

Channel Attention Module for Segmentation of 3D Hyperspectral Point Clouds in Geological Applications

Aldino Rizaldy, Pedram Ghamisi, Richard Gloaguen

Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology (HIF),
09599 Freiberg, Germany. - (a.rizaldy, p.ghamisi, r.gloaguen)@hzdr.de

Keywords: Channel Attention, Transformer, 3D Point Cloud, Hyperspectral, Geology

Abstract

We develop a Transformer-based model enhanced with a Channel Attention Module (CAM) to capture the inter-channel dependencies in 3D hyperspectral point cloud data for geological applications. We hypothesize that specific channels of hyperspectral data correspond to distinct mineral types, and therefore, exploiting the relationships among these channels is beneficial for our analysis. We evaluate our method using the newly released Tinto dataset, which consists of 3D hyperspectral point clouds featuring three different spectral ranges: Long Wave Infrared (LWIR), Short Wave Infrared (SWIR), and Visible-Near Infrared (VNIR). We explore four different CAMs from various networks—SENet, ECANet, CBAM, and DANet—and successfully integrate them into a CNN-based model to enhance feature representation. We specifically tailor the channel attention to our use of 3D hyperspectral point cloud data. Our experiments demonstrate significant improvements in performance after incorporating the CAM into our backbone model, which draws inspiration from the Point Cloud Transformer architecture and Vector Self-Attention mechanism. These results highlight the potential for further research into enhancing classification accuracy using hyperspectral data in geological applications. The code will be released on <https://github.com/aldinorizaldy/CAM-Transformer>.

1. Introduction

Although 2D hyperspectral data is extensively employed for mineral classification in numerous geological applications (Ghamisi et al., 2021), recent research advocates for the utilization of 3D data due to its superior ability to capture the complexity and variability of geological scenes (Afifi et al., 2024).

In scenarios where both hyperspectral data and 3D point clouds are available within the same geographical area, merging them into a 3D hypercloud becomes feasible. Unlike conventional 3D point cloud data, which typically comprises XYZ coordinates, a 3D hypercloud incorporates hyperspectral information atop these attributes, thereby expanding the dataset to include spectral data across potentially hundreds of bands.

Several 3D hypercloud datasets exist as open data for scientific purposes. For instance, the Maarmorilik dataset (Lorenz et al., 2022) captures the complex geological structure in West Greenland. More recently, the Tinto dataset (Afifi et al., 2024) has been released as a benchmarking dataset for geological applications using 3D hyperclouds.

Moreover, the task of semantic segmentation of 3D point clouds has been an active area of research in the computer vision community. Since the introduction of PointNet (Qi et al., 2017), deep learning models for point clouds have significantly evolved, culminating in the adoption of the Transformer architecture in recent models such as the Point Cloud Transformer (PCT) (Guo et al., 2021).

Given that hyperspectral data typically comprises hundreds of bands, leveraging channel-wise features has become a common approach to capture meaningful information. Historically, researchers employed 1D CNNs to extract channel-wise features from 2D hyperspectral data. This methodology evolved to incorporate 3D CNNs, enabling the simultaneous extraction

of spectral and spatial features. However, the application of convolution-based filters is limited by the size of the filters. Although these filters excel at extracting local patterns, they lack the capacity to capture broader contexts.

An alternative approach involves measuring the long-range interaction across all channels simultaneously. In this context, attention modules play a crucial role by computing the relationship of each channel to the remaining channels. Some studies have shown that exploiting channel-wise features for 3D point cloud classification, alongside spatial-wise features, enhances the performance of baseline models (Lin et al., 2021).

In our study, we aim to segment 3D hypercloud data for geological applications by focusing on the inter-dependency of feature channels. We argue that in mineral classification, certain channels exhibit distinct responses to specific mineral classes. Therefore, assessing the relationships among all channels proves advantageous for our task, and enable more precise and context-aware classification.

We utilized PCT as the backbone of our model. PCT incorporates four transformer layers as its core to ensure the extraction of rich features. In addition, Vector Self-Attention (VSA) was chosen for the Transformer layers due to its superior performance, as reported by Zhao et al. (2021). We further enhanced the model by integrating a Channel Attention Module (CAM) directly after each Transformer layer. This addition was motivated by our belief that addition of CAM facilitates the recalibration of feature maps to extract more meaningful features.

