belt's surface. Since the conveyor belt is apparent in every HS image, no matter the type of plastic, initially the number of *Background* pixels was  $\times 10$  higher than any other class, potentially leading the model to be biased in favor of this class. To handle this, a random subset of *Background* pixels was selected, ensuring its size will be equal to the maximum number of pixels of the plastic classes, i.e. 831,340 pixels.

**Test Set:** Regarding the test set, was prepared from different objects than the train set, but that belong to the same material categories. Additionally the dataset was selected in order to

capture four main cases i) *unmodified objects*, similar to the train set ii) *shredded objects*, the shape do not capture any information iii) Mixed-overlapping materials, where a material is into another material, Table II presents the dimensions of the test set.

TABLE II  
TEST SET WITHOUT BLACK SAMPLES.

Material	Objects	Pixel
HDPE	29	124,109
PET	20	100,445
PP	2	59,682
PS	29	116,704

TABLE III  
BLACK SAMPLES TEST SUBSET.

Material	Objects	Pixels
HDPE	19	31,649
PET	0	0
PP	14	25,280
PS	5	5,035

Furthermore, regarding the literature [38], [39], HS cameras (up to 1700nm) cannot capture information regarding plastics that are coloured black. In order to examine that an extra black plastic dataset was collected and presented in the Table III

#### B. Acquisition Pipeline

This section describes the process of converting the captured pixel values from 16-bit **unsigned** integers, in the range of [0, 4095] to compatible for deep learning and ready-to-used 32-bit floats pixel values in the range of [0., 1.].

Throughout the literature it is strongly suggested a) to use the **spectral calibration matrix**, that is specifically provided by the camera manufacturer, and post the HS image's multiplication with spectral calibration matrix, b) to **normalize** the calibrated data using **White** and **Black** reference.

**White reference:** A white reflection target is required with the key property of reflecting the incident radiation uniformly across all wavelengths. In this manner, the real maximum absolute pixel value a HS camera can capture, given the illumination source, is calculated and ultimately creates the white reference. Although should be taken into account that a white reflection target is often not commercially available and its cost grows exponentially with respect to its dimensions.

**Black reference:** On the other hand, the black reference is utilised to model the sensors' electronic noise, caused by the electrons' random movement due to the sensors' temperature. In this work, the black reference  $I_{\text{black}}(\lambda)$  was acquired, by closing the shutter of the camera and capturing 1000 lines, which were then averaged to create the black reference. For the maximum absolute pixel value the HS camera can capture, the assumption was made that it can be approximated by calculating the maximum pixel value within the train set.

By modeling that noise, one can subtract the black reference both from the HS image and the white reference, to acquire the noise-corrected version of both images. Equation 1 provides a mathematical formulation of the black-white reference normalization:

$$I_{\text{norm}}(x, y, \lambda) = \frac{I(x, y, \lambda) - I_{\text{black}}(\lambda)}{I_{\text{white}}(\lambda) - I_{\text{black}}(\lambda)} , \quad (1)$$

where  $I_{\text{norm}}(x, y, \lambda)$  is the normalized hyperspectral image at pixel  $(x, y)$  and wavelength  $\lambda$ ,  $I(x, y, \lambda)$  is the raw hyperspectral image at pixel  $(x, y)$  and wavelength  $\lambda$ ,  $I_{\text{black}}(\lambda)$  is

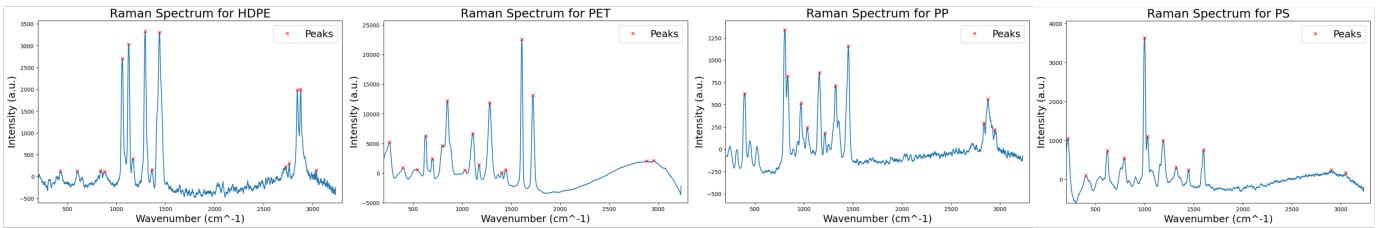


Fig. 3. The Raman spectra for the 4 material classes. For the HDPE spectrum, the peaks at  $1058\text{ cm}^{-1}$ ,  $1123\text{ cm}^{-1}$ ,  $1291\text{ cm}^{-1}$ ,  $1437\text{ cm}^{-1}$ , and the range from  $2843\text{-}2876\text{ cm}^{-1}$  are characteristic of high-density polyethylene (HDPE). These peaks correspond to the vibrations of the molecular structure of HDPE, specifically indicating the various stretching and bending vibrations of the C-H bonds. Accordingly, for the rest of the spectra, it is pointed out that two characteristic peaks at  $1607\text{ cm}^{-1}$  and  $1721\text{ cm}^{-1}$  were detected in the top-right plot and correspond to the vibrations of the phenyl group in the polyester. Additionally, the peaks in the range of  $1100\text{-}1200\text{ cm}^{-1}$  indicate the stretching vibrations of the C-O group. The peaks observed at  $970\text{ cm}^{-1}$ ,  $1034\text{ cm}^{-1}$ ,  $1360\text{ cm}^{-1}$ ,  $1453\text{ cm}^{-1}$ , and  $2946\text{ cm}^{-1}$  in the bottom-left plot are associated with the vibrations of the methyl group ( $\text{CH}_3$ ) in polypropylene, while the intense peak observed at  $1010\text{ cm}^{-1}$ , in the bottom-right plot, along with the peak at  $1598\text{ cm}^{-1}$ , suggest the presence of polystyrene.

the black reference image at wavelength  $\lambda$ ,  $I_{\text{white}}(\lambda)$  is the white reference image at wavelength  $\lambda$ .

The black reference  $I_{\text{black}}(\lambda)$  was acquired, by closing the shutter of the camera and capturing 1000 lines, which were then averaged to create the black reference. For the maximum absolute pixel value the HS camera can capture, the assumption was made that it can be approximated by calculating the maximum pixel value within the train set. Equation 2 provides an approximation of Equation 1 that does not require an expensive white reflection target:

$$I_{\text{norm}}(x, y, \lambda) = \frac{I(x, y, \lambda) - I_{\text{black}}(\lambda)}{M}, \quad (2)$$

where  $I_{\text{norm}}(x, y, \lambda)$  is the normalized hyperspectral image at pixel  $(x, y)$  and wavelength  $\lambda$ ,  $I(x, y, \lambda)$  is the raw hyperspectral image at pixel  $(x, y)$  and wavelength  $\lambda$ ,  $I_{\text{black}}(\lambda)$  is the black reference image at wavelength  $\lambda$ ,  $M$  is the dataset's maximum pixel value.

It should be noted that in HS image processing applications it is highly recommended to use the aforementioned steps for *spectral calibration* and *normalization* which often include expensive equipment and dependencies on the hardware manufacturer. However, as shown in Section VI, even if we totally ignore those steps, the learning performance of the proposed deep learning algorithm is not affected.

