Improved Mineral Detection via Spectral Attention U-Net with a Novel Hapke Layer: Accelerating Discovery of Critical Minerals for the Green Transition

Ryan Wang
Stanford University
Math and Computer Science
wangrj@stanford.edu

Chandra Suda
Stanford University
Math and Computer Science
csuda@stanford.edu

Abstract

The transition to green energy demands efficient discovery of critical minerals, yet traditional prospection methods remain costly and time-intensive. We present a novel deep learning approach for automated mineral detection from hyperspectral remote sensing data that accelerates this discovery process. Our method introduces two key innovations: (1) a spectral attention mechanism using squeeze-and-excitation blocks to learn inter-band dependencies in hyperspectral data, and (2) the first differentiable Hapke layer that embeds radiative transfer physics directly within a neural segmentation network. Unlike prior work that uses Hapke theory only for offline data augmentation, our physics-infused layer jointly optimizes mineral-specific scattering parameters with spatial features. We evaluate our approach on the Tinto dataset, achieving significant improvements over baseline methods. Our Hapke-enhanced U-Net attains 0.7811, 0.8014, and 0.7275 mIoU on LWIR, SWIR, and VNIR data respectively, representing on average a 4 percent gain over standard architectures. Additionally, we demonstrate spectral masked autoencoders for leveraging unlabeled hyperspectral data. This work establishes that coupling physics-based constraints with deep learning can substantially improve mineral segmentation accuracy, offering a scalable solution for accelerating critical mineral discovery essential to the green energy transition.

1. Introduction

As technology advances, the importance of efficiency and cost-effectiveness in mineral prospection cannot be overstated. Minerals such as rare earths are highly limited in supply, yet strictly essential for manufacturing high-tech products that have become indisposable in various applications. For example, minerals used in batteries are essential for the development of green energy, a crucial tool in the fight against climate change. Current methods of prospecting for new mineral deposits are tedious and costly, involving extensive geological surveys, drilling, and laboratory analysis. [5] first demonstrated the promise of a data-driven approach by introducing hyperspectral remote sensing in 1985. Hyperspectral images consist of hundreds of contiguous spectral bands for each ground pixel instead of the three bands for RGB images, ranging from the visible spectrum to NIR (near-infrared), SWIR (short-wave-infrared), and LWIR (long-wave-infrared). Moreover, hyperspectral data can easily be captured from airborne and spacecraft sensors. [14] showed that remote sensing data is immensely valuable in mineral prospection, and has already been used for the mapping of many minerals, such as certain types of clay, sulfate, and carbonate.

Our project uses various segmentation methods to automate the process of mineral prospection from hyperspectral images. Our input is a hyperspectral cube of shape H x W x B, where B is the number of bands, and we will output a label map of shape H x W that contains the geological class label for each pixel. Our main dataset is the Tinto dataset[1], a large hyperspectral scene with field-verified ground truth labels for geological classes. Our baseline models are an MLP and a U-Net, classic deep learning models used in segmentation. On top of these baselines, we experiment with several architectural modifications, including a spectral attention layer that learns associations between hyperspectral bands and a physics-infused layer based on the Hapke equations. These models utilize the VNIR, LWIR, and SWIR bands from the Tinto dataset to predict pixel-level labels. In addition, we also experiment with spectral masked autoencoders, a pretraining framework that makes use of unlabeled data, to address the difficulty of obtaining geologically labeled data. We use the Cuprite dataset, a classic benchmark hyperspectral dataset, for pretraining, and finetune on the Tinto dataset.

Our hope is to show through these methods that deep learning, paired with the increasing availability of hyperspectral data, can be an effective solution for mineral prospection.

2. Related Work

2.1. U-Nets with Squeeze-and-Excitation Blocks

AeroRIT [9] benchmarks several CNN segmentation architectures, including SegNet, U-Net, and Res-U-Net, on AeroRIT, a large-scale airborne hyperspectral dataset. The authors found that deeper backbones perform better on AeroRIT, and propose the use of squeeze-and-excitation blocks in the encoder. Adding squeeze-and-excitation blocks resulted in a 1-2% increase in mIoU, demonstrating the viability of squeeze-and-excitation blocks for hyperspectral segmentation. These results were promising to us because the AeroRIT dataset is similar to our Tinto dataset in that it contains detailed pixel-level annotations of complex classes (AeroRIT classes include cars, roads, and buildings, rather than simple land-cover classes). For our spectral U-Net, we implement a similar architecture that makes use of squeeze-and-excitation blocks. FuSENet [12] extends standard squeeze-and-excitation by proposing the use of dual squeeze operations, where both global average pooling and global max pooling are used in the squeeze operation, rather than global average pooling alone.