We reimplemented four different CAMs from different existing neural networks, namely Convolutional Block Attention Module (Woo et al., 2018), Squeeze-and-Excitation Network (Hu et al., 2020), Efficient Channel Attention Network (Wang et al., 2020), and Dual Attention Network (Fu et al., 2019) to investigate the performance of different mechanisms. Our experiments

on Tinto dataset (Afifi et al., 2024) show that inserting CAM modules directly after Transformer layers improved model performance on different range of hyperspectral data.

2. Related Work

2.1 Hyperspectral data for geological applications

Hyperspectral imaging has emerged as a powerful tool for geological applications, enabling researchers to analyze Earth's surface composition with high spectral and spatial resolution. In this section, we review relevant literature that demonstrates the utility of hyperspectral imagery in these domains.

Researchers typically employ two primary methodologies (Ghamisi et al., 2021, 2017): unsupervised and supervised approaches for classification of hyperspectral data. These methods represent distinct strategies for analyzing hyperspectral data and extracting meaningful information about the composition and characteristics of observed surfaces.

Unsupervised approaches. Guo et al. (2014) presented the Spatial Subspace Clustering (SpatSC) method, tailored for segmenting hyperspectral data from drill holes by seamlessly integrating subspace learning with the fused least absolute shrinkage and selection operator. This approach enhances interpretability, particularly in mineralogy differentiation, regardless of the presence or absence of a library of pure spectra. Meanwhile, Francky Fouedjio and Laukamp (2018) introduced geostatistical clustering to address the challenge of unsupervised grouping of drill hole samples into homogeneous domains. By leveraging spatial dependency between samples, this method significantly improves the spatial contiguity of clusters, leading to more accurate domain delineation in mineral deposits.

Subsequently, Shahi et al. (2019) capitalized on advanced subspace clustering algorithms, exploiting the self-representation property of hyperspectral data to achieve precise and reliable domain delineation in geological investigations, eliminating the need for manual intervention. Further advancements are seen in (Shahi et al., 2020a), where multi-sensor data fusion is employed to enhance domain delineation, combining spectral and spatial information from diverse imaging sensors. Additionally, Shahi et al. (2020b) introduces the hierarchical subspace-based clustering algorithm (HESCC) to address challenges in hyperspectral data, aiming to overcome computational complexity and automate cluster number estimation. Finally, Shahi et al. (2021) proposed a novel multi-sensor subspace-based clustering algorithm. Spatial information extraction is facilitated by a hidden Markov random field (HMRF) applied to RGB images, while spectral information is derived using an advanced sparse subspace clustering algorithm on HSIs.

Supervised approaches. Acosta et al. (2019) pioneered the use of supervised methods to improve drill-core hyperspectral data analysis, achieving enhanced speed and accuracy by fusing high-resolution mineralogical analysis with hyperspectral data. Expanding on this, Acosta et al. (2020) proposed a superpixel-based machine learning framework to integrate hyperspectral and geochemical data for core analysis, resulting in more accurate element abundance extrapolation along the core length.

Lorenz et al. (2021) explored hyperspectral imaging for geological mapping, employing advanced feature extraction algorithms

such as PCA, MNF, OTVCA, and WSRRR to improve classification accuracy and interpretability across various mapping campaigns. Additionally, Ghamisi et al. (2020) introduced a multi-sensor feature fusion approach using remote sensing and deep learning for mineral mapping, achieving exceptional classification accuracies and promoting sustainable mining practices.

2.2 Point cloud for geological applications

Point cloud data has attracted the attention of researchers for its potential in geological and mineralogical applications. Walton et al. (2016) introduced a machine learning algorithm designed for automated lithological classification within terrestrial lidar point clouds, aiming to address challenges related to resolution and accuracy commonly associated with mobile lidar platforms. On the other hand, Lorenz et al. (2018) delved into ground-based hyperspectral imaging for mapping near-vertical geological outcrops. Their study presents an adapted workflow that tackles challenges in long-range outcrop sensing by integrating atmospheric and topographic corrections, resulting in improved spectral mapping accuracy. This method enables the creation of large-scale 3D hyperclouds that merge hyperspectral data with 3D photogrammetric point clouds, offering comprehensive insights into geological structures and compositions.

Additionally, Guo et al. (2019) introduced an automatic extraction method for discontinuity trace mapping from 3D point clouds of natural rocky slopes, while Weidner et al. (2021) addressed the need for rapid point cloud interpretation methods in characterizing rock slope hazards using photogrammetry point clouds. The latter study presents a machine learning approach for classifying geological datasets into relevant categories. Moreover, Thiele et al. (2021) introduced a Python workflow, hylite, for creating hyperclouds that integrate geometric information with hyperspectral reflectance data. This approach facilitates data-driven mapping of geological exposures and mineralogy, demonstrating its effectiveness in mapping volcanic-hosted massive sulphide (VHMS) mineralization in an open-pit mine.