**Remark (Calibration on Training).** *The spectral calibration and the normalization as operations are a sequence of matrix multiplications, it seems by our experiments, that this transformation can be learned directly in the training process.*

### C. Ground truth mask generation

As in every supervised learning application, a set of ground truth labels is needed to ensure successful training of the neural network model. To this end, an AI-assisted methodology was deployed for the generation of the binary masks, presented in Figure 2, which later on will be utilised as training labels.

**Semi-Automated Segmentation:** The first step in the proposed methodology, is the creation of a false-colour RGB version of the HS image. To this end, the Standard Deviation of each channel in the original image was calculated, as a measure of its contrast. The three channels with the highest

contrast were selected and sorted in ascending order of wavelength, for each image, in order to create the false-colour RGB image. An adaptive histogram stretching algorithm was, also, applied to the respective RGB versions in order to further increase the contrast and make the objects' edges as sharp as possible without altering the spatial content of each image.

The false-colour, histogram stretched images were subsequently utilised for the generation of segmentation masks. A ViT model, namely SAM (Segment Anything) [40], for the semi-automated mask generation task. In detail, positive (pixel to be included in the mask) and negative (pixel to be excluded from the mask) points were given as prompts to the model in order to generate a first estimation of the mask. The predicted mask was then visually inspected and refined, when needed, aiming for precision maximization at the boundaries of the object. This procedure was repeated for every image and every object depicted within an image of the dataset and the final results can be seen in Figure 2.

**Labeling - Raman:** The final step of the ground truth generation is to assign a class to each of the aforementioned masks. To this end, Raman Spectroscopy was employed.

Raman spectroscopy is a powerful analytical technique used to observe vibrational, rotational, and other low-frequency modes in a system. It relies on inelastic scattering, or Raman scattering, of monochromatic light, typically from a laser. When light interacts with molecules, vibrations, or other excitations in the system occur shifting up or down the energy of the laser, creating peaks in the acquired spectrum and hence providing a fingerprint by which molecules can be identified [41], [42]. In Figure 3, an example of the spectrum for each of the 4 plastic types is presented, where the red marked peaks indicate the existence of each polymer in under examination the plastic sample.

In this manner, each sample annotated in the previous step was individually scanned with the Raman equipment and its spectrum was analysed in order to identify the indicative, for each class, peaks. The results of the Raman Spectroscopy analysis, were used as the class of each of the aforementioned masks.

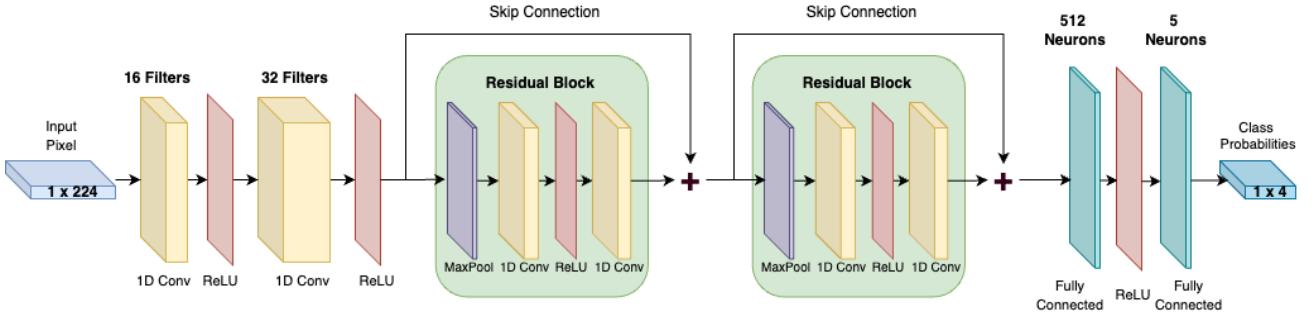


Fig. 4. The high-level architecture of the proposed Pixel-wise 1D Convolutional Hyperspectral (P1CH) Classifier .

## V. MODEL ARCHITECTURE & TRAINING

### A. Architecture

The proposed hyperspectral image classification model uses a **1D CNN architecture** to capture and process spectral information. The architecture uses a) 2 convolutional layers, b) 2 residual blocks, and c) 2 fully connected layers to accurately classify hyperspectral data. The architecture of the proposed Pixel-wise 1D Convolutional Hyperspectral Classifier is shown in Figure 4.

**Convolutions:** The input pixel of shape  $(1 \times 224)$  is passed through an initial convolutional layer with 16 size-3 filters and padding to preserve dimensions. Afterwards, a 32-filter, 3-size convolutional layer with padding follows. After each convolutional layer, a ReLU activation function introduces non-linearity and a max-pooling layer with a kernel size of 2 and stride of 2 reduces data dimensionality.

**Residuals:** The model uses two residual blocks for feature extraction and learning. The first residual block receives the output from the second convolutional layer and processes it through two convolutional layers with 64 filters each, maintaining a kernel size of 3. Batch normalization is also applied after each convolutional layer. This output is added to the block's input - through the utilisation of skip connection - and passed through a ReLU activation function. The second residual block follows a similar structure, but with 128 filters in each convolutional layer.

**Fully Connected:** The processed features are flattened and passed through a fully connected layer with 512 neurons, followed by a dropout layer. A second fully connected layer follows with its number of neurons set equal to the number of classes. Finally a soft-max layer provides the final classification probabilities.

### B. Data Pre-Processing

As mentioned in the *Architecture*, the model expects a 1-dimensional vector as input. In this work, the input vectors are the individual pixels of the dataset's images. Up to this point, however, the dataset consists of HS images, hence it is necessary to convert the images into a set of pixels.

To this end, a data handling pipeline was implemented to efficiently utilise the large volume of data encoded within HS images. In detail, a memory-mapped array uses the operating system's virtual memory capabilities to map a disk file directly

into the address space of the application, allowing for efficient, random access to large datasets without loading the entire file into memory. In this manner, by employing memory-mapped arrays out-of-core processing is achieved, which significantly reduces the memory footprint and improves the performance of data loading operations.

By exploiting the capabilities of memory-mapped arrays, the entire set of images, along with their respective ground truth masks are flattened across the spatial dimensions, thus creating the desired 1-dimensional vectors is 224 features each. The feature vectors are, subsequently, randomly shuffled and split in two subsets; the train and the validation set with ratios 90% and 10% respectively.

### C. Model Training

With the data preprocessing pipeline established, the focus can now shift on the model training phase, where the calibrated, and pre-processed data is being used to train the Pixel-wise 1D Convolutional Hyperspectral Classifier.

