Tinto: Multisensor Benchmark for 3-D Hyperspectral Point Cloud Segmentation in the Geosciences

Ahmed J. Afifi, Samuel T. Thiele, Aldino Rizaldy, Sandra Lorenz, Pedram Ghamisi, *Senior Member, IEEE*, Raimon Tolosana-Delgado, Moritz Kirsch, Richard Gloaguen, and Michael Heizmann

Abstract— The increasing use of deep learning techniques has reduced interpretation time and, ideally, reduced interpreter bias by automatically deriving geological maps from digital outcrop models. However, accurate validation of these automated mapping approaches is a significant challenge due to the subjective nature of geological mapping and the difficulty in collecting quantitative validation data. Additionally, many state-of-the-art deep learning methods are limited to 2-D image data, which is insufficient for 3-D digital outcrops, such as hyperclouds. To address these challenges, we present Tinto, a multisensor benchmark digital outcrop dataset designed to facilitate the development and validation of deep learning approaches for geological mapping, especially for nonstructured 3-D data like point clouds. Tinto comprises two complementary sets: 1) a real digital outcrop model from Corta Atalaya (Spain), with spectral attributes and ground-truth data and 2) a synthetic twin that uses latent features in the original datasets to reconstruct realistic spectral data (including sensor noise and processing artifacts) from the ground truth. The point cloud is dense and contains 32 42 964 labeled points. We used these datasets to explore the abilities of different deep learning approaches for automated geological mapping. By making Tinto publicly available, we hope to foster the development and adaptation of new deep learning tools for 3-D applications in Earth sciences. The dataset can be accessed through this link: https://doi.org/10.14278/rodare.2256.

Index Terms— Deep learning, digital outcrop, hypercloud, hyperspectral, point cloud, point cloud segmentation, remote sensing, synthetic data.

I. INTRODUCTION

THE need for annotated datasets to train and assess deep learning models has become essential in numerous advanced fields of research, including remote sensing and Earth observation [1]. Although there are several (ongoing) attempts to promote benchmarking and open science in the remote sensing field by developing exhaustive lists of available datasets [1],[1] evaluation servers (e.g., DASE[2]), cloud services (e.g., Amazon Web Services,[3] Microsoft's Planetary Computer,[4] and Radiant Earth's MLHub[5]), and benchmark datasets (e.g., [2], [3], [4], [5]), novel applications within the fields of geomorphology, and geology remain sparse and unclear. Also, generally applicable reference data in remote sensing and geosciences for evaluating machine learning approaches are still not available in sufficient quantity and quality [6], [7].

Hyperspectral remote sensing has emerged as a powerful tool for detecting subtle spectral differences in mineralogical composition. The ability to apply this technology from a range of platforms, including satellites, airplanes, autonomous vehicles, and tripods, supports geological mapping at various scales. Integrating surface spectral data with topographic data enables the creation of hyperclouds, which are geometrically and radiometrically accurate point cloud representations of the target. Depending on the analyzed range of the electromagnetic spectrum, a variety of minerals can be detected and mapped in their original 3-D context. The visible and near-infrared (VNIR) and shortwave infrared range (SWIR) are useful for detecting spectral features of alteration minerals, such as oxides and hydroxides. In contrast, the long-wave infrared region (LWIR) allows for the detection of many rock-forming minerals, such as quartz and feldspars. Integrating both information sources enables comprehensive lithological mapping and has numerous geological applications, including greenfield exploration of critical raw materials, geological mapping of open-pit and underground mines, compositional mapping of rock wastes and stockpiles, and soil contamination mapping in postmining landscapes.

Despite the vast potential of deep learning for hyperspectral mapping [8], it has found limited applicability for geoscientific applications, primarily due to the challenges associated with validation and the benchmarking of adapted algorithms [6]. Ground truth is difficult to establish for large-scale datasets,

[1]https://github.com/satellite-image-deep-learning/datasets
[2]http://dase.grss-ieee.org/
[3]https://registry.opendata.aws/
[4]https://planetarycomputer.microsoft.com/catalog
[5]https://www.mlhub.earth/

due to the limited accessibility of geological outcrops, their inherently complex and heterogeneous mineralogical composition as well as the traditionally rather subjective definition of lithological domains by geological experts based on mostly visual criteria. While simulated datasets have been used in the past, they do not yet capture the complexity and variability of real-world geological environments. The rather novel concept of 3-D digital spectral outcrops (e.g., hyperclouds) is causing further complications due to the unstructured and complex nature of spectral point cloud data, which is incompatible with most state-of-the-art algorithms. Current approaches partially solve these issues by simplifying the dataset either spatially (working in the 2-D image space [9]) or spectrally (segmentation based on selected features [10]) before applying segmentation in 2-D or 3-D, respectively. For both ways, the potential of the dataset is not fully exploited and important information might be lost in the process. Geological scenes are preferred over highly structured scenes for benchmarking deep learning models in geosciences due to their complexity, variability, and real-world relevance. Geological scenes exhibit various features like rock formations, mineralogy, and topography, creating a realistic and challenging testing environment for deep learning algorithms. Class boundaries are diffused in the geological scenes and the classes are highly mixed. Benchmarking geological scenes enhance the models' generalization capability and learning new patterns and features from the training data, allowing them to adapt to new unseen geological scenarios and making them more robust and reliable in real-world applications. Additionally, it enables researchers to understand deep learning models' performance in capturing geological information and aligns with the needs of geoscientific applications. Overall, benchmarking on geological scenes provides a robust evaluation environment and facilitates tailored deep learning solutions in geosciences. Lately, few 3-D hyperspectral datasets with geoscientific application context have been openly published (see [7], [11]). Nevertheless, none of them offers an adequate amount of ground truth data to qualify as a benchmark dataset.

In this contribution, we present a large and geologically complex but well-understood real-world benchmark dataset, and a synthetic (reconstructed) equivalent, designed for testing and comparing deep learning methods for hyperspectral geological mapping. The real-world dataset covers Corta Atalaya, an abandoned open pit mine within the Minas de Rio Tinto copper mining district in Andalusia, Spain. The hyperspectral data have been acquired using plane, drone, and tripod-based acquisition and cover the VNIR, SWIR, and LWIR range of the electromagnetic spectrum. Lithology class labels have been defined for the whole dataset based on a combination of detailed laboratory analysis and derived supervised classification [12] and were adjusted based on an expert interpretation of the geology. However, due to the complex nature of geological datasets, this labeling cannot be treated with 100% confidence. To address this shortcoming, we used the real-world benchmark dataset to derive a realistic synthetic dataset in which class labels (and associated spectral endmembers and abundances) are known with certainty. This approach allows us to develop data for which the class and abundance properties are known with confidence while retaining the spatial statistical properties and complexity of a real dataset.

To facilitate established and emerging deep learning approaches, we present these datasets both in 2-D raster form (as is conventional for remote sensing applications) and 3-D point cloud form (for emerging approaches that are beginning to move beyond the topological limitations imposed by 2-D rasters). Challenges and limitations associated with each data representation, and a selection of tools available for working with them, are discussed.

The rest of the paper is structured as follows. Section II reviews some related work and the available datasets for 3-D point cloud processing. Section III describes the proposed Tinto benchmark dataset and how it is collected, labeled, and synthesized in detail. In Section IV, we discuss the baseline deep learning models that are used to evaluate the Tinto dataset, the experimental setup, and the experimental outcomes. Finally, the conclusion and remarks are drawn in Section V.

II. BACKGROUND AND RELATED WORK

A plethora of sophisticated methodologies has been advanced to segment 3-D point clouds, often utilizing publicly available datasets to evaluate their efficacy. These methodologies encompass a broad spectrum of techniques, from traditional methods to cutting-edge machine learning and deep learning approaches [13]. In Section II-A and II-B, we provide a very brief overview of the mature field of some diverse point cloud segmentation approaches by highlighting a few relevant examples and presenting a comprehensive collection of 3-D datasets commonly employed for this specific task.

A. Point Cloud Segmentation

Traditional point cloud segmentation methods rely on strict hand-crafted geometric constraints and rules. The main goal of the segmentation process is to group 3-D points into nonoverlapping regions. The generated regions have common semantic meanings and geometric structures [14]. With the introduction of machine learning and deep learning models in solving 2-D tasks and the availability of large-scale labeled datasets, many researchers proposed machine/deep learning models to segment point clouds from the object and scene levels. Generally, deep learning models achieved remarkable performance compared to traditional and machine learning methods. Following, we discuss different point cloud segmentation methods.

1) Edge Based: Edge-based methods try to detect points close to the edge by calculating the rapid changes in the intensity (the feature associated with the points), normals, or the gradient. This will create boundaries between two different regions. Then, the points are grouped inside the same region where changes are small. These methods perform segmentation quickly but struggle to achieve accurate results when dealing with point clouds from large areas due to issues like noise and uneven point distribution [15], [16].

2) Region Growing: Region growing-based methods involve the random selection of seed points and the measure of geometrical or feature similarity between the seeds and neighboring points. Points with similar features are merged to create one region. This process is performed iteratively until all points are merged into similar regions. These methods were first applied to 2.5-D LiDAR data, and they were widely applied for the segmentation of building structures. Similar points that belong to the same region can be selected by comparing their features or calculating the Euclidean similarity. The segmented points are selected for example by fitting a plane to a number of points in a given volume and then points with the minimum distance to that plane are merged [17], [18], [19].

3) Shallow Supervised Machine Learning: Shallow supervised machine learning refers to nondeep algorithms that use labeled data to train a model. These methods allow classifying points in a cloud based on predefined features such as maximum likelihood based on support vector machine (SVM) [20], random forests (RFs) [21], and Bayesian discriminant classifiers [22]. Other groups of methods depend on statistical contextual models such as conditional random fields (CRFs) [23] and Markov random fields (MRFs) [24]. These methods focus on the statistics and the relational information of the points over different scales. Machine learning models applied for point cloud segmentation perform a neighborhood point selection, then feature extraction from the grouped points, feature selection to reduce the feature dimensionality and then segment the points semantically.