2.3 Deep learning on 3D Hypercloud

Decker and Borghetti (2022, 2023) both focus on the fusion of hyperspectral and lidar data, which offer complementary spectral and structural information. They introduced a composite fusion-style architecture for semantic segmentation, leveraging point-based CNN layers for processing lidar 3D point cloud and hyperspectral data. This approach achieves improved pixel accuracy compared to unimodal models.

Mitschke et al. (2023) explored the use of hyperspectral information to enhance semantic segmentation of 3D point clouds, particularly in ambiguous objects like streets, sidewalks, and cars. The study extends the RandLA-Net neural network to integrate hyperspectral histograms, aiming to improve segmentation accuracy for such objects.

Furthermore, (Afifi et al., 2024) addressed the challenge of validating deep learning approaches for geological mapping, especially for non-structured 3D data like point clouds. The study introduces Tinto, a multisensor benchmark digital outcrop dataset containing real and synthetic data with spectral attributes and ground-truth labels. This dataset aims to facilitate the development and validation of deep learning tools for 3D applications in Earth sciences, including semantic segmentation of point clouds for geological mapping.

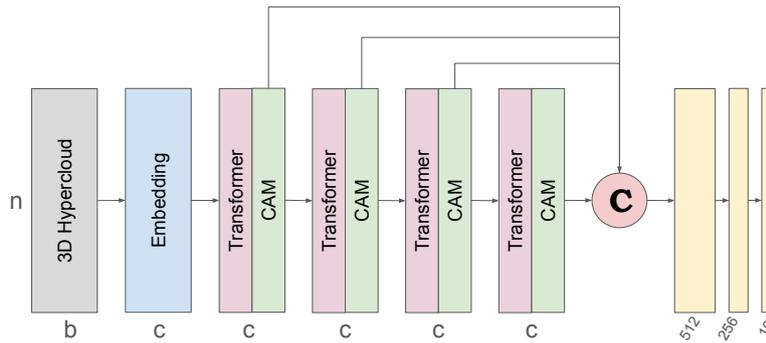


Figure 1. Network architecture.

3. Method

In this section we outline our selection of the network architecture and the Transformer layer, which we adopted from our prior work (Rizaldy et al., 2023). Subsequently, we provide a comprehensive description of the various CAMs that we incorporated into our work.

3.1 Network architecture

Our network’s entire operation adheres to the workflow of the PCT Guo et al. (2021), as depicted in Figure 1. The network consists of two main layers: the Transformer layer and the channel attention layer. Both layers function as feature encoders, with the Transformer layer leveraging spatial features of input points and the channel attention layer enhancing output point features by assessing the significance of each channel of the point features. Detailed explanations of both layers will follow in the subsequent sections.

Let $\hat{\mathbf{P}} \in \mathbb{R}^{n \times b}$ represent the hyperspectral point cloud data as the network’s input, where n and b denote the number of points and hyperspectral bands, respectively. Initially, we embed the input points via two fully-connected layers to yield $\mathbf{P} \in \mathbb{R}^{n \times c}$, where c is the dimension of the embedded features. For our implementation, we have chosen c to be 128. Following this, the embedded features pass through four consecutive blocks, each consisting of a Transformer layer followed by a CA layer. Skip connections are utilized to concatenate the outputs, resulting in 512-dimensional features from various stages. Finally, these enriched features are fed into the classification block, which employs several fully connected layers.

3.2 Transformer layer

The Transformer layer incorporates a self-attention mechanism designed to focus on pertinent information within the input features. This mechanism typically employs scalar dot products for computing an attention map, hence we called this method as scalar self-attention (SSA). PCT Guo et al. (2021) advances SSA by introducing scalar offset-attention (SOA), drawing inspiration from the Laplacian matrix. While scalar operation is conventionally employed in Transformer architectures, an alternative self-mechanism known as vector self-attention (VSA) Zhao et al. (2020) has been proposed.

Let \mathbf{P} be the input features for SSA, the process initiates with SSA transforming \mathbf{P} using linear projections or MLPs denoted as φ , ψ , and α . SSA then utilizes scalar dot products on the transformed features by φ and ψ to produce an attention map.