A dataloader was defined for each of the two sub-sets, each with a *batch size* equal to 640. The samples in the train set are shuffled in the beginning of each epoch, thus ensuring slightly different data distribution on batch-level in every iteration. The selected optimizer is Adam [43], and the initial learning rate was set to 0.001. To prevent the model being stuck to a local minima of the loss function, a learning rate (LR) scheduler was also implemented. The LR scheduler utilised in this work was *Cosine Annealing with Warmup* [44], [45]. The mathematical formulation of this scheduler is described in the following equation :

$$\eta_t = \begin{cases} \eta_{\text{init}} \frac{t}{T_w} & \text{if } t \leq T_w \\ \eta_{\text{min}} + \frac{1}{2}(\eta_{\text{max}} - \eta_{\text{min}}) \left(1 + \cos\left(\frac{(t-T_w)\pi}{T-T_w}\right)\right) & \text{if } t > T_w \end{cases},$$

where  $\eta_t$  is the learning rate at epoch  $t$ .  $\eta_{\text{init}}$  is the initial learning rate,  $\eta_{\text{max}}$  is the maximum learning rate,  $\eta_{\text{min}}$  is the minimum learning rate,  $T_w$  is the number of warmup epochs,  $T$  is the total number of epochs.

Since the model's aim is to classify each of the pixel in a HS image to their respective class, the *Cross-Entropy Loss Function* [46] was selected to be minimized throughout the training process as depicted in Equation 3.

$$\mathcal{L}_{\text{CE}} = - \sum_{i=1}^N \sum_{c=1}^C y_{ic} \log(\hat{y}_{ic}) , \quad (3)$$

where  $N$  is the number of samples,  $C$  is the number of classes.  $y_{ic}$  is the ground truth label (1 if sample  $i$  belongs to class  $c$ , otherwise 0),  $\hat{y}_{ic}$  is the predicted probability that sample  $i$  belongs to class  $c$ .

The number of training epochs  $T$  was set to 50, while a callback was also employed, in order to save to the model's checkpoint with the best accuracy score in the validation set, thus ensuring that after the training process is finished the final model will be the one with the best performance in the validation set. Moreover, it is noted that the initial and maximum learning rate  $\eta_{\text{init}}$ ,  $\eta_{\text{max}}$  respectively were set to 0.001,  $\eta_{\text{min}}$  was set to 0.0001, while warmup  $T_w$  was set to 10 epochs.

Post training, the model was able to achieve an accuracy score of **99.58%** in the train set, with the last epoch's loss being equal to **0.0157**, as seen in Figure 5.

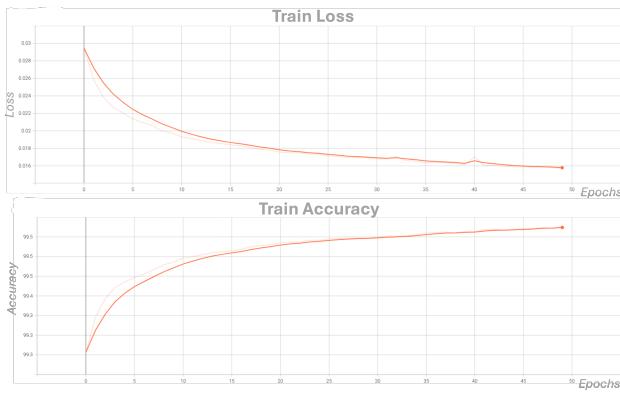


Fig. 5. Evolution of model's loss (top), and accuracy score (bottom) during training.

In Figure 6, one may notice that the model achieved peak performance in the validation set at epoch 35, with an accuracy score of **98.05%**.

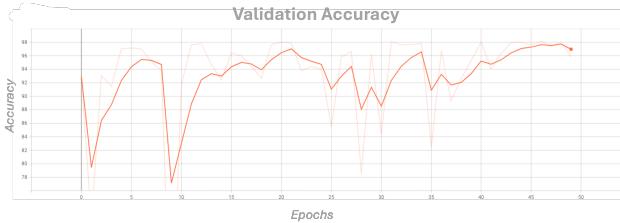


Fig. 6. The model's performance in the validation set, during training, with the best accuracy score achieved in epoch 35.

#### D. Post-Processing

The proposed model is capable of processing 330 lines of shape  $640 \times 224$  per second. In order to take advantage by the spatial domain and to acquire the final classification map, each of the predicted lines is accumulated in a 3D buffer until the whole set of lines post-processed together.

**Median filter:** a kernel size 5, is applied to the reconstructed classification map. In this manner, misclassifications that reassemble the *Salt & Pepper* noise are correct by

substituting them with the median value of the surrounding  $5 \times 5$  region, as described in Equation 4.

$$I'(x, y) = \text{median}\{I(i, j) \mid (i, j) \in W(x, y)\} , \quad (4)$$

where  $I(x, y)$  is the original image,  $I'(x, y)$  is the filtered image, and  $W(x, y)$  is the neighborhood window centered at  $(x, y)$ .

**Morphological opening and closing:** the application of such filters enhances the overall quality of the segmentation by removing small noise artifacts and refining object boundaries. The opening filter effectively eliminates small, isolated regions of misclassified pixels, while the closing filter fills in small gaps and smoothes the contours of classified regions, [47], [48]. Morphological filters presented respectively in Equations 5, 6

$$I \circ B = (I \ominus B) \oplus B , \quad (5)$$

where  $I$  is the original image,  $B$  is the structuring element,  $\ominus$  denotes the erosion operation,  $\oplus$  denotes the dilation operation.

$$I \bullet B = (I \oplus B) \ominus B , \quad (6)$$

where  $I$  is the original image,  $B$  is the structuring element,  $\oplus$  denotes the dilation operation,  $\ominus$  denotes the erosion operation.

Together, these morphological operations improve the structural integrity of the classification map, leading to more accurate and visually coherent segmentation results.

## VI. RESULTS & DISCUSSION

In this section the results of the proposed work on pixel-level material classification of hyperspectral images are presented, the structure of the sections is as follows a) overall performance, b) capacity to analyse randomly shredded objects, c) ability to distinguish mixed overlapping materials and d) limitations are discussed.

### Overall Performance

As discussed in Section IV the test-set presented in Table II includes images of all the HDPE, PET, PP, PS classes, which however have not been utilised in the training subset. In this manner, guarantee is provided that no same object or pixel is simultaneously evident in both train and test data.

The hyperspectral images of this dataset are used by P1CH classifier in order to predict the material class of each pixel. Figure 7 shows the original image as well as the ground truth and the prediction of each pixel.

To quantify the model's performance, the accuracy score was calculated for the whole test set in pixel level thus providing a detailed evaluation on the model's classification capacity. The overall accuracy achieved by the model is **97.44%**. The confusion matrix is presented in Figure 8, summarizing the classification performance among different material classes. The cell values in the confusion matrix are row-wise normalized, i.e. normalized with respect to the total

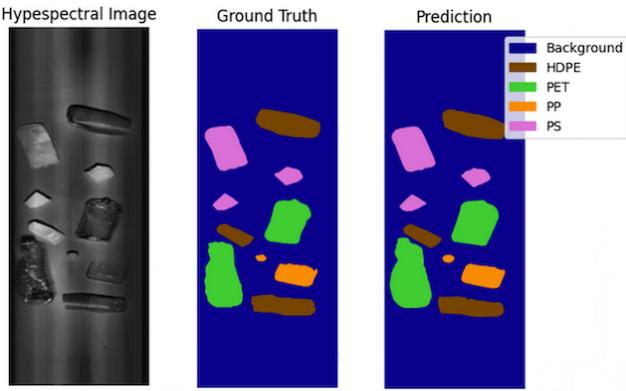


Fig. 7. The false-colour version of the HS images, the ground truth mask, as well as the generated classification map from the proposed model.

number of samples in each class. By observing the confusion matrix of Figure 8 is easy to notice that less than 1% of the total pixels of each class is falsely classified as another material and most of the errors are misclassifications between materials and the Background.