4) Deep Learning: Deep learning has become the most influential and hottest technique in different research fields such as computer vision, medical imaging, autonomous driving, and robotics. Deep learning is a special branch of machine learning where the models are deeper; more complex and the extracted features generally have higher dimensions than the ones extracted from traditional machine learning methods. Applicable methods for applying deep learning on 3-D data depend on how the data are represented. With multiview data, a normal 2-D convolutional neural network (CNN) can be easily applied, such as the MVCNN model [25]. Voxel-based data can be used with 3-D CNN, where the normal 2-D CNN can be easily extended to 3-D [26]. The drawback of using voxel-based representation is the memory and the computation cost to train a model. To overcome the voxel-based and multiview methods, models that can be applied directly on point cloud data were recently proposed as a promising solution. Models applied directly on point clouds such as the pioneer model PointNet [27] were followed by the improved version PointNet++ [28] and the dynamic graph CNN (DGCNN) [29]. Other approaches of point cloud segmentation in the field of remote sensing and 3-D laser scanning can be found in [30], [31], and [32].

The research on point cloud segmentation using deep learning is a hot research topic. Different models are proposed with either sophisticated layers [33], [34], [35] to deal with the point cloud as an unordered set or with a simple architecture using MLP as a backbone of the model [36] to achieve improved performance with less computation and memory cost.

B. Available 3-D Datasets

To our knowledge, no 3-D hyperspectral benchmark datasets have been published for a geoscientific application context yet. Available benchmark datasets for point cloud segmentation deal with indoor scenes, such as Stanford large-scale 3-D indoor spaces (S3DIS) [37] and Semantic3D.Net [38], or urban scenes, such as Sydney Urban Objects Dataset [39], Toronto-3-D [40], and SemanticKITTI [41], for semantic segmentation. Other datasets focus on the objects, for instance segmentation or object parts segmentation, such as ShapeNet [4] and ModelNet40-C [42]. In most cases, point cloud attributes are limited to the 3-D coordinates (X, Y, Z), intensity, or RGB color values. The Maarmorilik dataset [7] is an open-source 3-D hyperspectral dataset capturing the complex geology of the Black Angel Mountain in Maarmorilik, West Greenland, alongside a detailed and interactive tutorial documenting relevant processing workflows for hypercloud data. It includes RGB and VNIR-SWIR hyperspectral data but does not provide ground truth and thus cannot be defined as a benchmark.

In [43] and [44], similar work has been conducted by generating hyperspectral point clouds, but the datasets are not available to the public. The MUUFL Gulfport dataset [45] and GRSS18 dataset [46], on the other hand, provide the ground truth and are publicly open. They consist of LiDAR and hyperspectral data. However, the ground truth is in a 2-D format rather than in a 3-D point format as offered by our dataset. Some works [10], [47], [48], [49] have also been done investigating the utilization of hyperspectral data with LiDAR for classification task but neglecting the use of deep learning. Moreover, these works [47], [48], [49] performed the classification in a 2-D style. The work in [50] investigated the use of smartphone and UAV photogrammetry to assess rock slope hazards in mountainous regions. Different datasets were created according to lighting conditions, slope morphologies, and seasons. They were manually labeled and were classified using RF into geologically relevant categories. The research compares 12 different point cloud feature sets, finding that feature sets focused on geometry, slope, and texture perform significantly better than those incorporating absolute color features, which are sensitive to lighting changes and struggle to distinguish between geological materials. More recent works [51], [52], [53] use deep learning models for hyperspectral point cloud segmentation and proved that hyperspectral data improved the performance of the models. These results motivated us to employ various deep learning models in our work.

To the best of the authors' knowledge, the Tinto dataset will be the first-ever dataset that provides the following features:) a 3-D point cloud of a real outcrop; 2) the corresponding ground truth; 3) the same scene captured using different sensors (RGB, VNIR, SWIR, and LWIR); 4) hyperspectral information attached to each point in the point cloud; 5) two types of corresponding synthetic data (clean and noisy data); and 6) 2-D views of the scene from three different directions.

III. TINTO DATASET

A. Data Acquisition and Correction

Several steps of the acquisition and processing of the Corta Atalaya hyperclouds have been previously described

by Thiele et al. [12] and Kirsch et al. [54]. To summarize briefly, a tripod-mounted hyperspectral Specim AisaFenix camera was used to capture oblique VNIR and SWIR imagery from three locations on the edge of the Corta Atalaya open-pit. Each of these rasters was then back-projected onto a dense 3-D point cloud derived from 488 RGB photographs (captured using a Nikon D850 DSLR camera and Nikkor 85-mm f/1.8G lens) using the structure from motion multiview stereo method implemented in Agisoft Metashape Professional v1.6. Atmospheric effects, which result largely from: 1) the spectral signature of sunlight; 2) interactions between this light and the atmosphere; and 3) uneven illumination across the complex surface of the Corta Atalaya mine, were corrected using the method described by Thiele et al. [55] during the back-projection step.

The long-wave infrared hyperspectral data were collected in August 2020 during a larger hyperspectral airborne mapping campaign [54]. A Hyper-Cam FTIR hyperspectral camera from Telops was deployed, covering the electromagnetic spectrum between 7.7 and 11.8 mm. The collected raw data were processed using a standard workflow: The individual data cubes were orthorectified using a 2.5-m resolution Lidar-based terrain model, acquired within the same campaign, and subsequently stitched to a mosaic (average ground sampling distance of 1.2 m) using Telops' Reveal Airborne geolocation tool (version 2). Initial radiometric correction to at-sensor radiance was performed using the Telops Reveal Calibrate Software Version 5.2.8. Atmospheric correction was done by an In-Scene Atmospheric Compensation algorithm, while the separation of temperature and emissivity was performed based on emissivity normalization from the radiance data [54]. The resulting calibrated emissivity mosaic was then sampled onto the same dense 3-D point cloud that was used with the VNIR-SWIR data to create an LWIR hypercloud to be included in this benchmark dataset. Fig. 1 visualizes an overview of the Tinto benchmark dataset with the various datasets it contains [56]. The 3-D visualization of the Tinto point clouds on Potree [57] can be accessed through this link: https://www.hzdr.de/FWG/FWGE/Hyperclouds/Tinto.html

B. Data Labeling and Synthetic Twin

Geological maps, i.e., classifications that show the spatial distribution of rock types, are generally subjective interpretations of the map author due to: 1) cover by vegetation or soil; 2) ambiguous rock type definitions; and 3) the thematic or purpose the author has in mind when creating the map. For example, a map intended to constrain geotechnical aspects of a mine would often differ significantly from one made to quantify the composition and distribution of ore, due to subjective choices made when defining and classifying different rock units. While this ambiguity is an important justification for automated methods offered by, e.g., machine learning approaches, which can improve objectivity while simultaneously allowing for data to be reprocessed for various purposes, it presents significant challenges when developing meaningful and reliable benchmarks. In this contribution, we have mitigated these challenges using two radically different approaches: 1) deriving a manually vetted but largely
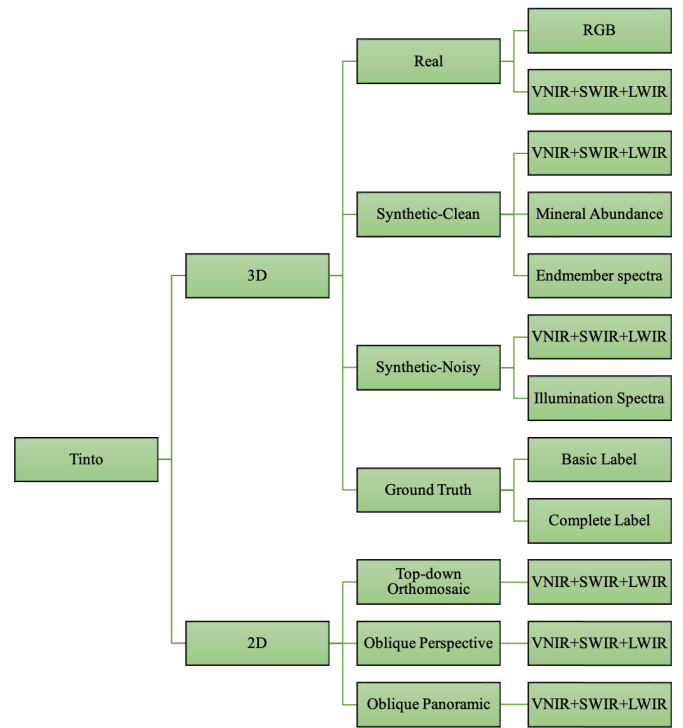

Fig. 1. Overview of the Tinto benchmark structure and the various datasets it contains.

data-driven classification for ground-truthing purposes and 2) back-calculating a realistic dataset (synthetic twin) from this classification result to derive a hyperspectral dataset for which the original rock composition is known for each pixel. These two approaches are described in the following sections.

1) Geological Ground Truth: In the first approach, which aims at a meaningful ground-truth classification for the real hyperspectral data, we have integrated and synthesized hyperspectral information, sample mineralogy, field mapping, and published geological understanding of Corta Atalaya. The spectral classification results of [12] were used as a base for the ground truth and manually corrected where field data, ground sampling and expert interpretation of the high-resolution photogrammetric model showed clear mislabeling. These results (Fig. 2) were subsequently checked by mine geologists on-site, resulting in a labeled dataset that we consider to be as accurate as practically possible for geological applications.

Several classes in this classification are spectrally and geologically related (e.g., classes defined by the presence of different but related alteration minerals [12]). Lumping these together, we derive a simplified classification containing six rock types. While we encourage people to use the full label set, this simplified version could be useful for evaluating approaches that perform poorly with a large number of classes (e.g., unsupervised methods).