Having ρ as the softmax function, this attention map is subsequently utilized to scale the features transformed by α . Formally, the output attention feature $\mathbf{X} \in \mathbb{R}^{n \times c}$ using SSA can be defined as:

$$\mathbf{X} = \rho \left(\varphi(\mathbf{P}) \psi(\mathbf{P})^T \right) \alpha(\mathbf{P}) \quad (1)$$

SOA makes a slight modification to SSA by computing the offset (difference) between the attention features \mathbf{X} and the input features \mathbf{P} to derive the output features. Conversely, VSA measures the relationship between the transformed features using a subtraction operator to create the attention map. It then computes the attention features through element-wise multiplication denoted by \odot . Given γ as a non-linear function, the feature vector $\mathbf{x}_i \in \mathbf{X}$ using VSA is defined by Zhao et al. (2021) as:

$$\mathbf{x}_i = \sum_{\mathbf{p}_j \in \mathbf{P}} \rho(\gamma(\varphi(\mathbf{p}_i) - \psi(\mathbf{p}_j))) \odot \alpha(\mathbf{p}_j) \quad (2)$$

where i is the central point and j denotes the neighboring points within a set of local points.

In our methodology, rather than employing the widely-used SSA, we opt for VSA, as recommended in our previous research Rizaldy et al. (2023), where VSA has exhibited superior performance compared to SSA. One key advantage of VSA lies in its exploitation of individual feature channels, in contrast to SSA which operates on entire feature vectors Zhao et al. (2021).

3.3 Channel attention layer

In our implementation, we calculate channel attention features by utilizing the output features of each Transformer layer as the input for the channel attention layer. Several studies have proposed various CAMs to harness channel attention features Woo et al. (2018); Hu et al. (2020); Fu et al. (2019); Wang et al. (2020). Here, we delve into the mechanisms of each module and present performance results on 3D hypercloud data in the following section. Subsequently, we introduce our channel attention module, developed based on our initial investigations, aimed at addressing some limitations of existing methods. Figure 2 shows the whole mechanism of different CAMs.

Squeeze-and-Excitation (SE) module Hu et al. (2020). As the name suggests, the SE module consists of two key operations: squeeze and excitation. Global information is captured through global average pooling (GAP) by reducing spatial dimensions to a single value per channel. This channel descriptor encodes the importance of each channel. The SE module then

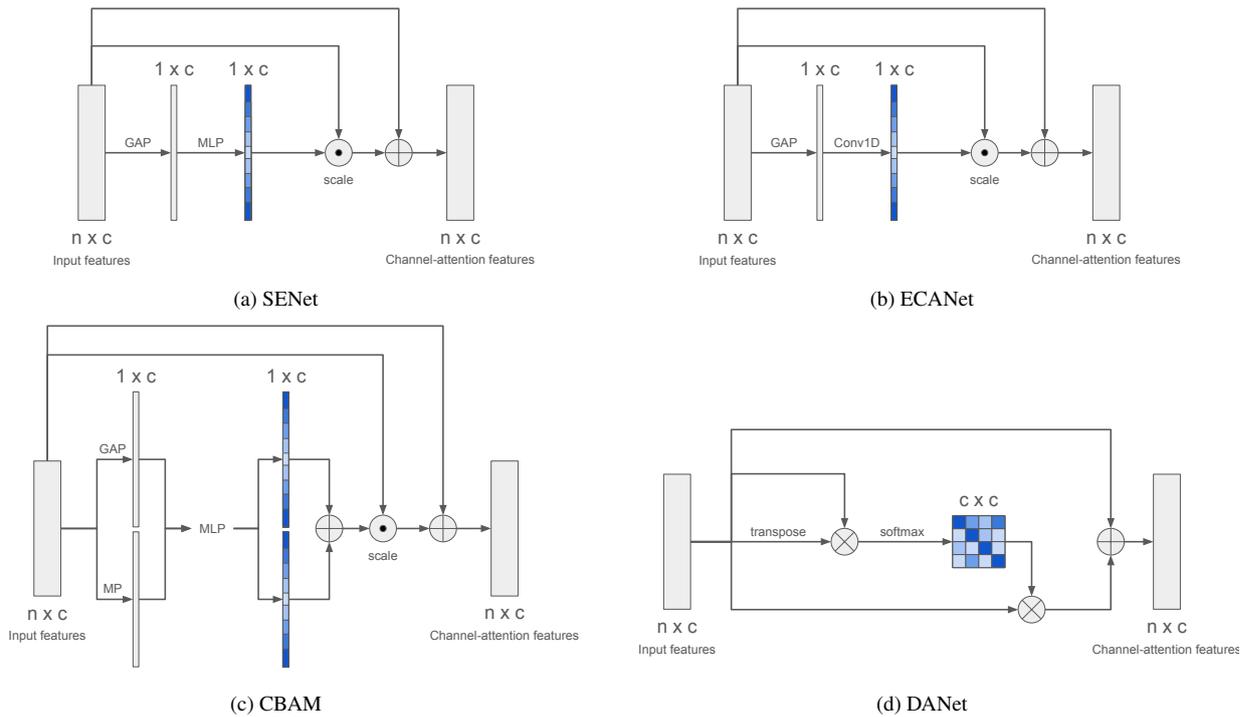


Figure 2. Visual comparison between different channel attention modules and the baselines.