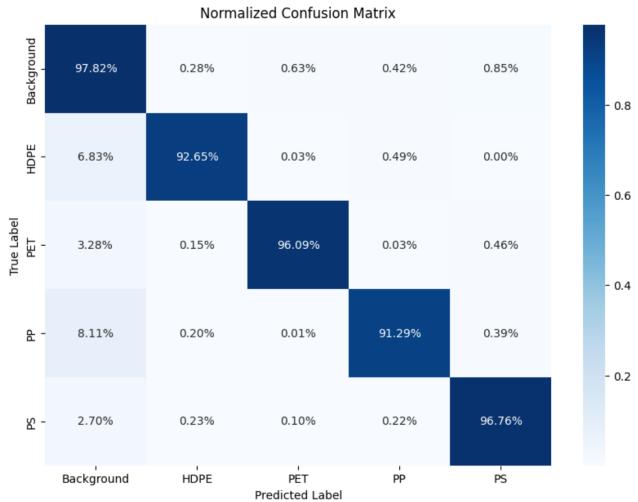


Fig. 8. Confusion Matrix describing models performance in classifying materials on pixel-level.

To further validate the performance of the proposed P1CH model, the state-of-the-art HybridSN methodology [35], [49] was replicated and evaluated on the same test dataset. This approach ensures a direct and fair comparison under identical conditions. The overall accuracy achieved by the HybridSN model on this dataset is significantly lower, at **21.81%**. In contrast, the P1CH model achieves an accuracy of **97.44%**, along with superior recall and Kappa score. These results, summarized in Table IV, underline the robustness and reliability of the proposed P1CH model in pixel-level material classification tasks.

The performance metrics summarized in Table IV further underscore the effectiveness of the proposed P1CH classifier. Notably, the accuracy achieved by P1CH (**97.44%**) is significantly higher than that of the HybridSN model (**21.81%**), highlighting its superior capacity to correctly classify pixels

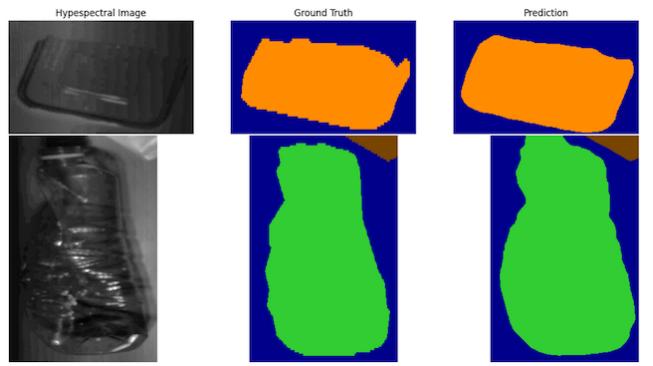


Fig. 9. Two zoomed crops of Figure 7. In the first row the PP sample originally located on the bottom right-side of Figure 7 is depicted with the GT and Predicted masks. In the second row the PET sample originally located on the left-side of Figure 7 is presented.

TABLE IV  
PERFORMANCE METRICS COMPARISON BETWEEN P1CH AND HYBRIDSN MODELS

Metric	P1CH	HybridSN
Mean Accuracy (%)	97.44	21.81
Mean Recall (%)	85.99	7.86
Mean Kappa Score	0.9295	-0.0795
Mean Inference Time (sec)	5.06	34.29

across the test set. In addition to accuracy, the recall value of P1CH (**85.99%**) demonstrates its robustness in identifying material classes with high consistency, in stark contrast to the HybridSN model, which achieved a recall of only **7.86%**. The Kappa score further illustrates the reliability of the P1CH model, with a value of **0.9295** compared to the negative score produced by HybridSN, reflecting its struggles with pixel-level classification tasks.

Beyond classification accuracy, the computational efficiency of P1CH is a notable advantage. The proposed model achieves an inference speed of **200 lines per second (5.06 seconds per image)**, far surpassing HybridSN's **29 lines per second (34.29 seconds per image)**. This substantial difference not only demonstrates the lightweight design of P1CH but also makes it a practical choice for almost real-time applications, where rapid and reliable pixel-level classification is essential. With these results, P1CH emerges as a robust, accurate, and highly efficient solution for hyperspectral image pixel-wise classification in demanding environments.

A visual comparison of predictions by both models, on the test set, is presented in Figure 10. The HybridSN model struggles with pixel-level classification accuracy, particularly at object boundaries, and often misclassifies a large portion of pixels. In contrast, the P1CH model demonstrates precise segmentation and classification, even in challenging scenarios involving overlapping objects or small irregular shapes. These findings validate the effectiveness of the P1CH model and its potential applicability in real-world hyperspectral imaging tasks.

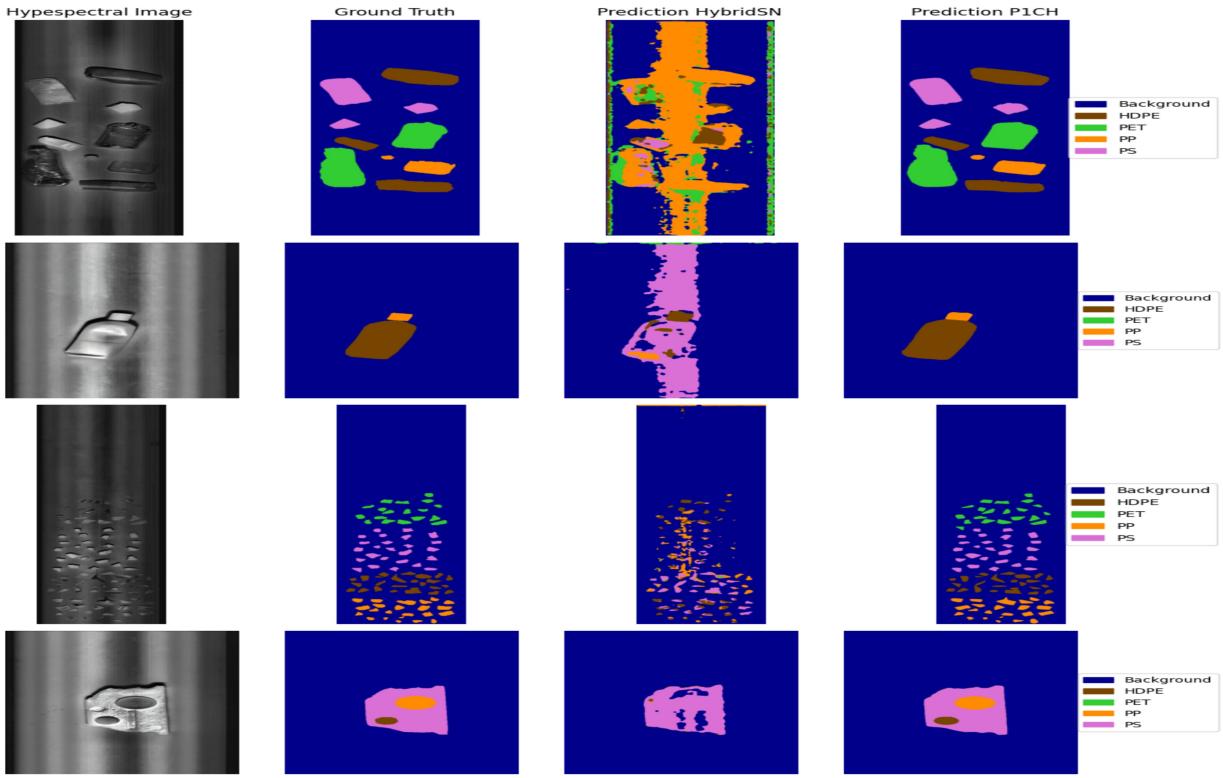


Fig. 10. From left to right: The false-colour HS image, the Ground Truth mask, the classification maps generated by HybridSN model ,as well as the generated classification map, by the P1CH Classifier, in the aforementioned scenarios.