We have also defined a suggested training subset (Fig. 2) to ensure consistent results between studies. This has been selected such that: 1) it covers all classes in the dataset and 2) matches with what could be realistically achieved in practice, with training data distributed along three bench-traverses that are typical for geological mapping in open-pit environments. Note that this geometry results in a highly imbalanced
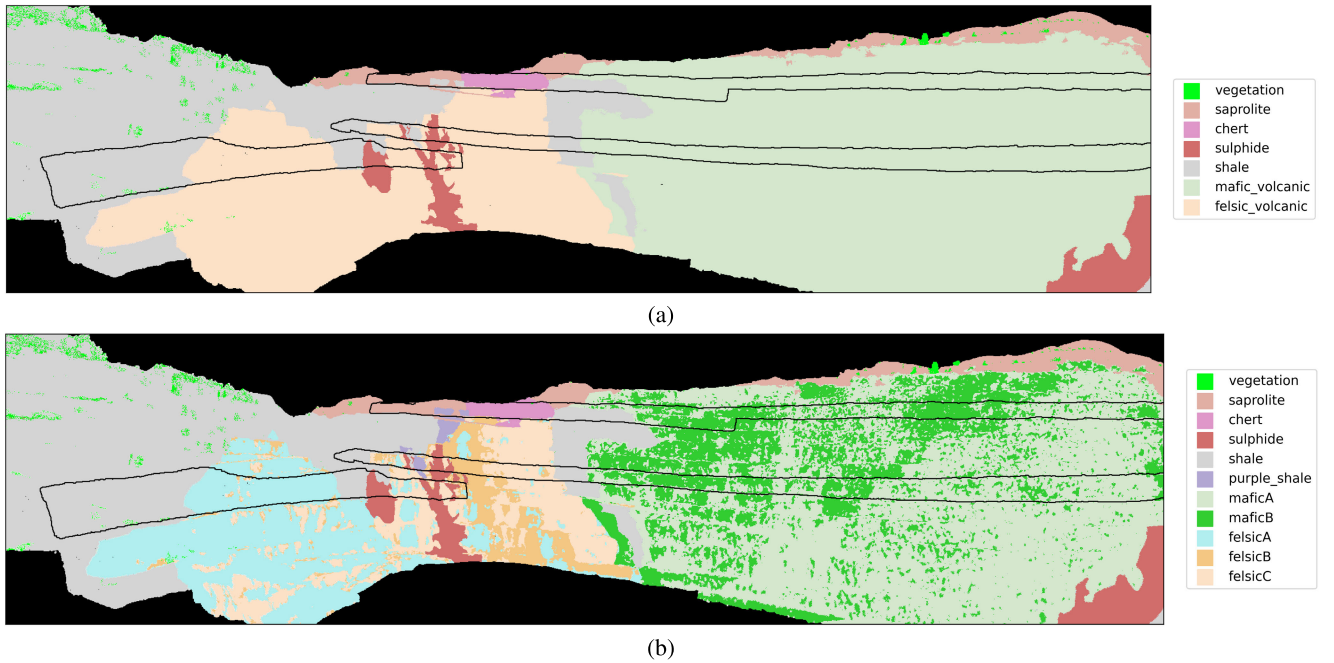
Fig. 2. (Top) Simplified and (Bottom) complete ground-truth labels provided for this benchmark dataset. The suggested training subset is outlined in black and follows traverses that match roughly with how a field geologist would collect data: (a) basic (six-lithology) segmentation and (b) complete (ten-lithology) segmentation.

TABLE I

NUMBER OF POINTS PER CLASS IN THE TRAINING/TESTING SPLIT FOR THE BASIC-LABEL GROUND TRUTH. VEGETATION CLASS IS EXCLUDED DURING TRAINING AND TESTING

Class	Training	Testing	Color
Vegetation*	55,179		
Saprolite	28,231	309,540	
Chert	22,007	27,711	
Sulphide	13,095	128,264	
Shale	49,518	987,790	
Mafic_volcanic	143,861	1,037,782	
Felsic_volcanic	41,256	398,730	

TABLE II

NUMBER OF POINTS PER CLASS IN THE TRAINING/TESTING SPLIT FOR THE COMPLETE-LABEL GROUND TRUTH. VEGETATION CLASS IS EXCLUDED DURING TRAINING AND TESTING

Class	Training	Testing	Color
Vegetation*	55,179		
Saprolite	28,229	309,263	
Chert	22,046	28,953	
Sulphide	13,066	129,085	
Shale	44,321	946,634	
Purple_shale	5,138	40,498	
MaficA	81,092	742,868	
MaficB	62,745	293,460	
FelsicA	20,790	166,804	
FelsicB	6,783	85,818	
FelsicC	13,758	146,434	

training set, a common challenge for hyperspectral classification problems. Tables I and II present the number of points per class in the basic and complete ground truth, respectively.

2) Synthetic Twin: Potential issues associated with remaining biases or inconsistencies in the ground-truth labels have been addressed by generating an entirely synthetic suite of spectral data by forward modeling. These share the same labels as the real dataset, as well as several latent variables and spatial relationships, but are derived using a spectral mixing model and a spatial distribution of mineral abundances simulated using spectral proxies and sample measurements for each class from [12]. We suggest that these synthetic spectra are suited for comparing learning approaches, as the ground truth is known with certainty, while the real spectra can be used to evaluate performance on realistic data. The procedure followed to generate this synthetic twin is outlined below.

First, three latent features known to correlate with specific mineral abundances (spectral proxies) were extracted from the real dataset using established minimum wavelength mapping and band-ratio techniques [12]. These were normalized to have a mean of zero and standard deviation of one and assembled into a vector $\mathbf{L}$ containing the latent feature at every point, ensuring that spatial associations present in the real dataset (and potentially informative for deep machine learning methods) are preserved in the synthetic one.

Next, mineral abundances from X-ray diffraction measurements on the ground-truth samples [12] were used to define a mean composition for each class. To ensure the synthetic abundances sum to one, the so-called additive log ratio (ALR) transformation [58] was used. As reference phase, an abundant phase was chosen, generally quartz. Sulfide was used for the massive sulfide class. Hence, the ALR transformed abundance $\boldsymbol{\alpha}$ of the remaining phases was computed for each point x by

$$\alpha_{i,j}(x) = \log\left(\frac{\widehat{A}_{i,j}}{\widehat{A}_{0,j}}\right) + \sigma \mathbf{M}_i^T \cdot \boldsymbol{\Lambda}_j(x), \quad i = 1, 2, \ldots, n \tag{1}$$

where $\widehat{A}_{i,j}$ denotes the average abundance of mineral $i \in \{0, 1, \ldots, n\}$ in class j—$\widehat{A}_{0,j}$ being then the abundance of the reference mineral for class j—, the vector $\boldsymbol{\Lambda}_j(x)$ contains
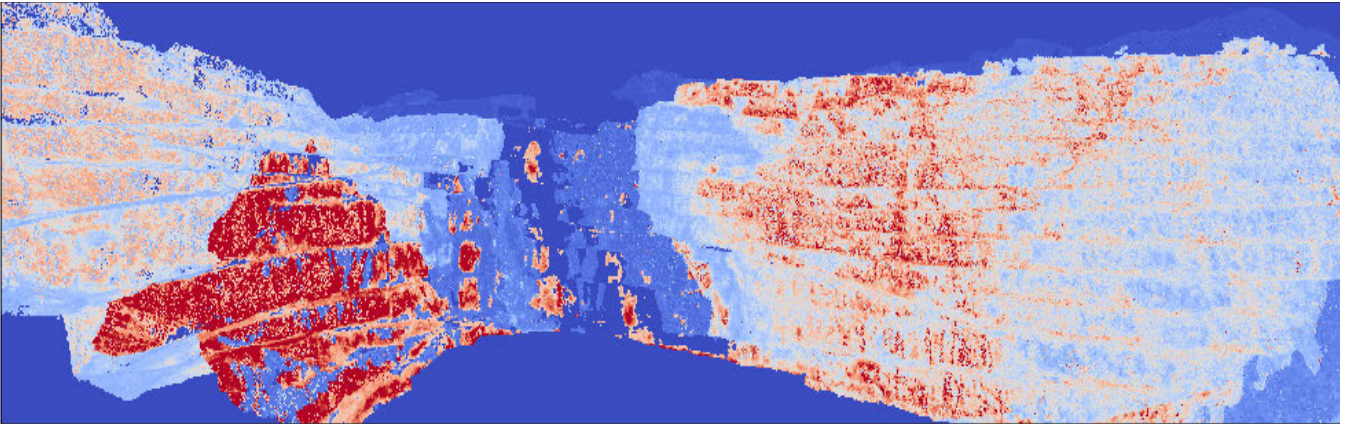
Fig. 3. Example of a simulated mineral abundance map (chlorite in this case). These were used to derive synthetic reflectance spectra with realistic spatial variations.

the values of the three latent variables described previously for class j at location x, and $\mathbf{M}_i$ contains a manually defined mapping vector that determines the contribution of each latent variable to the log-abundance of the ith mineral. Finally, σ scales the log standard deviation of the mineral abundances within each class, and was kept at a constant value of 0.3 after some experimentation.

A vector of closed abundances $\mathbf{A}(x)$ was then calculated for each point x by inverting the additive log transform [58]

$$\mathbf{A}_j(x) = \frac{\exp\begin{bmatrix} 0 & \alpha_{1,j}(x) & \ldots & \alpha_{n,j}(x) \end{bmatrix}}{\left\| \exp\begin{bmatrix} 0 & \alpha_{1,j}(x) & \ldots & \alpha_{n,j}(x) \end{bmatrix} \right\|_1} \quad (2)$$

with $\exp(\cdot)$ the componentwise exponential function, resulting in a set of realistic mineral abundance maps (Fig. 3). Following the real geology exposed in Corta Atalaya, a degree of endmember variability was then introduced by splitting the abundance of three mineral groups (muscovite, chlorite, and clay) into compositional endmembers, based on the position of the 2200-, 2250-, and 2160-nm absorption features, respectively. This extended the number of phases in $\mathbf{A}$ from seven to ten, a realistic degree of complexity for geological outcrops. A pure endmember spectral library $\mathbf{E}$ assembled using spectra from the USGS [59] was then used to derive a synthetic reflectance spectra $\mathbf{S}$ for each data point, assuming linear mixing

$$\mathbf{S} = \mathbf{E} \cdot \mathbf{A}. \quad (3)$$

These synthetic reflectance spectra, and the mineral abundances used to derive them, are also included in the benchmark dataset, and could be used for testing, e.g., endmember identification and unmixing methods.