2.2. Physics-Infused (Hapke) Methods

The Hapke model [2] is a radiative-transfer formulation that describes how incident light is scattered by a particulate surface. It introduces physical parameters such as the single-scattering albedo ω, the phase-function asymmetry g, grain-size–dependent attenuation d, and an opposition-surge term (B_0, h) that captures enhanced retro-reflection at small phase angles. Originally developed for lunar and planetary spectroscopy, Hapke's theory has become a standard tool for interpreting laboratory reflectance spectra of minerals and regoliths.

Existing hyperspectral-learning papers exploit Hapke *only* as an offline data-augmentation engine: spectra are synthetically perturbed under plausible illumination geometries to enlarge the training set. No prior work has embedded a differentiable Hapke module directly inside a neural segmentation network. We therefore propose the first **Hapke Layer**: a learnable, end-to-end component that injects the governing physics into the feature pipeline. By coupling mineral-specific Hapke parameters with abundance estimates produced by the network, we enable joint optimisation of radiative-transfer physics and spatial context—an approach that is novel in both remote-sensing and computer-vision literature.

2.3. Spectral Masked Autoencoders

Spectral-MAE [4] introduces a self-supervised pretraining framework for hyperspectral image classification. The training involves randomly masking a subset of bands in unlabeled hyperspectral images, enabling the encoder to learn features from unlabeled data. Once the encoder is trained using this framework, it can then be finetuned using a lightweight classifier head. This self-supervised approach is especially valuable for limited-data applications, making it perfect for mineral segmentation tasks. We attempted to implement a similar self-supervised training framework to demonstrate how it can be applied to mineral segmentation. LO-SST [8] builds on the idea from Spectral-MAE, reducing computational overhead by pruning the layers that are contributing the least through the use of learned importance scores.

3. Data

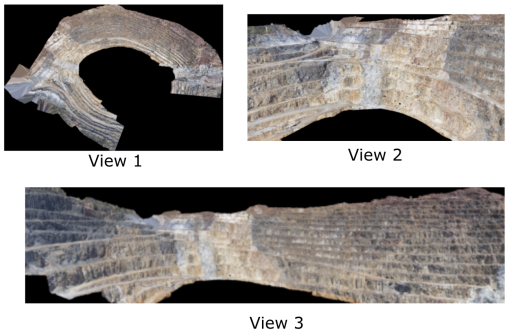

Figure 1. RGB images of the three views in the Tinto dataset. The three views are the same scene from different viewpoints.

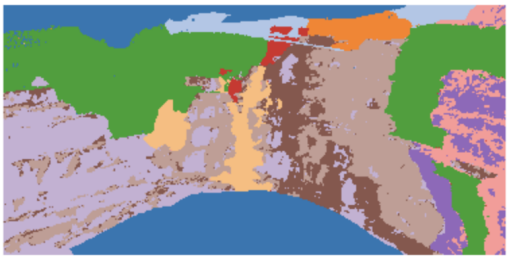

Figure 2. Ground-truth labels for view 2 of the Tinto dataset. In the label map, each color represents a different geological class.

Our main dataset is the Tinto dataset, a public dataset consisting of several 2D hyperspectral views of a real-world scene in Corta Atalaya, an open-pit mine in Andalusia, Spain. The dataset contains three views of the scene from different angles containing the same points. Our models are trained on the second view, a vertical landscape perspective, because it has the largest proportion of non-background pixels. Each pixel in the scene is labeled as vegetation or one of 10 geological classes, which include saprolite, chert, and sulphide. According to the original Tinto paper, the labels are laboratory-tested and expert-verified, and can be regarded as mostly reliable [1].

The dataset includes readings from LWIR, SWIR, and VNIR hyperspectral ranges, with 126, 141, and 51 bands

respectively. The second view, which we used to train our models, has three hyperspectral cubes of dimensions 512 x 1024 x B, where B = 126, 141, and 51 for LWIR, SWIR, and VNIR. The total number of pixels is 524,288. Due to the facts that mineral detection applications of deep learning are relatively new and obtaining labeled data is difficult, we were not able to find a larger public dataset even after extensive research. However, due to the wide range of models tested in the original Tinto paper [1] (on 3D point cloud data derived from the 2D rasters), including transformer-based models, we are confident that the dataset is sufficiently large for our models.

In addition to the Tinto dataset, we also utilize the Cuprite hyperspectral dataset for our self-supervised pretraining tests. This dataset consists of a single hyperspectral cube with 224 bands obtained by the NASA Jet Propulsion Laboratory, using SWIR and LWIR sensors to map out an area in Cuprite, Nevada. The dataset is unlabeled and contains 512 x 614 = 314,368 pixels at a resolution of 20 m per pixel.