### Focusing on the borders

Moreover, it can be calculated that 97.45% of the error comes in the borders between objects and the background, as suggested by the relatively low recall score. After more careful inspection is not easy to conclude that the error is actually on the predictions of P1CH and not on the ground truth masks. Figure 9 presents two zoomed-in crops of Figure 7, where it is clearly shown that in both cases the P1CH has the capacity to generate much smoother and precise masks for the respective object while simultaneously predicting correctly the material class. Comparing, also, the zoomed crops one can realize the mistake made on the top right side of the PP sample's mask, as well as at the PET bottle's spout. In both cases, P1CH was capable of predicting a more precise than the semi-automatically generated mask.

Therefore, given the aforementioned observations if the misclassifications on the borders are excluded from error calculations, the updated error rate is 0.0653% and hence the total pixel-wise accuracy of P1CH classifier is **99.94%**!

### Shredded materials

Beyond overall accuracy metrics, evaluating the model's effectiveness in classifying randomly shaped plastic samples is critical, as these shapes present significant challenges for RGB computer vision models and the same time is needed in various real-life applications. In this manner, an ensemble of plastic samples was collected and shredded in small, irregular pieces, and then placed on the conveyor belt. P1CH

classifier demonstrates remarkable capabilities in classifying on pixel-level the different material classes, clearly proving its superior performance compared to traditional RGB-based instance segmentation algorithms.

Traditional RGB models often struggle with such high variability due to their reliance on surface-level features like colour, texture and shape, which are limited in the case of small shredded samples. On the contrary, the proposed hyperspectral imaging approach is trained to purely exploit the rich spectral content of each pixel in the HS image, hence successfully tackling the challenge of small and irregularly shaped objects. Another key advantage of the proposed model is its resilience to noise and artifacts commonly found in RGB images. While traditional models can be easily misled by variations in lighting and surface texture, the spectral information utilised by the proposed model provides a more stable basis for classification.

In Figure 11 the HS and the respective RGB images of the shredded plastic samples of HDPE, PET, PP and PS are depicted. Along with the aforementioned images, the Ground Truth and the Predicted masks are presented in the same figure. The achieved accuracy in this specific image is 98.9%, with all the misclassifications falling under the borders case described in the previous section.

Visual comparison between the ground truth and the predicted mask indicates a high level of agreement in the classification of the plastic fragments. Each class—HDPE, PET, PP, and PS—is distinctly identified and accurately located in the predicted mask. In detail, even though the samples have

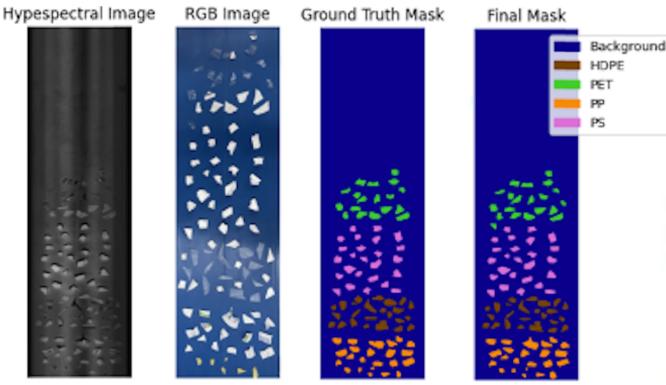


Fig. 11. From left to right: The false-colour HS image, the RGB equivalent image, the Ground Truth mask, as well as the generated classification map, by the P1CH Classifier, in a challenging, cluttered scene, where the objects are small with irregular shapes and similar textures.

similar colours regardless of their class, being either white or transparent, the model was able to correctly classify all the samples to their respective category, as indicated in the ground truth mask. This result underscores the robustness of P1CH is **invariant to colour and to size** of the objects by solely relying on the spectral content of each individual pixel.

Even though the points just presented are important, the most impressive observation from this experiment is the model's ability to detect objects that were not even visually identified during the labeling process of the HS images. Figure 12 specifies two regions in the HS image, denoted with red and green boxes respectively, in which there are two small shards of PET samples. Inspecting the zoomed-in crops presenting in Figure 12. It is almost impossible to detect those objects, yet the proposed HS approach clearly identified these two objects as PET. This outcome is only possible through the analytical processing of the spectral information of pixels in those regions, resulting in very precise masks and ultimately correctly classified samples.

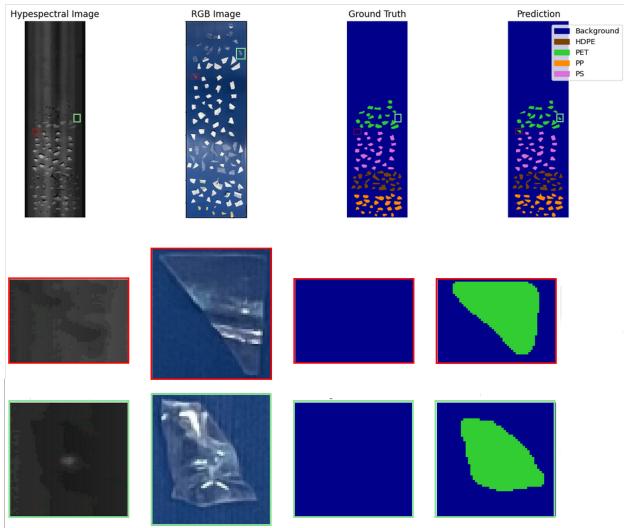


Fig. 12. Zoomed-in crops of Figure 11, highlighting two small PET shards that were mistakenly omitted from the labelling process (denoted with red and green boxes), and yet were detected by the model.

### Mixed overlapping materials

In addition to the challenge posed by small irregularly shaped objects, conventional RGB-based detection models often fail to accurately identify distinct samples that are either attached to one another or overlapping. In this subsection, an analysis is conducted to evaluate the model's capacity to precisely segment samples of different material class, while accurately classifying at the same time each individual pixel to its respective class. To this end, two experiments were executed, and in Figure 13 the HS and RGB images and the respective masks of the utilised objects are depicted. To further highlight the proposed model's ability to accurately classify materials and segment their instances within an image, a fine-tuned on the specific classes RGB-based instance segmentation model was also employed to segment the RGB images. The predictions of the RGB-based model are depicted in the 4th column of Figure 13.