3) Degraded Twin: In reality, sensor noise and other unwanted effects (e.g., atmospheric and topographic distortions, coating, vegetation) mean that no dataset will contain perfect reflectance spectra. Hence, as a final step, the synthetic reflectance spectra were degraded to simulate realistic measurement, preprocessing and data-correction procedures. First, the reflectance spectra were converted to at-target radiance estimates using the two-light-source atmospheric model described by Thiele et al. [55] and the Oren-Nayar BRDF [60]. Simulating the real acquisition procedure, these radiance data

were projected onto 2-D rasters using three different camera poses, and path-radiance added to the corresponding spectra proportional to the target-sensor distance, resulting in three at-sensor radiance rasters. For the LWIR dataset, light emitted by the target (and by air between the target and the sensor) was also calculated, noting that emissivity $= 1 -$ reflectance following Kirchhoff's law, and added to the at-sensor radiance.

Each raster was then transformed according to the inverse of the sensor-specific lens calibration and converted to digital numbers by dividing by the lab-determined spectral calibration values. Sensor noise was added using dark-current data acquired during the acquisition of the real hyperspectral data, resulting in a set of three simulated raw rasters with realistic noise.

A degraded synthetic reflectance dataset (Fig. 1) was then derived by correcting the simulated raw data using the same routine as was applied to the real data (cf., Section III-A).

C. Accompanying 2-D Data

Although this manuscript focuses on 3-D point cloud data attributed with reflectance spectra to create hyperclouds, it is worth noting that we have included a set of 2-D rasters derived by projecting the class labels, real, synthetic, and degraded spectra onto nadir, oblique perspective and oblique panoramic views (Fig. 4). These will not be discussed further here, but could serve as a useful benchmark for image segmentation or unmixing methods.

IV. EVALUATION

A. Baseline Models

Many deep learning models designed for processing raw point clouds are primarily focused on classification and segmentation tasks. Point coordinates are typically the most common input to the network, and in some cases, normals and RGB values can also be incorporated. Some models proposed sophisticated layers to effectively process point clouds directly [33], [34], [35]. They learn the geometrical features of the points to perform the classification and segmentation tasks. Tinto dataset includes multiple sources of information, but the emphasis is on learning hyperspectral information for
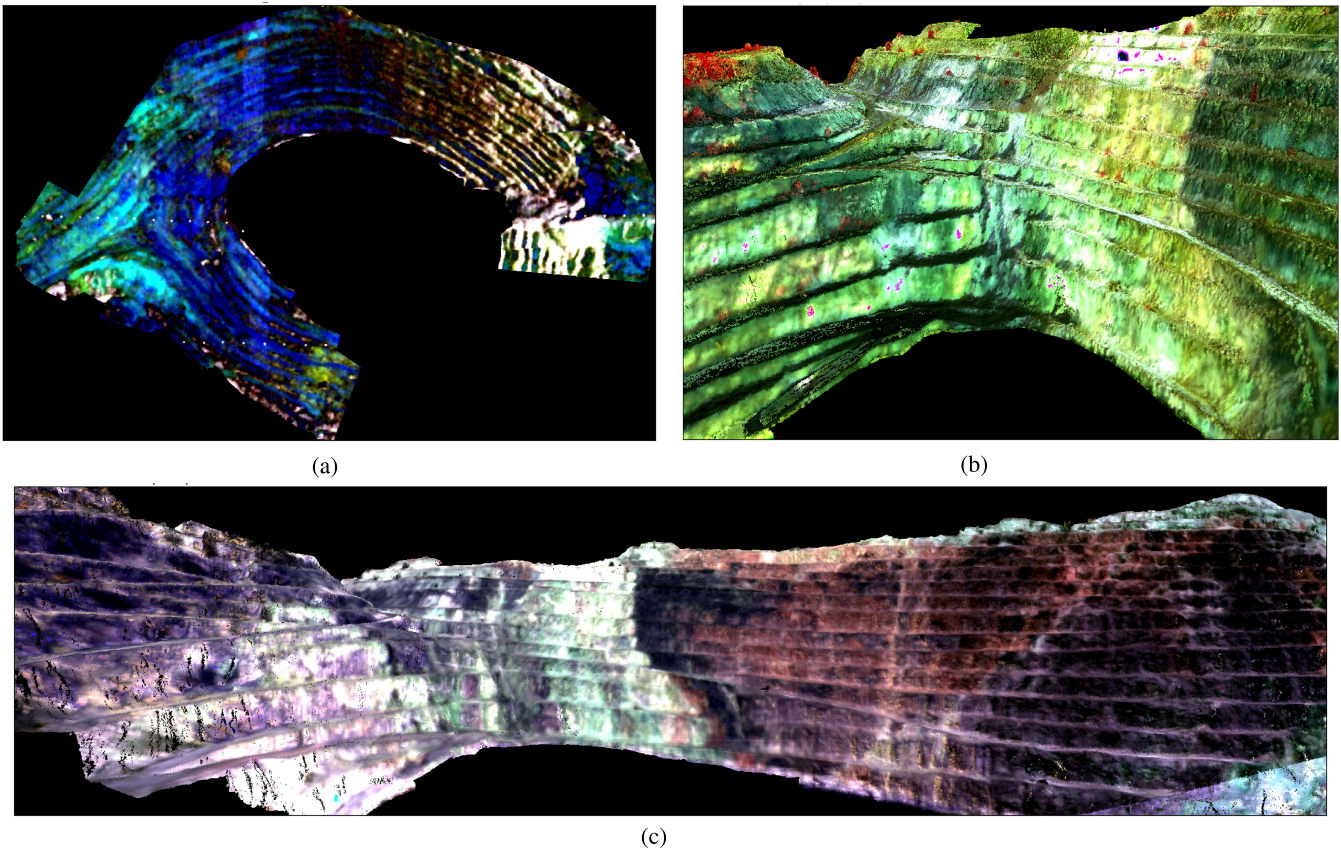
Fig. 4. False-color visualizations of the (Top left) real LWIR, (Top right) VNIR, and (Bottom) SWIR hyperspectral datasets from the three viewpoints used to derive the Tinto2D benchmark images: (a) View1: nadir orthoimage (LWIR: 10114.0, 9181.0, 8545.0 nm); (b) View2: oblique perspective (VNIR: 850.0, 650.0, 525.0 nm); and (c) View3: oblique panorama (SWIR: 2200.0, 2250.0, 2350.0 nm).

point cloud segmentation. To evaluate the Tinto dataset on deep learning models, different deep learning models designed for point cloud processing were selected and trained from scratch. They are categorized into three different architectures according to the main building layers. The first category is the multilayer perceptron (MLP)-based models [61] where the MLPs are the main building layers in the models. The second category is the convolutional-based models where convolutional-like layers are used to process the point clouds. The third category is the transformer-based models where the natural language processing (NLP) Transformer models are adopted and modified to process raw point clouds for different tasks. Following, we will discuss each category and explain the models related to the category.

1) MLP-Based Models: MLP-based models for point cloud processing refer to architectures that utilize MLPs as their fundamental building blocks for analyzing and extracting information from 3-D point cloud data. The first baseline model is a ten-layer MLP. The model was implemented to classify the point cloud according to its hyperspectral information. The output layer is equal to the number of classes. Our network consists of ten hidden layers between the input and the output layers to extract useful features and help in the classification process. The hidden layers have to extract features of different sizes. The input layer takes the hyperspectral information of the point as input information. The output layer has the same size as the number of classes

in the ground truth. Another MLP-based model is the PointNet model [27]. It is the first pioneering deep learning model for direct point cloud processing. As the points are unordered in the point cloud data, it processes each point in isolation through a shared MLP to extract local features. Specifically, PointNet applies point-wise operations using several MLP layers to extract independent features separately and uses max-pooling operation to capture the global features of the point cloud. The aggregated global feature extracted from the point cloud can be used for various tasks such as classification and segmentation. The drawback of the PointNet model is that features are learned independently and then the global feature is aggregated. So, the local structure of the point cloud between points is not captured. To overcome this limitation, PointNet++ [28] was proposed as a hierarchical network. PointNet++ consists of three main layers: the sampling layer, the grouping layer, and the PointNet-based learning layer. The sampling layer uses the farthest point sampling algorithm to select centroids. The grouping layer uses the selected centroids to find the nearest neighbor points of each centroid. The PointNet layer is then applied on the local region to learn and extract the feature vector. This process is repeated in a hierarchical form and the points' resolution is reduced as the network goes deeper. In the last layer, the global feature is produced. However, the calculated KNN is not updated as the input goes deeper into the network. The DGCNN [29] creates dynamic graphs that capture the relationships between

TABLE III
PERFORMANCE OF THE BASELINE MODELS ON THE TINTO DATASET

Model	Real Data			Clean Synthetic Data			Noisy Synthetic Data		
	LWIR	SWIR	VNIR	LWIR	SWIR	VNIR	LWIR	SWIR	VNIR
MLP	38.8	71.5	60.4	92.7	93.6	90.5	75.1	77.2	73.1
PointNet [27]	43.2	67.7	**62.1**	96.7	96.4	96.0	82.5	84.9	83.2
PointNet++ [28]	43.0	71.5	60.7	96.8	95.9	96.2	**83.9**	85.3	83.1
DGCNN [29]	42.7	68.9	59.1	**97.1**	**96.9**	**96.8**	83.5	**86.0**	**83.5**
PointCNN [34]	**47.3**	59.7	58.0	90.1	95.0	87.7	76.9	69.7	69.3
ConvPoint [62]	45.1	67.7	55.4	92.9	89.7	92.9	76.8	77.3	74.1
PT [63]	45.3	60.2	51.2	84.7	95.3	80.1	64.9	80.4	71.0
PCT [64]	43.3	**71.8**	61.2	96.9	96.4	94.0	82.6	85.2	82.8

points. It utilizes EdgeConv layers to extract the characteristics of individual points while employing nonlinear MLPs on neighboring points. Subsequently, the edges are iteratively adjusted in subsequent layers based on the input from prior layers, creating a dynamic workflow.