For preprocessing, we apply the vegetation mask given in the dataset to exclude vegetation pixels from our data. We divide the data into an 80/10/10 split for train/val/test, keeping the ratios of each class the same. We also experiment with normalizing the data by spectral channel. In addition, we augment both datasets with random flips, rotations, and Gaussian jitter on sampled patches.

Class	Train	Val	Test	Total
2 (saprolite)	11,856	1,482	1,482	14,824
3 (chert)	8,164	1,020	1,021	10,206
4 (sulphide)	15,168	1,896	1,896	18,963
5 (shale)	92,348	11,543	11,544	115,634
6 (purple shale)	3,194	399	400	3,993
7 (MaficA)	19,524	2,440	2,441	24,405
8 (MaficB)	18,505	2,313	2,314	23,132
9 (FelsicA)	66,904	8,363	8,364	83,634
10 (FelsicB)	36,522	4,565	4,566	45,654
11 (FelsicC)	72,568	9,071	9,071	90,710

Table 1. Number of pixels per class in our training, validation, and test sets. Class 0 (vegetation) and class 1 (background) are removed from the train, val, and test sets during preprocessing.

4. Methods

For our baselines, we implement an MLP and a U-Net pixel classifier. In addition, we experiment with a novel Spectral U-Net architecture, which makes use of squeeze-and-excitation blocks to improve the representational power of our U-Net by modeling interdependencies between hyperspectral bands [6]. Finally, we also test out a self-supervised pretraining model that utilizes spectral masked autoencoders to derive hyperspectral image features. Since there is no standard number of bands for hyperspectral datasets, all of our models are trained from scratch.

4.1. Baselines: MLP and U-Net

Our first baseline is a vanilla MLP that takes in flattened data and outputs class scores for each pixel. We chose an MLP as our first baseline because it was one of the baseline models included in the original Tinto paper [1], indicating that it is a viable option for hyperspectral image segmentation.

Our second baseline is a U-Net, a classic architecture for semantic segmentation [11]. We chose it for its versatility in various applications and its proven effectiveness on limited data. As in the classic architecture, our baseline U-Net has a symmetric encoder-decoder architecture with skip connections. The convolutional and max-pooling layers in the encoder allow the model to learn image features using spatial context, which is crucial for predicting mineral classes. Skip connections preserve high-resolution details and feed them into the decoder, ensuring that these details are not thrown away by the model after downsampling. We utilize a 1x1 convolutional layer as the prediction head to generate pixel-level labels in our U-Net.

4.2. Spectral U-Net

We modify our baseline U-Net with spectral attention, a block specialized for hyperspectral data based on the "squeeze-and-excitation" mechanism proposed in [6]. Whereas regular convolutional layers treat all channels equally, squeeze-and-excitation blocks allow for channel recalibration by learning weights for each channel. They consist of two components: the "squeeze" and "excitation."

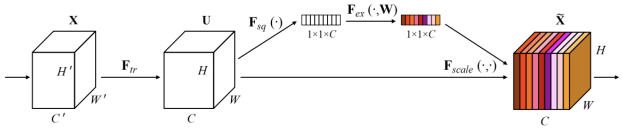

Figure 3. A squeeze-and-excitation block [6]

First, the "squeeze" layer takes in an H x W x C input u (usually from a convolutional layer), then applies global average pooling for each channel, producing a vector z of shape (C,), where

$$z_c = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{1}$$

for each class c. This operation effectively condenses the global "importance" of channel c into a single number z_c.

Next, z is fed into the "excitation" layer, which consists of two fully connected layers with a ReLU nonlinearity, and an elementwise sigmoid applied to the output. The first

fully connected layer reduces the input dimension C to a smaller "bottleneck" dimension, and the second layer expands it back to dimension C. Finally, the output from the excitation layer is a vector $e = \begin{bmatrix} s_1 & s_2 & \cdots & s_C \end{bmatrix}$ used as the "excitation weights," and each element $u_c(i, j)$ from channel C in the original input is multiplied by its corresponding excitation weight s_c.

Squeeze-and-excitation style blocks are lightweight and can be added to convolution-based models to improve performance with minimal overhead [3][6]. We experiment with adding squeeze-and-excitation blocks in our U-Net to improve its representational power in modeling associations between bands. These blocks are called "spectral attention" blocks because they are learning to "pay attention" to different spectral bands. We believe the use of spectral attention blocks will enable our model to learn the relative importance of different bands for predicting different geological classes.

4.3. Hapke Layer

We augment our U-Net with a physics-informed Hapke layer that explicitly models bidirectional reflectance by way of radiative-transfer theory. To our knowledge this is the first instance of the full Hapke formalism embedded as a differentiable module in any deep-learning architecture; earlier work used Hapke only to create synthetic spectra offline for data-augmentation. [2].