In the first experiment, a common, commercially available, light blue shampoo bottle was selected. From a visual point of view, as seen in the RGB image, the body of the shampoo container and its lid appear almost identical making it impossible, even for a human, to realize that those two parts are different types of plastic, i.e. HDPE and PP. From the first row in Figure 13, one can easily observe that the RGB-based model was able to identify the contour of greater complex of objects, but not the individual parts, i.e. body and lid. Moreover, the RGB-based model failed to predict the class of both samples, assigning them to the PET class. Finally, due to shadows and non uniform illumination of the scene, the RGB generated mask lacks in precision as it includes pixels of the background too. On the contrary, the proposed HS approach, not only generated highly precise masks for both the body and the lid, but also is capable of classifying those two parts in their respective class. The overall accuracy of the prediction for this specific test case is 98.8%, with disagreements in a pixel's class being evident only between background and not the material classes.

In the second experiment, the case of overlapping materials with very similar (white) colour was examined. To this end, a white PS flat surface was selected, on top which a white PP (top right) and a white HDPE (bottom left) lid were placed. The respective masks and images are depicted in the second row of Figure 13. Looking at the RGB image, no distinctive textures can be detect for the two lids, while some minor changes in texture may be identified between the lids and the PS surface. Therefore, given the uniformity in colour and the very low variance in texture, as anticipated, the RGB-based model fail to detect all three objects apparent in the scene. Once again, the generated mask contains the greater complex of samples with no regard to the two lids. Moreover, the class prediction of the RGB-based is inaccurate since it considers the PS surface as PP. In contrast to the RGB case, the proposed HS model effectively utilises the rich spectral signature encoded in each pixel being able to precisely segment all three components of the image, while classifying materials on pixel-level with 99.54% accuracy.

The results presented above underscore a pivotal advance-

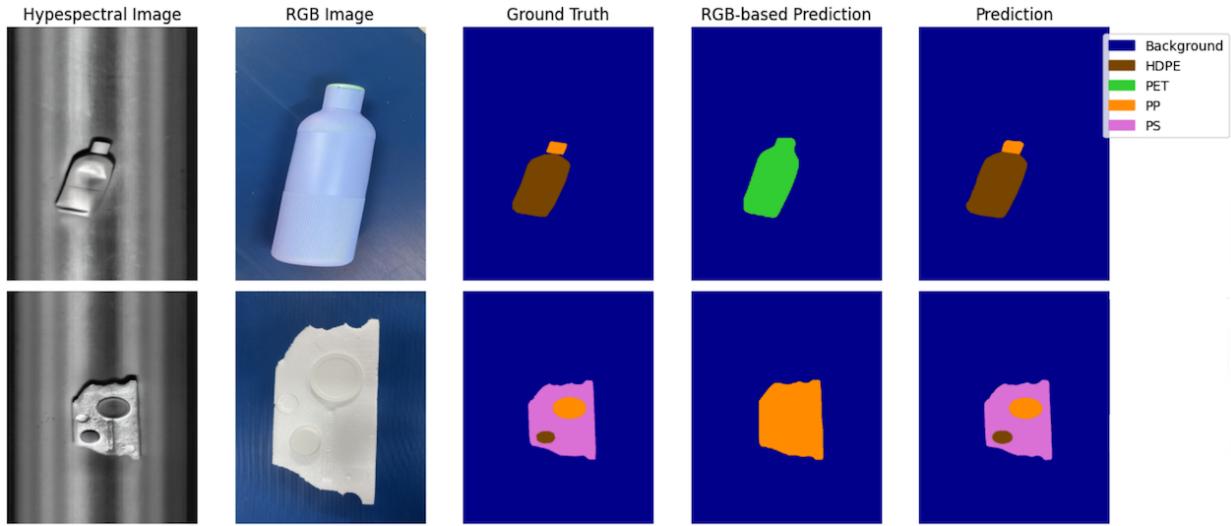


Fig. 13. From left to right: The false-colour HS image, the RGB equivalent image, the Ground Truth mask, the classification maps generated by an RGB-based model using the RGB equivalent images, as well as the generated classification map, by the P1CH Classifier, in the scenario of mixed or overlapping materials.

ment in material classification and segmentation capabilities. The most remarkable achievement demonstrated here is the proposed HS model's ability to accurately identify and classify distinct, overlapping samples with complex boundaries. This marks a significant breakthrough, as conventional RGB-based models consistently fail under these conditions, misidentifying materials and producing imprecise masks. The superior performance of P1CH classifier achieving up to **99.54%** accuracy even in the presence of overlapping objects, illustrates a transformative improvement in hyperspectral imaging applications, paving the way for more sophisticated and reliable material detection and sorting systems in real-world environments.

### Limitations

Despite the impressive results acquired in the previous experiments, this work intendeds to also underline the limitations of the HS imaging in classifying materials. To this end, the last experiment involves the analysis of dark-coloured, irregularly shaped samples. For the needs of this experiment, black and dark-coloured HDPE, PP and PS objects were cut in random small fractions and placed on top of the conveyor belt. The respective images and masks are presented in Figure 14. This subsection delves into the acquired results of the analysis of dark samples, discussing also the reasons why the model performs poorly in that case.

Examining the predictions, in the last column of Figure 14 it is concluded that the model performs very poorly in the case of black plastics. In detail, only PP samples were correctly classified, although lacking in mask precision. The PS samples were not completely undetected, while some HDPE samples were identified by the model but once again misclassified as PS or PP. The performance of the proposed model in terms of accuracy without taking into consideration the background is equal to merely **39.69%**.

The confusion matrix presented in Figure 15 specifically summarizes the models performance in the case of black

samples. In detail, it is confirmed that all the PS pixels were classified as background, i.e. they were not detected by the model at all. Moreover, from this confusion matrix one can see that indeed the PP samples were accurately classified, with the models precision in that case dropping to 75% since HDPE samples are also misclassifies as PP. Finally, for the HDPE case, only **4.01%** of the total samples were correctly classified, with the model's recall for the specific case being equal to **46%**.

At this point, it of paramount importance to explain the reasons that lead in this drop in the model's performance. Specifically, this phenomenon should mainly be attributed to the nature of dark-coloured materials, rather than being considered a model's deficiency. In detail, the observed dark or even black colour of an object is the result of complete or almost complete absorption of the incident radiation. In this manner, the reflected radiation, which is captured by the HS cameras' sensor, is of very low intensity; hence resulting in a very weak digital signal. Therefore, given the low amplitude of the captured signal, the PSNR is consequently also low, and so is the variability of the spectrum. All the above, result in very similar -almost identical-, noisy features of the model's input vector, thus rendering the model incapable of properly analyzing each pixel's spectrum and ultimately correctly classifying them in their respective classes. An example of the spectrum of a black and a white plastic of the same material are presented in Figure 16. Can be easily observed that the spectrum of the black-colored object is much more noisy with 10 times lower intensity.

Although, it should be highlighted that black plastics of any material class were totally absent from the training set. The exploration of techniques that would probably lead to mitigate the pure performance of computer vision with HS imaging on black objects is beyond the scope of this paper, but opens and intriguing research direction on the intersection between material science and artificial intelligence.