2) Convolutional-Based Models: The success of CNNs for image classification inspired researchers to use a custom convolutional-like operation for point clouds. PointCNN [34] is designed to process the point cloud with the ability to efficiently learn and capture patterns from point clouds without relying on predefined grids or structures. It achieves this by introducing a unique convolution operation called X-Conv that adapts to the local geometry of the points by learning the weights of the input features and then permuting the points into canonical order, making it highly effective for tasks like 3-D object classification and segmentation. ConvPoint [62] is another model where a customized continuous convolution operation was introduced to process the unordered point clouds. This operation can be extended easily to build a CNN model to process the point clouds similar to 2-D CNNs. The convolution operation separates the kernel into spatial and feature parts. A unit sphere is used to select the location of the spacial parts randomly and the weighted function of the layers is learned by a simple MLP.

These proposed approaches and others underscore the adaptability and effectiveness of custom convolutional-like operations in extracting meaningful features from point cloud data, further expanding the horizons of 3-D data analysis.

3) Transformer-Based Models: Inspired by the popularity of the transformer models in NLP, researchers proposed different Transformer models to process the point clouds. It is under the assumption that the point cloud format suits the self-attention operator due to invariance to permutation and cardinality. Point transformer (PT) [63] proposed a transformer model with self-attention layers as a set operator that can process the point clouds for various tasks. The attention layers learn the relationship between selected central points and the corresponding neighboring points. Those attention layers serve as the backbone for the feature encoder block, which gradually downsamples the number of points in each consecutive layer. Point cloud transformer (PCT) [64] is permutation invariant and it enhances the input embedding by applying the farthest point sampling for centroids selecting and nearest neighbors calculating. PCT proposed an attention mechanism where the final output features from the attention layer are the offset

features that are the difference between the input and the original attention features.

In summary, transformer-based models tailored for point clouds leverage the self-attention mechanism to exploit the unique characteristics of point cloud data, demonstrating promising potential for a wide range of applications in 3-D perception and analysis.

B. Implementation Details

We implemented, trained, and evaluated the MLP model on TensorFlow [65]. The weights of the model were initialized using the Xavier initialization method [66]. For the remaining models, we used the original implementation codes from the GitHub repositories, all with Adam optimizer [67], except DGCNN and PCT used stochastic gradient descent (SGD) [68] with the momentum of 0.9. The dataset contains 3 187 785 points (excluding the vegetation class). For the complete label scenario, the dataset is split into a training set (2 97 968 points) and a testing set (28 89 817 points) with the ratio of 10% and 90%, respectively. The inputs were the hyperspectral information of the points. The input size depends on the sensor used to acquire the data (VNIR = 51 bands, SWIR = 141 bands, and LWIR = 126 bands). The learning rate was set to 0.001, except PCT used 0.1. All models used a batch size of 16, except PointNet and PointNet++ used 24 and DGCNN used 32. The models were implemented in PyTorch [69], and only PointCNN was applied in TensorFlow.

C. Experimental Results

We conducted experiments on the testing split of the Tinto dataset to assess the performance of the baseline models. All hyperspectral point clouds of the VNIR, SWIR, and LWIR (the real, clean synthetic, and noisy synthetic) are utilized for the quantitative and qualitative evaluation. The baseline models trained and tested on the Tinto dataset are MLP, PointNet [27], PointNet++ [28], DGCNN [29], PointCNN [34], ConvPoint [62], PT [63], and PCT [64].

First, we trained the baseline models separately on each sensor data using the training set of real data with complete labels. The trained models were then evaluated on the testing set and their accuracies were computed. Table III (real data) reports the overall accuracy of the baseline models on the real data for the complete labels. The results indicated that the PointCNN model achieved the highest accuracy on the LWIR data, the PCT model achieved the highest accuracy on
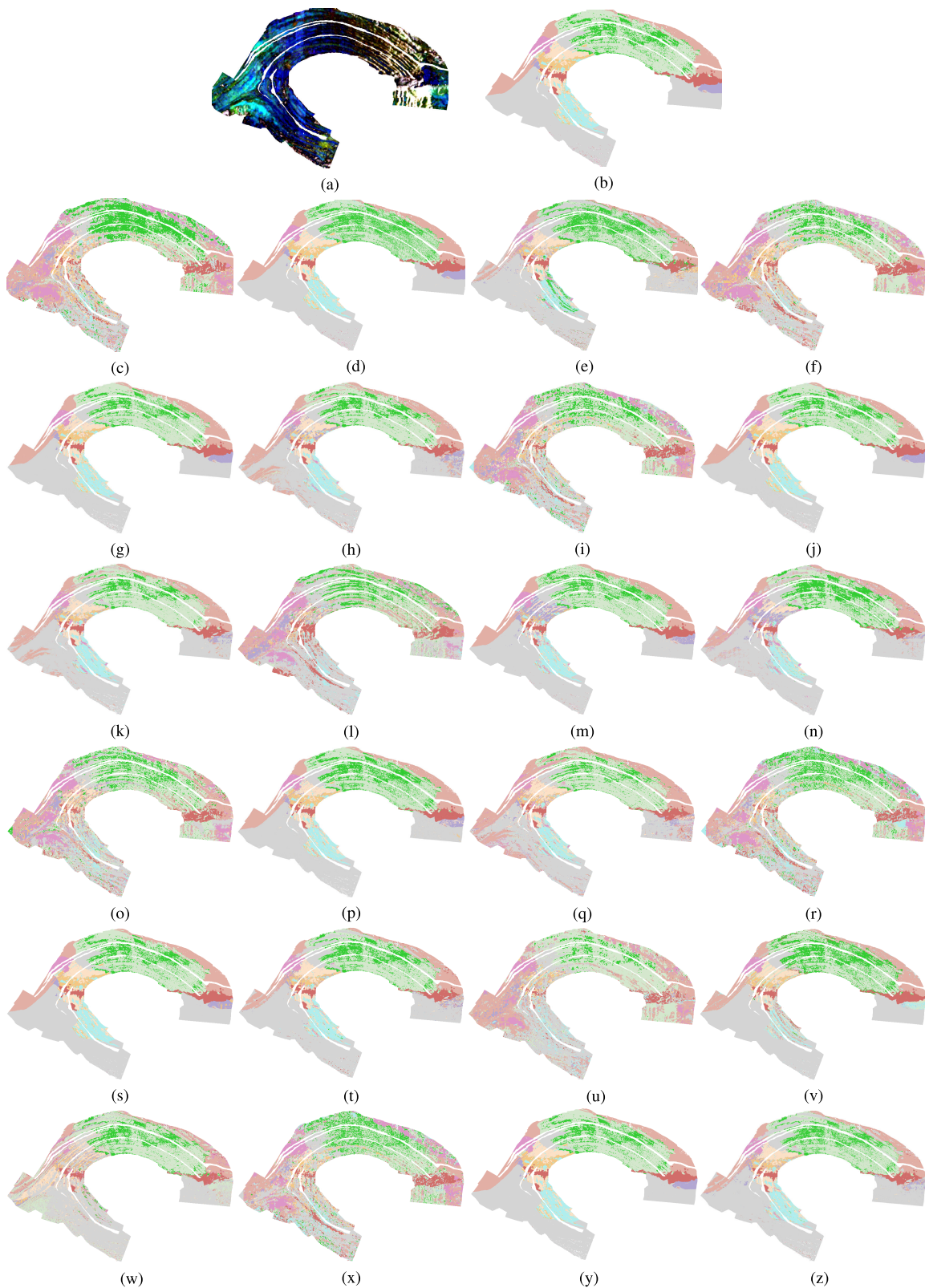
Fig. 5. Segmentation results of the baseline models on the LWIR test in various scenarios: (a) LWIR; (b) ground truth; (c) LWIR MLP real; (d) LWIR MLP clean; (e) LWIR MLP noisy; (f) LWIR PointNet real; (g) LWIR PointNet clean; (h) LWIR PointNet noisy; (i) LWIR PointNet2 real; (j) LWIR PointNet2 clean; (k) LWIR PointNet2 noisy; (l) LWIR PointCNN real; (m) LWIR PointCNN clean; (n) LWIR PointCNN noisy; (o) LWIR ConvPoint real; (p) LWIR ConvPoint clean; (q) LWIR ConvPoint noisy; (r) LWIR DGCNN real; (s) LWIR DGCNN clean; (t) LWIR DGCNN noisy; (u) LWIR PT real; (v) LWIR PT clean; (w) LWIR PT noisy; (x) LWIR PCT real; (y) LWIR PCT clean; and (z) LWIR PCT noisy.