models channel-wise dependencies using an MLP that results in channel-specific weights to recalibrate feature responses. Given \mathbf{X} as the input of the SE module, channel attention features $\mathbf{CA} \in \mathbb{R}^{n \times c}$ are computed in Equation 4.

Efficient channel attention (ECA) module Wang et al. (2020). The ECA module is designed to enhance the efficiency of channel attention. This module replaces the computationally intensive global operations with a more efficient 1D convolution with a kernel size k . Unlike the SE module, the ECA module does not involve dimensionality reduction. The 1D convolution operates locally on the channel domain with a coverage of k , enabling the module to capture channel-wise dependencies efficiently. Using Conv1D as the 1D convolution operator, the ECA module computes \mathbf{CA} as in Equation 5.

Convolutional block attention module (CBAM) (Woo et al., 2018). CBAM consists of two attention modules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). In this work, our focus is on CAM. CAM operates similarly to the SE module, with the key difference being its utilization of two pooling strategies: Global Average Pooling (GAP) and Max Pooling (MP) within a two-branch mechanism. These two branches are processed through a shared MLP to compute two sets of attention weights, which are then summed and applied to the input features. Using MP as the max pooling operator, the CAM computes \mathbf{CA} in Equation 6

Dual Attention Network (DANet) Fu et al. (2019). Similar to CBAM, DANet introduces a Channel Attention Module (CAM) and a Position Attention Module (PAM). Again, our focus lies on CAM. In contrast to previous approaches using MLPs or Convolution operators, DANet employs a self-attention mechanism to compute the attention map $\mathbf{A} \in \mathbb{R}^{c \times c}$. This enables DANet to capture global cross-channel relations. The output of DANet simply computes \mathbf{A} using matrix multiplication directly

with the input features \mathbf{X} :

$$\mathbf{A} = \text{Softmax}(\mathbf{X}^T \mathbf{X}) \quad (3)$$

Following the self-attention concept, DANet proceeds with the second matrix multiplication of \mathbf{A} and \mathbf{X} to derive the channel attention features \mathbf{CA} . Additionally, a scale parameter β is introduced to modulate the channel attention weights as can be seen in Equation 7.

$$\mathbf{CA}_{SE} = \sigma(\text{MLP}(\text{GAP}(\mathbf{X}))) + \mathbf{X} \quad (4)$$

$$\mathbf{CA}_{ECA} = \sigma(\text{Conv1D}(\text{GAP}(\mathbf{X}))) + \mathbf{X} \quad (5)$$

$$\mathbf{CA}_{CBAM} = \sigma(\text{MLP}(\text{GAP}(\mathbf{X})) + \text{MLP}(\text{MP}(\mathbf{X}))) + \mathbf{X} \quad (6)$$

$$\mathbf{CA}_{DANet} = \beta(\mathbf{A}\mathbf{X}) + \mathbf{X} \quad (7)$$

4. Results and Discussion

For our evaluation, we employed the Tinto dataset (Afifi et al., 2024). This dataset provides 3D point cloud data, with each point linked to three distinct hyperspectral sensors: long-wave infrared (LWIR), short-wave infrared (SWIR), and visible-near infrared (VNIR) features. Specifically we used synthetic noisy data of the Tinto dataset due to the accurate ground truth label. Situated in the open-pit mine Corta Atalaya, Minas de Riotinto, Spain, the Tinto dataset includes ground truth information for 10 mineral classes relevant to geological applications.