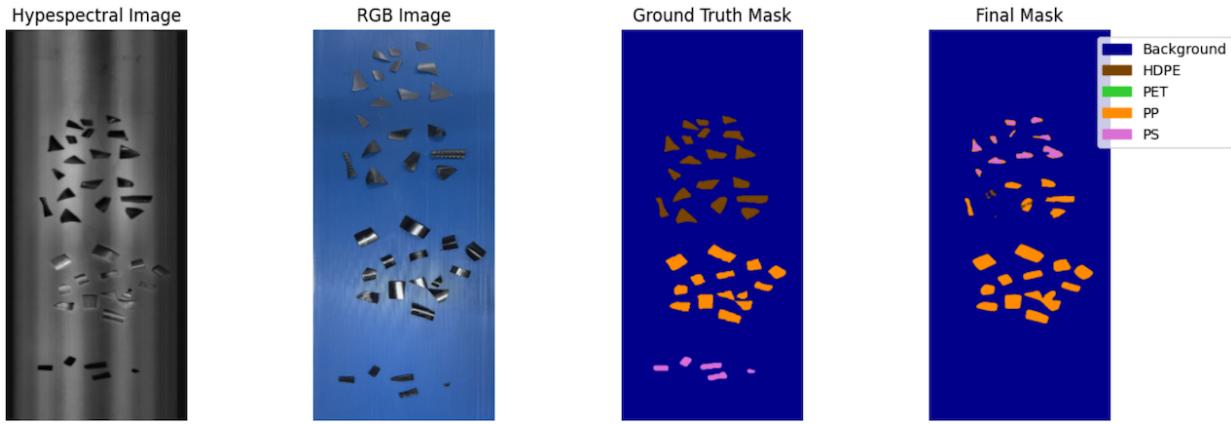


Fig. 14. From left to right: The false-colour HS image, the RGB equivalent image, the Ground Truth mask, as well as the generated classification map, by the P1CH Classifier, in the case of black plastics.

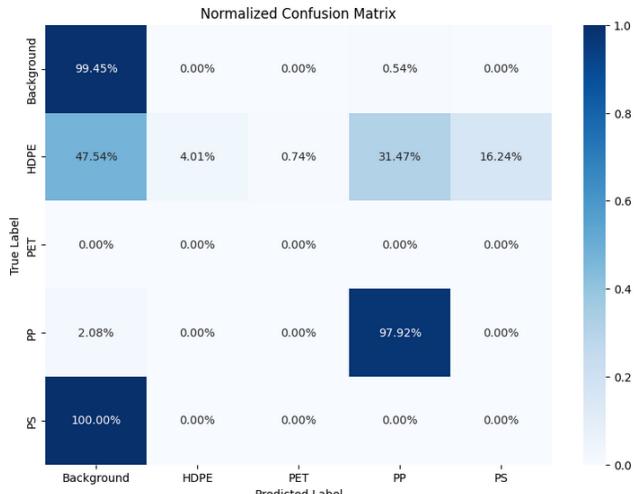


Fig. 15. Normalized Confusion Matrix for the predictions of the model in the case of black and dark-coloured samples.

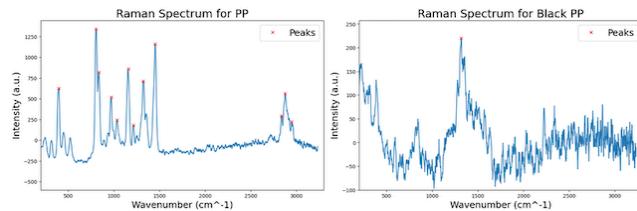


Fig. 16. A comparison between the spectra acquired from a black (left) and a white (right) PS sample. A significant drop in the signal's amplitude can be observed in the case of black PS. Also one can notice that most of the indicative peaks for PS are non-existent in the case of black PS.

## VII. CONCLUSION & FUTURE WORK

In this paper a lightweight *1D Convolutional Hyperspectral Classification*, system, P1CH, is proposed. The developed model generates highly detailed and precise classification maps. Unlike conventional RGB methods, the proposed algorithm utilises the spectral information encoded in hyperspectral image's pixel, allowing to detect and correctly classify objects in challenging scenarios.

For the training and validation purposes two sets of hyperspectral images were generated. Both of the dataset splits contain images of *HDPE*, *PET*, *PP* and *PS* samples. Each of the images in the training set, contain explicitly objects of one class, while the images in the test set aim to reproduce challenging conditions, hence depicting very small irregularly shaped objects, overlapping objects, and mixed materials with similar to identical texture. A simplified, cost-efficient spectral calibration and normalization technique is, also, proposed in this work that do not requires specialized hardware. The model was validated on the aforementioned test set and it achieved an overall accuracy of **99.54%**. Moreover, to explore the limitations of the P1CH Classifier the scenario of classifying black samples was explored, where it was noticed that the model struggled to properly classify the samples.

On the one hand, this work presents the capabilities of computer vision with Hyperspectral imaging. On the other hand raised limitations and problems that are absolutely worth systematic research effort. Further investigation on the problem of black or dark-coloured samples. Study the capabilities on more extensive spectral range. Moreover, the model's robustness in various illumination conditions is set to be further examined. Furthermore, dimensionality reduction and compression problems have to be examined, as well as the effects such operations might have on the learning performance. Finally, the utilisation of the spatial information encoded in hyperspectral images, shall also be examined in future studies.

## REFERENCES

- [1] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *Ieee Access*, vol. 6, pp. 14 118–14 129, 2018.
- [2] B. Lu, P. D. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sensing*, vol. 12, no. 16, p. 2659, 2020.
- [3] M. Díaz, R. Guerra, P. Horstrand, E. Martel, S. López, J. F. López, and R. Sarmiento, "Real-time hyperspectral image compression onto embedded gpus," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 8, pp. 2792–2809, 2019.
- [4] N. Sharma, M. S. Waseem, S. Mirzaei, and M. Hefeeda, "Mobspectral: Hyperspectral imaging on mobile devices," in *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, 2023, pp. 1–15.