the SWIR data, and the PointNet model achieved the highest accuracy on the VNIR data. Most baseline models were proposed and designed to capture the geometric information from 3-D point clouds of shapes and performed well with objects and scenes that can be segmented into grids. These models mainly trained on the coordinates of the points and
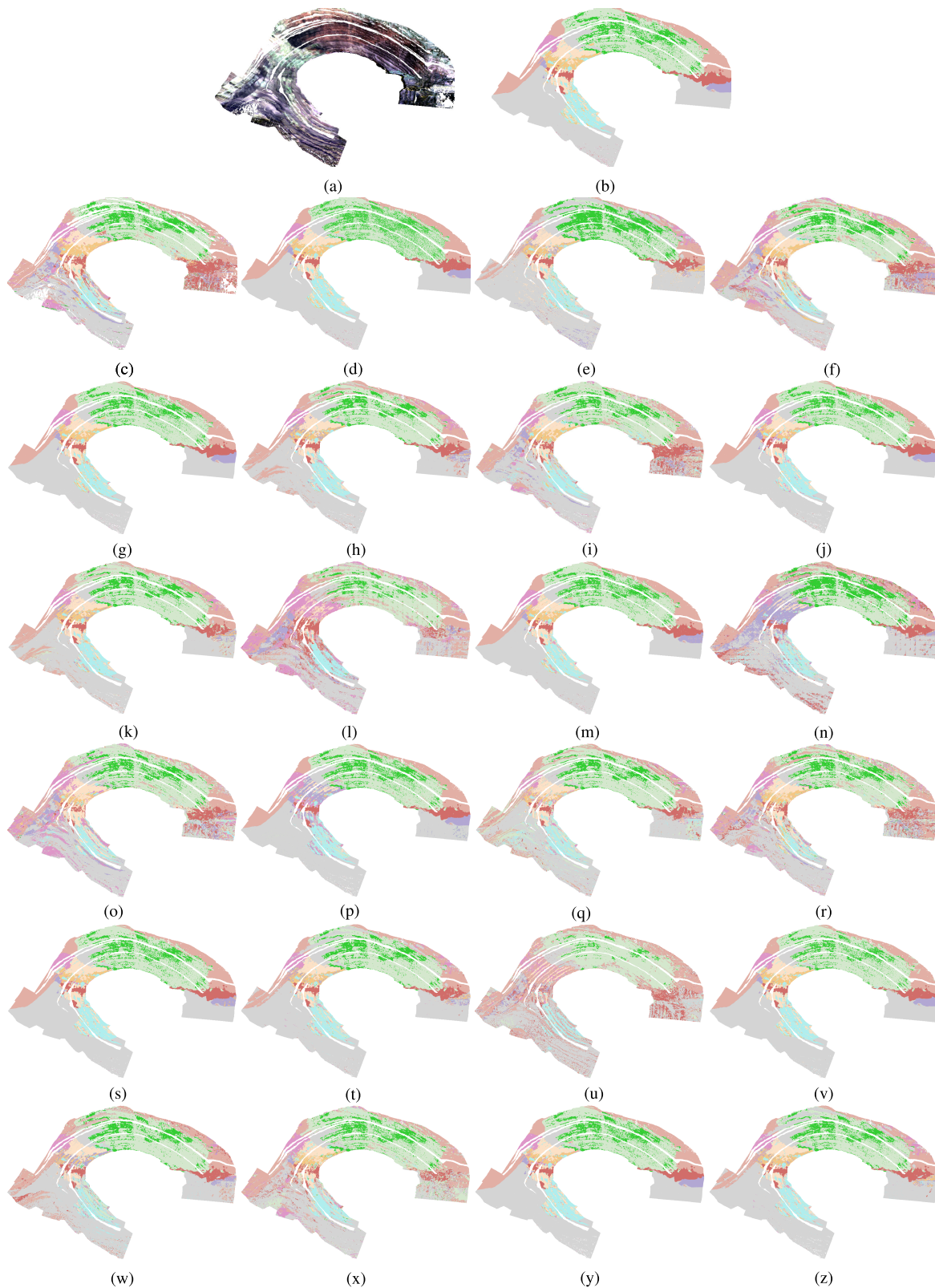
Fig. 6. Segmentation results of the baseline models on the SWIR test in various scenarios: (a) SWIR; (b) ground truth; (c) SWIR MLP real; (d) SWIR MLP clean; (e) SWIR MLP noisy; (f) SWIR PointNet real; (g) SWIR PointNet clean; (h) SWIR PointNet noisy; (i) SWIR PointNet2 real; (j) SWIR PointNet2 clean; (k) SWIR PointNet2 noisy; (l) SWIR PointCNN real; (m) SWIR PointCNN clean; (n) SWIR PointCNN noisy; (o) SWIR ConvPoint real; (p) SWIR ConvPoint clean; (q) SWIR ConvPoint noisy; (r) SWIR DGCNN real; (s) SWIR DGCNN clean; (t) SWIR DGCNN noisy; (u) SWIR PT real; (v) SWIR PT clean; (w) SWIR PT noisy; (x) SWIR PCT real; (y) SWIR PCT clean; and (z) SWIR PCT noisy.

other information (e.g., normals and RGB values) as input features to extract geometric information from the point cloud. However, we modified the baseline models and trained them on the hyperspectral data only. Their performance is lower on this dataset compared to the models' performance on other datasets. Moreover, it's worth mentioning that the accuracy of
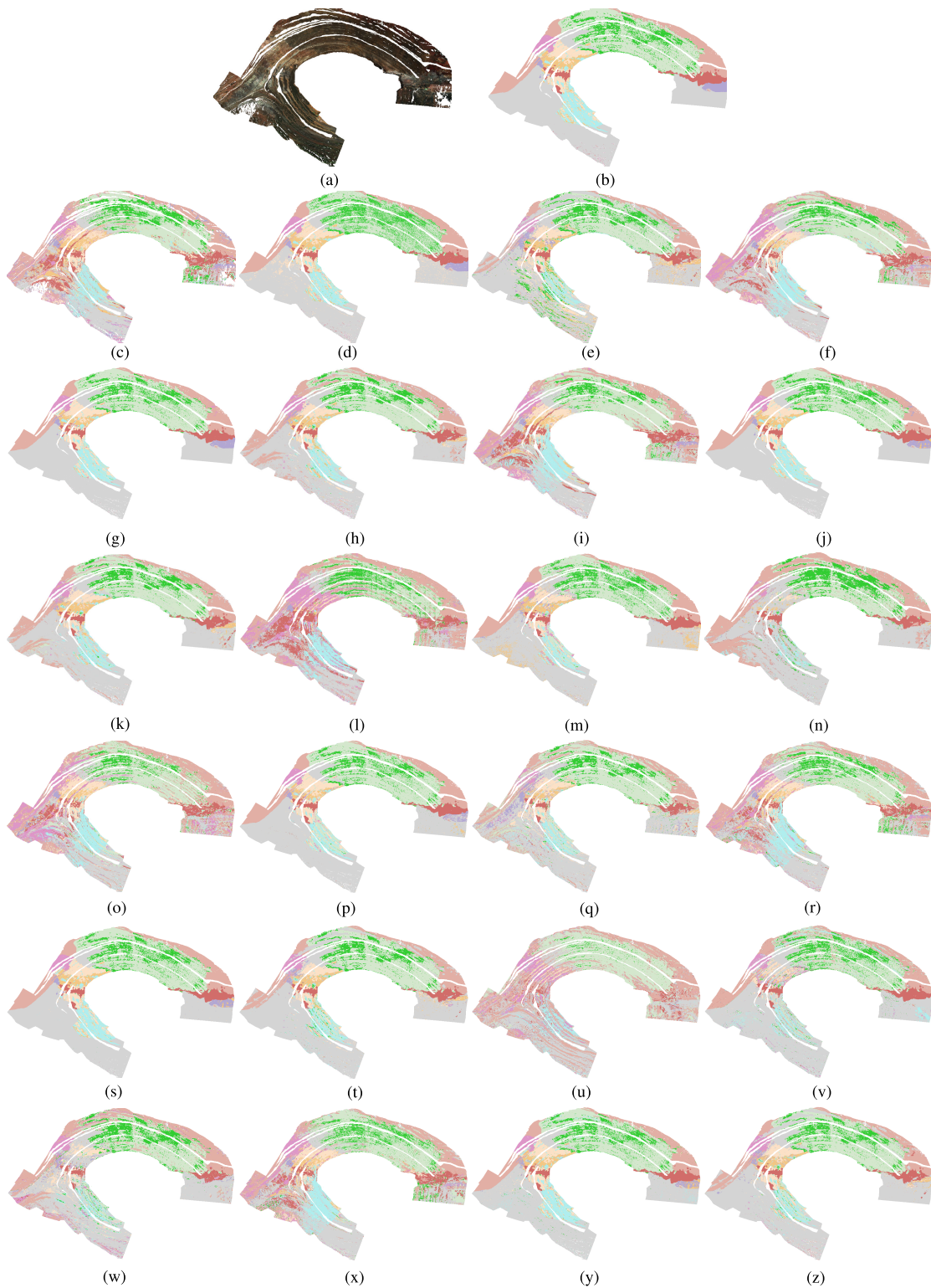
Fig. 7.   Segmentation results of the baseline models on the VNIR test in various scenarios: (a) VNIR; (b) ground truth; (c) VNIR MLP real; (d) VNIR MLP clean; (e) VNIR MLP noisy; (f) VNIR PointNet real; (g) VNIR PointNet clean; (h) VNIR PointNet noisy; (i) VNIR PointNet2 real; (j) VNIR PointNet2 clean; (k) VNIR PointNet2 noisy; (l) VNIR PointCNN real; (m) VNIR PointCNN clean; (n) VNIR PointCNN noisy; (o) VNIR ConvPoint real; (p) VNIR ConvPoint clean; (q) VNIR ConvPoint noisy; (r) VNIR DGCNN real; (s) VNIR DGCNN clean; (t) VNIR DGCNN noisy; (u) VNIR PT real; (v) VNIR PT clean; (w) VNIR PT noisy; (x) VNIR PCT real; (y) VNIR PCT clean; and (z) VNIR PCT noisy.

the labels associated with the real data is not guaranteed to be 100%, which can impact the models' performance to some extent. This is because of the nature of the dataset as there are no sharp boundaries between the classes as the rocks are highly overlapped in reality.