4.1 Quantitative valuation of different CAMs

Table 1 and 2 present the evaluation results across three different hyperspectral sensors, comparing them with the baseline method. Our assessment utilized Overall Accuracy and F1-score as the accuracy metrics. Overall, we observed that the

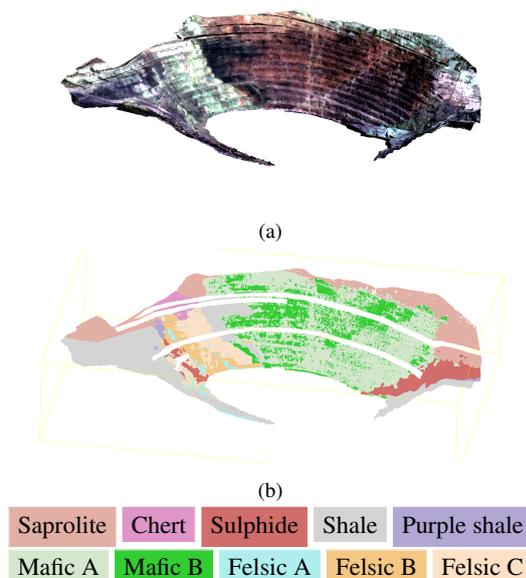


Figure 3. (a) 3D hypercloud of Tinto dataset and (b) the ground truth.

Table 1. Evaluation of different channel attention modules.

Module	Overall Accuracy		
	LWIR	SWIR	VNIR
PCT	82.6	85.2	82.8
PCT+VSA	82.9	86.1	84.0
PCT+VSA+SENet	82.1	86.8	85.8
PCT+VSA+ECANet	83.9	86.8	84.7
PCT+VSA+CBAM	82.9	86.9	85.7
PCT+VSA+DANet	83.3	86.3	84.6

inclusion of CAMs led to an improvement in model accuracy. However, none of the modules stood out as distinctly superior. SENet and CBAM modules demonstrated impressive performance with the SWIR and VNIR datasets, while the ECANet module showcased notable results with the LWIR and SWIR datasets.

Conversely, the performance of the DANet module, which employs a mechanism similar to self-attention, exhibited lower accuracy than the other CAMs. One reason for this could be the lack of learning capability in the DANet module, as the channel attention weight is directly computed from the input features without incorporating learnable weights, unlike other modules. Additionally, the single-head attention mechanism of DANet contrasts with the multi-head attention typically observed in the modern Transformer-based model, limiting its ability to learn different attention weights.

Table 2. Evaluation of different channel attention modules.

Module	FI score		
	LWIR	SWIR	VNIR
PCT	83.1	85.8	83.0
PCT+VSA	82.8	86.0	83.7
PCT+VSA+SENet	81.6	86.4	85.3
PCT+VSA+ECANet	83.3	86.4	84.5
PCT+VSA+CBAM	81.6	86.3	85.3
PCT+VSA+DANet	82.7	86.0	84.4

4.2 Visual inspection on prediction points

Figure 4 plots the prediction points. As depicted, the application of CAMs consistently mitigated false positive errors, particularly for the Purple shale and Mafic B classes. Moreover, the outcomes from various CAMs exhibited remarkable similarity, indicating that the selection of CAMs may not significantly influence our task. The key insight lies in the utilization of CAMs to enhance channel-wise feature representation.

4.3 Visualization of the channel attention features

We extend our analysis of the learned channel attention features by visualizing them. We took a sample of points belonging to Purple shale class. The output of CAMs are 128-dimensional feature vectors. In Figure 5, the learned features across different layers are depicted. It is evident that CAMs highlight specific channels of the learned features, with a particularly strong emphasis on certain channels in the last layer.

5. Conclusion

In summary, our study introduces CAMs tailored for hyperspectral data analysis. Given the assumption that distinct spectra within hyperspectral data correspond to different mineral classes, leveraging channel-wise features becomes imperative. Our investigation into various CAMs revealed their overall effectiveness in our task. However, none emerged as significantly superior, highlighting the need for further exploration and development of specifically designed CAMs for hyperspectral data in future research.

References

Acosta, I. C. C., Khodadadzadeh, M., Tolosana-Delgado, R., Gloaguen, R., 2020. Drill-Core Hyperspectral and Geochemical Data Integration in a Superpixel-Based Machine Learning Framework. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 4214-4228.

Acosta, I. C. C., Khodadadzadeh, M., Tusa, L., Ghamisi, P., Gloaguen, R., 2019. A Machine Learning Framework for Drill-Core Mineral Mapping Using Hyperspectral and High-Resolution Mineralogical Data Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(12), 4829-4842.

Afifi, A. J., Thiele, S. T., Rizaldy, A., Lorenz, S., Ghamisi, P., Tolosana-Delgado, R., Kirsch, M., Gloaguen, R., Heizmann, M., 2024. Tinto: Multisensor Benchmark for 3-D Hyperspectral Point Cloud Segmentation in the Geosciences. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-15.