- [5] O. Lim, S. Mancini, and M. Dalla Mura, "Feasibility of a real-time embedded hyperspectral compressive sensing imaging system," *Sensors*, vol. 22, no. 24, p. 9793, 2022.
- [6] B. Przestrzelski, E. Reddy, and S. Lord, "Integrating experiential with technical: How materials science modules can help redefine the traditional engineering canon," in *ASEE Annual Conference*, 2018.
- [7] K. Jang, T. Heo, and S. Jeong, "Classification option for korean traditional paper based on type of raw materials, using near-infrared spectroscopy and multivariate statistical methods," *BioResources*, 2020.
- [8] D. P. Penumuru, S. Muthuswamy, and P. Karumbu, "Identification and classification of materials using machine vision and machine learning in the context of industry 4.0," *Journal of Intelligent Manufacturing*, vol. 31, pp. 1229–1241, 2020.
- [9] A. Chowdhury, E. Kautz, B. Yener, and D. Lewis, "Image driven machine learning methods for microstructure recognition," *Computational Materials Science*, vol. 123, pp. 176–187, 2016.
- [10] L. Zhang and Z. Li, "Machine learning for materials classifications from images," in *Journal of Physics: Conference Series*, vol. 2369, 2022, p. 012081.
- [11] L. Zhang and S. Shao, "Image-based machine learning for materials science," *Journal of Applied Physics*, 2022.
- [12] J. Weiß and A. Santra, "Material classification using 60-ghz radar and deep convolutional neural network," *International Radar Conference*, 2019. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9079136/>
- [13] H. Zheng, L. Fang, M. Ji, M. Strese, and Y. Özer, "Deep learning for surface material classification using haptic and visual information," *IEEE Transactions on Automation Science and Engineering*, 2016. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7530831/>
- [14] P. Zoumpoulis, F. K. Konstantinidis, G. Tsimiklis, and A. Amditis, "Smart bins for enhanced resource recovery and sustainable urban waste practices in smart cities: A systematic literature review," *Cities*, vol. 152, p. 105150, 2024.
- [15] ———, "Advancing urban waste management using industry 5.0 principles: A novel smart bin," in *2024 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4. 0 & IoT)*. IEEE, 2024, pp. i–vi.
- [16] O. Adedeji and Z. Wang, "Intelligent waste classification system using deep learning convolutional neural network," *Procedia Manufacturing*, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2351978919307231>
- [17] S. Azimi, D. Britz, M. Engstler, M. Fritz, and F. Mücklich, "Advanced steel microstructural classification by deep learning methods," *Scientific Reports*, 2018. [Online]. Available: <https://www.nature.com/articles/s41598-018-20037-5>
- [18] F. K. Konstantinidis, S. Sifnaios, G. Arvanitakis, G. Tsimiklis, S. G. Mouroutsos, A. Amditis, and A. Gasteratos, "Multi-modal sorting in plastic and wood waste streams," *Resources, Conservation and Recycling*, vol. 199, p. 107244, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0921344923003786>
- [19] M. Shaikh and B. Thörnberg, "Impact of water vapour on polymer classification using in situ short-wave infrared hyperspectral imaging," *Journal of Spectral Imaging*, 2022. [Online]. Available: [https://www.impopen.com/download.php?code=I11\\_a5](https://www.impopen.com/download.php?code=I11_a5)
- [20] A. Shaban, "Determination of concrete properties using hyperspectral imaging technology: A review," *Science Journal of Physics*, 2013. [Online]. Available: <https://www.sjpub.org/sjcea/sjcea-102.pdf>
- [21] G. Capobianco, F. Prestile, and S. Serranti, "Hyperspectral imaging-based approach for the in-situ characterization of ancient roman wall paintings," in *International Conference on Image Processing Theory, Tools and Applications*, 2015.
- [22] A. Polak, T. Kelman, P. Murray, and S. Marshall, "Hyperspectral imaging combined with data classification techniques as an aid for artwork authentication," *Journal of Cultural Heritage*, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1296207417301218>
- [23] J. Banumathi and A. Muthumari, "An intelligent deep learning based inception model for hyperspectral image analysis and classification," *Computational Materials*, 2021.
- [24] S. Li, W. Song, L. Fang, and Y. Chen, "Deep learning for hyperspectral image classification: An overview," *IEEE Transactions on Geoscience and Remote Sensing*, 2019.
- [25] L. Windrim, A. Melkumyan, and R. Murphy, "Pretraining for hyperspectral convolutional neural network classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2018.
- [26] Y. Chen, K. Zhu, L. Zhu, X. He, and P. Ghamisi, "Automatic design of convolutional neural network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2019. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8703410/>
- [27] B. Pan, Z. Shi, and X. Xu, "R-vcanet: A new deep-learning-based hyperspectral image classification method," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017. [Online]. Available: <https://levir.buaa.edu.cn/publications/RVACNet-R2.pdf>
- [28] R. Venkatesan and S. Prabu, "Hyperspectral image features classification using deep learning recurrent neural networks," *Journal of Medical Systems*, 2019. [Online]. Available: <https://link.springer.com/article/10.1007/s10916-019-1347-9>
- [29] F. Xiong, J. Zhou, and Y. Qian, "Material based object tracking in hyperspectral videos," *IEEE Transactions on Image Processing*, 2020. [Online]. Available: <https://arxiv.org/pdf/1812.04179.pdf>
- [30] L. Medus, M. Saban, and J. Francés-Villora, "Hyperspectral image classification using cnn: Application to industrial food packaging," *Food Control*, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0956713521001006>
- [31] N. Okada, Y. Maekawa, N. Owada, K. Haga, A. Shibayama, and Y. Kawamura, "Automated identification of mineral types and grain size using hyperspectral imaging and deep learning for mineral processing," *Minerals*, vol. 10, no. 9, 2020. [Online]. Available: <https://www.mdpi.com/2075-163X/10/9/809>
- [32] C. Ni, Z. Li, X. Zhang, X. Sun, Y. Huang, L. Zhao, T. Zhu, and D. Wang, "Online sorting of the film on cotton based on deep learning and hyperspectral imaging," *IEEE Access*, vol. 8, pp. 93 028–93 038, 2020.
- [33] S. Zhu, L. Zhou, P. Gao, Y. Bao, Y. He, and L. Feng, "Near-infrared hyperspectral imaging combined with deep learning to identify cotton seed varieties," *Molecules*, vol. 24, no. 18, 2019. [Online]. Available: <https://www.mdpi.com/1420-3049/24/18/3268>
- [34] A. Picon, P. Galan, A. Bereciartua-Perez, and L. B. del Valle, "Hyperspectral dataset and deep learning methods for waste from electric and electronic equipment identification (weee)," 2024. [Online]. Available: <https://arxiv.org/abs/2407.04505>
- [35] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "Hybridsn: Exploring 3d-2d cnn feature hierarchy for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, 2020.
- [36] D. J. Brady, *Optical imaging and spectroscopy*. John Wiley & Sons, 2009.
- [37] S. Sifnaios, I. Zorzos, G. Arvanitakis, F. K. Konstantinidis, G. Tsimiklis, and A. Amditis, "Exploration and mitigation of the impact of lighting conditions on multi-spectral image classification," in *2023 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2023, pp. 1–6.
- [38] O. Tamin, E. Moung, J. Dargham, F. Yahya, and S. Omatu, "A review of hyperspectral imaging-based plastic waste detection state-of-the-arts," *International Journal of Electrical and Computer Engineering*, vol. 13, pp. 3407–3419, 06 2023.
- [39] M. Lindell, "Pixel classification of hyperspectral images," Ph.D. dissertation, Uppsala University, 2018. [Online]. Available: <https://uu.diva-portal.org/smash/get/diva2:1216370/FULLTEXT01.pdf>
- [40] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.
- [41] E. Smith and G. Dent, *Modern Raman Spectroscopy: A Practical Approach*. John Wiley & Sons, 2019.
- [42] D. A. Long, "The raman effect: a unified treatment of the theory of raman scattering by molecules," *John Wiley & Sons*, 2002.
- [43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [44] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.
- [45] ———, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
- [46] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [47] J. Serra, "Image analysis and mathematical morphology," *Academic Press, Inc.*, 1982.
- [48] F. De la Torre and J. Serra, "Morphological filtering," *SPIE Proceedings Vol. 0534, Image Processing*, 1984.
- [49] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, "Attention-based adaptive spectral–spatial kernel resnet for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7831–7843, 2021.