To address the issue of inaccurate ground truth labels, the dataset has a synthetic part where each point in the point cloud has a synthetic hyperspectral feature and is associated with a correct class label. We trained the baseline models on the training set and evaluated them on the testing set of

various sensors. Table III (clean synthetic data) reports the overall accuracy of the baseline models on the clean synthetic data for the complete labels. We found that the performance of all models on the clean synthetic data scored higher accuracies compared to the performance on the real data with a large margin. Interestingly, the majority of models performed similarly on clean synthetic data from different hyperspectral sensors. Our observation is that precise ground truth data can greatly enhance model performance and enable accurate segmentation of the point cloud. Also, Table III shows that the models that consider the neighboring points when extracting the features scored the best accuracies. The DGCNN model, which relies on the local neighboring points to create and update the graph for learning the point features, outperforms other models on all sensors data. Then, it is followed by PCT and PointNet++ models. Both models learn the local features by considering the neighboring points and merge them with the global features computed by the self-attention layer for PCT and the max-pooling layer for PointNet++. Our experiment demonstrated that those models are better suited for our synthetic dataset.

In order to increase the realism of the synthetic data and challenge the models further, we added real noise information (sensor noise and processing artifacts) to the synthetic point cloud. We evaluated the baseline models on the noisy synthetic data and found that their accuracy decreased compared to those trained on clean synthetic data. Table III (noisy synthetic data) reports the overall accuracy of the baseline models on the noisy synthetic data for the basic and complete labels. While all models achieved higher performance compared to the models trained on the real data, DGCNN and PointNet++ models are consistently the leading models on the synthetic data. The DGCNN model achieved the highest accuracy on the SWIR and VNIR data while PointNet++ achieved the highest accuracy on the LWIR data.

In conclusion, the baseline models proved that they can be adapted to learn hyperspectral information and perform the point cloud segmentation task on the geological data. This opens a new direction of applying deep learning models to generate segmented maps on the geological data using hyperspectral information and propose new models that can fuse information from different sources.

Figs. 5–7 showcase the qualitative outcomes of segmented point clouds generated by the trained baseline models using the testing split of the dataset on the LWIR, SWIR, and VNIR data, respectively, in different scenarios (real data, clean synthetic data, and noisy synthetic data). These illustrations highlight the baseline models' ability to produce segmented point clouds with a reasonably high degree of accuracy. When applied to clean synthetic data, most models excel in accurately segmenting the point cloud. However, it's crucial to acknowledge that real-world data typically contains noise, and our models exhibit decreased performance when noise is introduced into the synthetic data. Furthermore, the presence of imbalanced data can lead to misclassifications, particularly in instances where certain classes have a disproportionately smaller number of training samples compared to others. This issue becomes particularly noticeable in the case of the

purple shale class, both in real data and noisy synthetic data scenarios.

V. CONCLUSION

This article introduces the first-ever fully labeled multisensor hyperspectral benchmark dataset and we expect it to be a valuable resource for researchers working on point cloud segmentation in geosciences. The dataset is comprehensive and diverse, covering real and synthetic data with different levels of labeling for the classes. The Tinto dataset is a suitable benchmark for developing and evaluating point cloud segmentation algorithms for geological applications. We believe that the Tinto dataset will serve as a benchmark for future studies and contribute to the development of innovative solutions for point cloud segmentation in Earth sciences. Interestingly, deep models that leverage information from neighboring points demonstrate superior performance, thanks to their ability to extract both local and global features for point cloud segmentation. The strength of this dataset is that it is versatile and allows testing architecture robustness under different conditions (e.g., noise and data quality). Overall, the Tinto benchmark dataset represents a significant contribution to the field and holds promise for a broad range of applications in Earth sciences such as mineral exploration, geological mapping, and natural resources management.

REFERENCES

[1] M. Schmitt, P. Ghamisi, N. Yokoya, and R. Hänsch, "EOD: The IEEE GRSS Earth observation database," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2022, pp. 5365–5368.

[2] O. Ghorbanzadeh, Y. Xu, P. Ghamisi, M. Kopp, and D. Kreil, "Landslide4Sense: Reference benchmark data and deep learning models for landslide detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022, doi: 10.1109/TGRS.2022.3215209.

[3] J. A. Hurt, G. J. Scott, D. T. Anderson, and C. H. Davis, "Benchmark meta-dataset of high-resolution remote sensing imagery for training robust deep learning models in machine-assisted visual analytics," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2018, pp. 1–9.

[4] A. X. Chang et al., "ShapeNet: An information-rich 3D model repository," 2015, *arXiv:1512.03012*.

[5] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 5901–5904.

[6] P. Ghamisi et al., "The potential of machine learning for a more responsible sourcing of critical raw materials," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8971–8988, 2021.

[7] S. Lorenz et al., "Three-dimensional, km-scale hyperspectral data of well-exposed Zn–Pb mineralization at black angel mountain, Greenland," *Data*, vol. 7, no. 8, p. 104, Jul. 2022.

[8] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[9] M. Ibrahim, N. Akhtar, K. Ullah, and A. Mian, "Exploiting structured CNNs for semantic segmentation of unstructured point clouds from LiDAR sensor," *Remote Sens.*, vol. 13, no. 18, p. 3621, Sep. 2021.

[10] B. Chen et al., "Hyperspectral LiDAR point cloud segmentation based on geometric and spectral information," *Opt. Exp.*, vol. 27, no. 17, pp. 24043–24059, 2019.

[11] C. Laukamp, M. Haest, and T. Cudahy, "The Rocklea dome 3D mineral mapping test data set," *Earth Syst. Sci. Data*, vol. 13, no. 3, pp. 1371–1383, Mar. 2021.

[12] S. T. Thiele et al., "Multi-scale, multi-sensor data integration for automated 3-D geological mapping," *Ore Geol. Rev.*, vol. 136, Sep. 2021, Art. no. 104252.

[13] E. Grilli, F. Menna, and F. Remondino, "A review of point clouds segmentation and classification algorithms," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 339–344, Feb. 2017.

[14] Y. Xie, J. Tian, and X. X. Zhu, "Linking points with labels in 3D: A review of point cloud semantic segmentation," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 38–59, Dec. 2020.

[15] T. Rabbani, F. Van Den Heuvel, and G. Vosselmann, "Segmentation of point clouds using smoothness constraint," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 36, no. 5, pp. 248–253, 2006.

[16] E. Castillo, J. Liang, and H. Zhao, "Point cloud segmentation and denoising via constrained nonlinear least squares normal estimates," in *Innovations for Shape Analysis: Models and Algorithms*. Berlin, Germany: Springer, 2013, pp. 283–299.

[17] P. J. Besl and R. C. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 2, pp. 167–192, Mar. 1988.

[18] A.-V. Vo, L. Truong-Hong, D. F. Laefer, and M. Bertolotto, "Octree-based region growing for point cloud segmentation," *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 88–100, Jun. 2015.

[19] A. Nurunnabi, D. Belton, and G. West, "Robust segmentation in laser scanning 3D point cloud data," in *Proc. Int. Conf. Digit. Image Comput. Techn. Appl. (DICTA)*, Dec. 2012, pp. 1–8.

[20] J. Zhang, X. Lin, and X. Ning, "SVM-based classification of segmented airborne LiDAR point clouds in urban areas," *Remote Sens.*, vol. 5, no. 8, pp. 3749–3775, Jul. 2013.

[21] N. Chehata, L. Guo, and C. Mallet, "Airborne LiDAR feature selection for urban classification using random forests," *Laserscanning*, vol. XXXVIII-3/W8, pp. 207–212, 2009.

[22] S. Srivastava, M. R. Gupta, and B. A. Frigyik, "Bayesian quadratic discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, no. 6, pp. 1277–1305, 2007.

[23] E. H. Lim and D. Suter, "3D terrestrial LiDAR classifications with super-voxels and multi-scale conditional random fields," *Comput.-Aided Des.*, vol. 41, no. 10, pp. 701–710, Oct. 2009.

[24] Y. Lu and C. Rasmussen, "Simplified Markov random fields for efficient semantic labeling of 3D point clouds," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 2690–2697.

[25] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 945–953.

[26] D. Maturana and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 922–928.

[27] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 77–85.

[28] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5099–5108.

[29] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Trans. Graph.*, vol. 38, no. 5, pp. 1–12, Oct. 2019.

[30] Y. Lin, G. Vosselman, and M. Y. Yang, "Weakly supervised semantic segmentation of airborne laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 187, pp. 79–100, May 2022.

[31] R. Huang, Y. Xu, and U. Stilla, "GraNet: Global relation-aware attentional network for semantic segmentation of ALS point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 177, pp. 1–20, Jul. 2021.

[32] Y. Lin, G. Vosselman, Y. Cao, and M. Y. Yang, "Local and global encoder network for semantic segmentation of airborne laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 176, pp. 151–168, Jun. 2021.

[33] W. Wu, Z. Qi, and L. Fuxin, "PointConv: Deep convolutional networks on 3D point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9613–9622.

[34] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: Convolution on X-transformed points," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 820–830.

[35] G. Cai et al., "Spatial aggregation net: Point cloud semantic segmentation based on multi-directional convolution," *Sensors*, vol. 19, no. 19, p. 4329, Oct. 2019.

[36] X. Ma, C. Qin, H. You, H. Ran, and Y. Fu, "Rethinking network design and local geometry in point cloud: A simple residual MLP framework," 2022, *arXiv:2202.07123*.

[37] Q. Xu, X. Sun, C.-Y. Wu, P. Wang, and U. Neumann, "Grid-GCN for fast and scalable point cloud learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5660–5669.

[38] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys, "Semantic3D.Net: A new large-scale point cloud classification benchmark," 2017, *arXiv:1704.03847*.

[39] A. Quadros, J. P. Underwood, and B. Douillard, "An occlusion-aware feature for range images," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 4428–4435.

[40] W. Tan et al., "Toronto-3D: A large-scale mobile LiDAR dataset for semantic segmentation of urban roadways," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 797–806.

[41] J. Behley et al., "SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9296–9306.

[42] J. Sun, Q. Zhang, B. Kailkhura, Z. Yu, C. Xiao, and Z. M. Mao, "Benchmarking robustness of 3D point cloud recognition against common corruptions," 2022, *arXiv:2201.12296*.