Decker, K. T., Borghetti, B. J., 2022. Composite Style Pixel and Point Convolution-Based Deep Fusion Neural Network Architecture for the Semantic Segmentation of Hyperspectral and Lidar Data. *Remote Sensing*, 14(9).

Decker, K. T., Borghetti, B. J., 2023. Hyperspectral Point Cloud Projection for the Semantic Segmentation of Multimodal Hyperspectral and Lidar Data with Point Convolution-Based Deep Fusion Neural Networks. *Applied Sciences*, 13(14).

Francky Fouedjio, E. J. H., Laukamp, C., 2018. Geostatistical clustering as an aid for ore body domaining: case study at the Rocklea Dome channel iron ore deposit, Western Australia. *Applied Earth Science*, 127(1), 15-29.

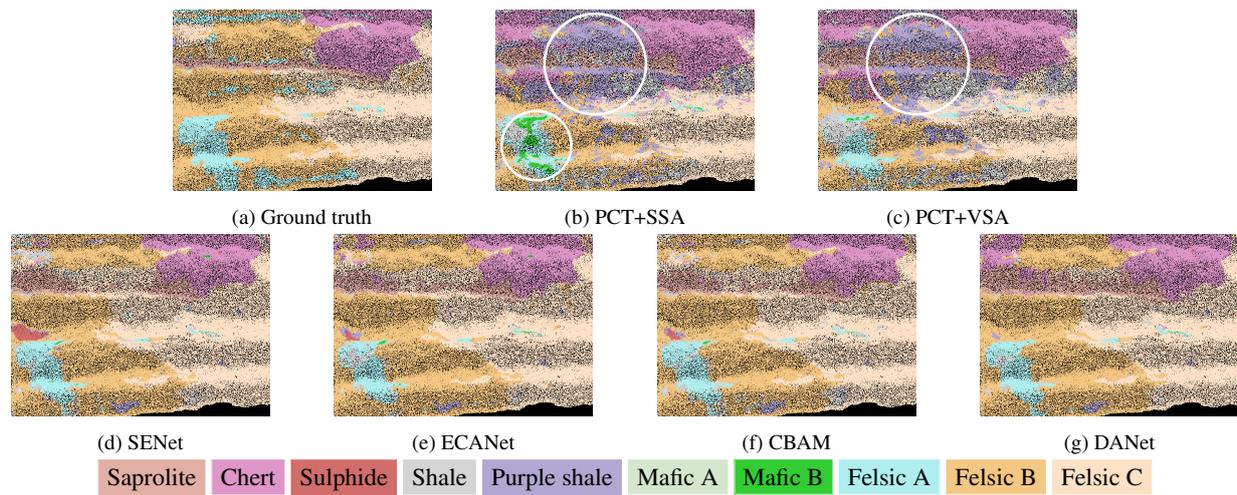


Figure 4. Visual comparison between different channel attention modules and the baselines.

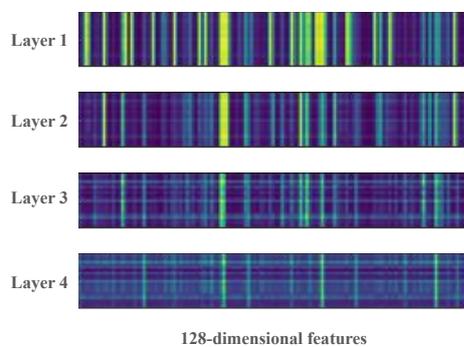


Figure 5. Visualization of channel attention features.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3141–3149.

Ghamisi, P., Li, H., Jackisch, R., Rasti, B., Gloaguen, R., 2020. Remote sensing and deep learning for sustainable mining. *IG-ARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 3739–3742.

Ghamisi, P., Plaza, J., Chen, Y., Li, J., Plaza, A. J., 2017. Advanced Spectral Classifiers for Hyperspectral Images: A review. *IEEE Geoscience and Remote Sensing Magazine*, 5(1), 8-32.

Ghamisi, P., Shahi, K. R., Duan, P., Rasti, B., Lorenz, S., Booyens, R., Thiele, S., Contreras, I. C., Kirsch, M., Gloaguen, R., 2021. The Potential of Machine Learning for a More Responsible Sourcing of Critical Raw Materials. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 8971-8988.

Guo, J., Liu, Y., Wu, L., Liu, S., Yang, T., Zhu, W., Zhang, Z., 2019. A geometry- and texture-based automatic discontinuity trace extraction method for rock mass point cloud. *International Journal of Rock Mechanics and Mining Sciences*, 124, 104132.