[43] A. López, J. M. Jurado, J. R. Jiménez-Pérez, and F. R. Feito, "Generation of hyperspectral point clouds: Mapping, compression and rendering," *Comput. Graph.*, vol. 106, pp. 267–276, Aug. 2022.

[44] M. Brell, K. Segl, L. Guanter, and B. Bookhagen, "3D hyperspectral point cloud generation: Fusing airborne laser scanning and hyperspectral imaging sensors for improved object-based information extraction," *ISPRS J. Photogramm. Remote Sens.*, vol. 149, pp. 200–214, Mar. 2019.

[45] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "MUUFL Gulfport hyperspectral and LiDAR airborne data set," Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570, 2013.

[46] Y. Xu et al., "Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1709–1724, Jun. 2019.

[47] M. Weinmann and M. Weinmann, "Fusion of hyperspectral, multispectral, color and 3D point cloud information for the semantic interpretation of urban environments," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 1899–1906, Jun. 2019.

[48] M. Weinmann and M. Weinmann, "Geospatial computer vision based on multi-modal data—How valuable is shape information for the extraction of semantic information?" *Remote Sens.*, vol. 10, no. 2, p. 2, Dec. 2017.

[49] M. Weinmann, P. M. Maier, J. Florath, and U. Weidner, "Investigations on the potential of hyperspectral and Sentinel-2 data for land-cover/land-use classification," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 4, pp. 155–162, Sep. 2018.

[50] L. Weidner, G. Walton, and A. Krajnovich, "Classifying rock slope materials in photogrammetric point clouds using robust color and geometric features," *ISPRS J. Photogramm. Remote Sens.*, vol. 176, pp. 15–29, Jun. 2021.

[51] K. T. Decker and B. J. Borghetti, "Composite style pixel and point convolution-based deep fusion neural network architecture for the semantic segmentation of hyperspectral and LiDAR data," *Remote Sens.*, vol. 14, no. 9, p. 2113, Apr. 2022.

[52] K. T. Decker and B. J. Borghetti, "Hyperspectral point cloud projection for the semantic segmentation of multimodal hyperspectral and LiDAR data with point convolution-based deep fusion neural networks," *Appl. Sci.*, vol. 13, no. 14, p. 8210, Jul. 2023.

[53] I. Mitschke, T. Wiemann, F. Igelbrink, and J. Hertzberg, "Hyperspectral 3D point cloud segmentation using RandLA-Net," in *Proc. Int. Conf. Intell. Auton. Syst.* Springer, 2022, pp. 301–312.

[54] M. Kirsch, S. Lorenz, S. Thiele, and R. Gloaguen, "Characterisation of massive sulphide deposits in the Iberian pyrite belt based on the integration of digital outcrops and multi-scale, multi-source hyperspectral data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2021, pp. 126–129.

[55] S. T. Thiele, S. Lorenz, M. Kirsch, and R. Gloaguen, "A novel and open-source illumination correction for hyperspectral digital outcrop models," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022, doi: 10.1109/TGRS.2021.3098725.

[56] A. Afifi et al., "Multisensor benchmark for 3D hyperspectral point cloud segmentation in the geosciences," Apr. 2023, doi: 10.14278/rodare.2256.

[57] M. Schütz et al., "Potree: Rendering large point clouds in web browsers," Technische Universität Wien, Vienna, Austria, 2016.

[58] J. Aitchison, *The Statistical Analysis of Compositional Data* (Monographs on Statistics and Applied Probability). London, U.K.: Chapman & Hall, 1986, p. 416.

[59] R. F. Kokaly et al., "USGS spectral library version 7 data: US geological survey data release," United States Geol. Surv. (USGS), Reston, VA, USA, 2017, vol. 61.

[60] M. Oren and S. K. Nayar, "Generalization of Lambert's reflectance model," in *Proc. 21st Annu. Conf. Comput. Graph. Interact. Techn.*, 1994, pp. 239–246.

[61] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," *WSEAS Trans. Circuits Syst.*, vol. 8, no. 7, pp. 579–588, 2009.

[62] A. Boulch, "ConvPoint: Continuous convolutions for point cloud processing," *Comput. Graph.*, vol. 88, pp. 24–34, May 2020.

[63] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, "Point transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 16239–16248.

[64] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. R. Martin, and S.-M. Hu, "PCT: Point cloud transformer," *Comput. Vis. Media*, vol. 7, no. 2, pp. 187–199, Jun. 2021.

[65] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *Proc. OSDI*, Savannah, GA, USA, vol. 16, 2016, pp. 265–283.

[66] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

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

[68] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proc. 30th Int. Conf. Mach. Learn. (ICML)*, vol. 28, 2013, pp. III-1139–III-1147.

[69] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 8026–8037.

Samuel T. Thiele received the M.Sc. degree in geology from the University of Western Australia, Perth, WA, Australia, in 2015.

Following this, he was awarded a Westpac Future Leader scholarship to pursue a Ph.D. project on the application of digital mapping technologies to volcanic landslides and magma plumbing systems. Since 2019, he has worked at the Helmholtz-Institute Freiberg for Resource Technology, Freiberg, Germany, with a focus on outcrop and drillcore hyperspectral remote sensing techniques and their application to minerals exploration and digital outcrop mapping.

Aldino Rizaldy received the M.Sc. degree in geoinformation from Faculty ITC, University of Twente, Twente, The Netherlands, in 2018. He is currently pursuing the Ph.D. degree with the Helmholtz-Institute Freiberg for Resource Technology, Freiberg, Germany.

His main research interests are deep learning, computer vision, 3-D point cloud, LiDAR, and photogrammetry.

Sandra Lorenz received the Ph.D. degree "Doctor rerum naturalis" in geosciences from Technical University Bergakademie Freiberg, Freiberg, Germany, in 2019.

She is a Research Associate with the Helmholtz-Institute Freiberg for Resource Technology, Freiberg, where she currently leads the research group for digital processing within the Department of Exploration. Her main research interest lies in the development of spectral imaging as noninvasive material characterization method at different scales, with a particular focus on spectral image and point-cloud processing and the implementation of autonomous platforms for hyperspectral surveys.
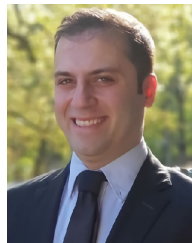
Pedram Ghamisi (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2015.

He currently serves as the Head of the Machine Learning Group at the Helmholtz-Institute Freiberg for Resource Technology, Freiberg, Germany. He also holds the position of a Research Professor and a Senior Principal Investigator, leading the AI for Remote Sensing (AI4RS) team, at the Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna, Austria. In addition to his roles in Germany and Austria, he holds a Visiting Professorship at Lancaster University, Lancaster, U.K. He specializes in deep learning for remote sensing applications.

Ahmed J. Afifi received the bachelor's and M.Sc. degrees in computer engineering from the Islamic University of Gaza (IUG), Gaza, Palestine, in 2008 and 2011, respectively, and the Ph.D. degree (Dr.-Ing.) from Technische Universität Berlin, Berlin, Germany, in 2021.

He is currently a Post-Doctoral Researcher with the Helmholtz-Institute Freiberg for Resource Technology (HIF), Freiberg, Germany, and Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, where he is working on 3-D point cloud classification and segmentation. His research interests include computer vision, deep learning, 3-D object reconstruction from a single image, and medical image analysis.

Raimon Tolosana-Delgado is an Engineering Geologist with the University of Barcelona, Barcelona, Spain, and the Technical University of Catalunya, Barcelona, a Doctor in environmental physics and technology with the University of Girona, Girona, Spain, and got a professorial qualification in mathematical geosciences with Technische Universität Bergakademie Freiberg, Freiberg, Germany. He is a Specialist in geostatistics and compositional data analysis, and works at the Helmholtz-Institute Freiberg, Freiberg, for resource technology applying stochastic modeling and data analysis to ore body modeling, geometallurgy, and mineral processing modeling.

Moritz Kirsch received the M.Sc. degree in geology from Technische Universität Bergakademie Freiberg, Freiberg, Germany, in 2008, and the Ph.D. degree from Universidad Nacional Autónoma de México, Mexico City, Mexico, in 2012.

Since 2016, he has been working as a Research Associate with the Helmholtz-Institute Freiberg for Resource Technology (HIF), Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany, where he leads a group on innovative applications with the Department of Exploration. His research interests lie at the intersection of geology and technology, with a focus on leveraging the latest digital tools, such as hyperspectral imaging and 3-D modeling, to characterize mineral assemblages and understand geological processes.

Michael Heizmann received the M.Sc. degree in mechanical engineering and the Ph.D. degree in automated visual inspection from the University of Karlsruhe, Karlsruhe, Germany, in 1998 and 2004, respectively.

Since 2016, he has been a Full Professor in mechatronic measurement systems and the Director with the Institute of Industrial Information Technology (IIIT), Karlsruhe Institute of Technology (KIT), Karlsruhe.

His research interests include machine vision, image processing, image and information fusion, measurement technology, machine learning, artificial intelligence, and their applications.

Richard Gloaguen received the Ph.D. degree "Doctor Communitatis Europae" in marine geosciences from the University of Western Brittany, Brest, France, in collaboration with the Royal Holloway University of London, London, U.K., and Göttingen University, Göttingen, Germany, in 2000.

He was a Marie Curie Post-Doctoral Research Associate at the Royal Holloway University of London from 2000 to 2003. He led the Remote Sensing Group at University Bergakademie Freiberg, Freiberg, Germany, from 2003 to 2013. Since 2013, he has been leading the division "Exploration Technology" at the Helmholtz-Institute Freiberg for Resource Technology, Freiberg. He is currently involved in UAV-based multisource imaging, laser-induced fluorescence, and noninvasive exploration. He is the Head of the Exploration Department at the Helmholtz-Institute Freiberg for Resource Technology. His research interests focus on multisource and multiscale remote sensing integration using computer vision and machine learning.