Guo, M.-H., Cai, J.-X., Liu, Z.-N., Mu, T.-J., Martin, R. R., Hu, S.-M., 2021. Pct: Point cloud transformer. *Computational Visual Media*, 7, 187–199.

Guo, Y., Gao, J., Li, F., 2014. Spatial subspace clustering for drill hole spectral data. *Journal of Applied Remote Sensing*, 8(1), 083644. <https://doi.org/10.1117/1.JRS.8.083644>.

Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2020. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), 2011-2023.

Lin, Y., Vosselman, G., Cao, Y., Yang, M. Y., 2021. Local and global encoder network for semantic segmentation of Airborne laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 176, 151-168.

Lorenz, S., Ghamisi, P., Kirsch, M., Jackisch, R., Rasti, B., Gloaguen, R., 2021. Feature extraction for hyperspectral mineral domain mapping: A test of conventional and innovative methods. *Remote Sensing of Environment*, 252, 112129.

Lorenz, S., Salehi, S., Kirsch, M., Zimmermann, R., Unger, G., Vest Sørensen, E., Gloaguen, R., 2018. Radiometric Correction and 3D Integration of Long-Range Ground-Based Hyperspectral Imagery for Mineral Exploration of Vertical Outcrops. *Remote Sensing*, 10(2).

Lorenz, S., Thiele, S. T., Kirsch, M., Unger, G., Zimmermann, R., Guarnieri, P., Baker, N., Sørensen, E. V., Rosa, D., Gloaguen, R., 2022. Three-Dimensional, Km-Scale Hyperspectral Data of Well-Exposed Zn-Pb Mineralization at Black Angel Mountain, Greenland. *Data*, 7(8).

Mitschke, I., Wiemann, T., Igelbrink, F., Hertzberg, J., 2023. Hyperspectral 3d point cloud segmentation using randlanet. I. Petrovic, E. Menegatti, I. Marković (eds), *Intelligent Autonomous Systems 17*, Springer Nature Switzerland, Cham, 301–312.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.

Rizaldy, A., Afifi, A. J., Ghamisi, P., Gloaguen, R., 2023. Transformer-based models for hyperspectral point clouds segmentation. *2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–5.

Shahi, K. R., Ghamisi, P., Jackisch, R., Rasti, B., Scheunders, P., Gloaguen, R., 2021. A multi-sensor subspace-based clustering algorithm using rgb and hyperspectral data. *2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–5.

Shahi, K. R., Ghamisi, P., Rasti, B., Jackisch, R., Scheunders, P., Gloaguen, R., 2020a. Data Fusion Using a Multi-Sensor Sparse-Based Clustering Algorithm. *Remote Sensing*, 12(23).

Shahi, K. R., Khodadadzadeh, M., Tolosana-delgado, R., Tusa, L., Gloaguen, R., 2019. The application of subspace clustering algorithms in drill-core hyperspectral domaining. *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–5.

Shahi, K. R., Khodadadzadeh, M., Tusa, L., Ghamisi, P., Tolosana-Delgado, R., Gloaguen, R., 2020b. Hierarchical Sparse Subspace Clustering (HESSC): An Automatic Approach for Hyperspectral Image Analysis. *Remote Sensing*, 12(15).

Thiele, S. T., Lorenz, S., Kirsch, M., Cecilia Contreras Acosta, I., Tusa, L., Herrmann, E., Möckel, R., Gloaguen, R., 2021. Multi-scale, multi-sensor data integration for automated 3-D geological mapping. *Ore Geology Reviews*, 136, 104252.

Walton, G., Mills, G., Fotopoulos, G., Radovanovic, R., Stancliffe, R., 2016. An approach for automated lithological classification of point clouds. *Geosphere*, 12(6), 1833-1841. <https://doi.org/10.1130/GES01326.1>.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020. Eca-net: Efficient channel attention for deep convolutional neural networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11531–11539.

Weidner, L., Walton, G., Krajinovich, A., 2021. Classifying rock slope materials in photogrammetric point clouds using robust color and geometric features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 176, 15-29.

Woo, S., Park, J., Lee, J.-Y., Kweon, I. S., 2018. Cbam: Convolutional block attention module. *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII*, Springer-Verlag, Berlin, Heidelberg, 3–19.

Zhao, H., Jia, J., Koltun, V., 2020. Exploring self-attention for image recognition. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10073–10082.

Zhao, H., Jiang, L., Jia, J., Torr, P. H., Koltun, V., 2021. Point transformer. *Proceedings of the IEEE/CVF international conference on computer vision*, 16259–16268.