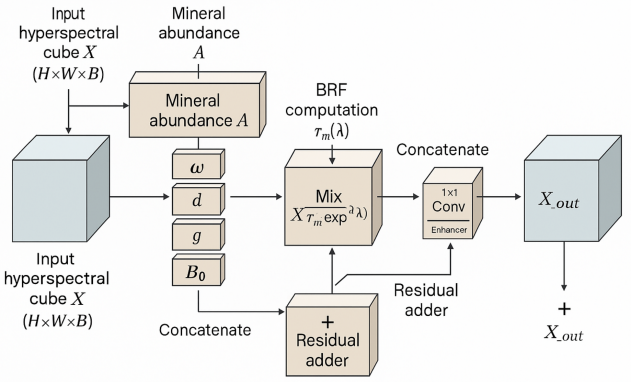

Figure 4. Our novel Hapke layer architecture

Learnable physical parameters. For each mineral class m the layer optimisms five sets of parameters:

- *Single-scattering albedo* $\omega_m \in \mathbb{R}^B$ — the fraction of incident photons scattered (rather than absorbed) at each wavelength;
- *Grain-size term* $d_m \geq 0$ — larger grains lengthen optical paths and deepen absorptions;
- *Phase-function asymmetry* $g_m \in [-1, 1]$ — positive values bias forward scattering, negative values bias backward scattering in the Henyey–Greenstein function;
- *Opposition-surge amplitude* $B_{0,m}$ and *width* h_m — together describe the non-linear brightening that occurs when illumination and viewing directions coincide.

Abundance estimation. A 1×1 convolution followed by a soft-max produces mineral abundances $A \in \mathbb{R}^{H\times W\times M}$ with $\sum_m A_m(i, j) = 1$. Each $A_m(i, j)$ can be interpreted as the probability that pixel (i, j) is dominated by mineral m; these weights will mix the mineral prototypes that the Hapke equations generate (full implementation details provided in Appendix 8).

Differentiable bidirectional reflectance factor. For mineral m the bidirectional reflectance factor (BRF) is

$$r_m(\lambda) = \frac{\omega_m(\lambda)}{4\pi(\mu_0 + \mu)} \Big[p(g_m)\big(1 + B(g_m)\big) + H(\mu_0, \omega_m)\, H(\mu, \omega_m) - 1 \Big], \tag{2}$$

where $\mu_0 = \cos i$ and $\mu = \cos e$ are the incidence and emergence cosines. The term $p(g_m)$ is the single-parameter Henyey–Greenstein phase function, $B(g_m)$ models the opposition peak, and H is the Chandrasekhar H-function that approximates multiple scattering. Closed-form approximations are used so that gradients propagate. (full formulas and derivations provided in Appendix 8).

Physics mixing and residual fuse. Mineral contributions combine linearly—weighted by abundances and attenuated by grain size—into

$$X_{\text{hapke}}(i, j, \lambda) = \sum_{m=1}^{M} A_m(i, j)\, X(i, j, \lambda)\, r_m(\lambda)\, \mathrm{e}^{-d_m \lambda}. \tag{3}$$

The tensor $[X, X_{\text{hapke}}]$ is compressed by a 1×1 enhancement conv and merged residually:

$$X_{\text{out}} = X + \alpha\, f_{\text{enh}}\big([X, X_{\text{hapke}}]\big),$$

Interaction with the Spectral-Attention U-Net. Placing the Hapke layer *before* the encoder ensures that every convolution and squeeze-and-excitation block operates on spectra already regularized by physical law. Spatial filters therefore learn *where* minerals change while the Hapke layer constrains *how* spectra may vary, yielding complementary supervision. This coupling helped us improve generalization whenever illuminations or grain sizes shift between train and test scenes, a common scenario in airborne surveys of geological areas.

Mathematical intuition. Equation (3) resembles a soft dictionary lookup: abundances A_m pick mineral prototypes

while the BRF r_m warps them according to viewing geometry and multiple scattering. Because ω, d, g, B_0, and h remain trainable, the network can refine prototypes to sensor-specific calibration yet is discouraged from drifting into non-physical regions of spectrum space. Gradients through A_m sharpen class assignment; gradients through the Hapke parameters adapt the prototypes themselves. The small gating factor α keeps the residual numerically stable during early epochs.

Geological intuition. Constraining the Hapke parameters to physically plausible ranges ($0 \leq \omega \leq 1$, $|g| \leq 1$) steers the network toward mineral-realistic spectra and away from over-fitting to noise in narrow bands. Subtle absorption shoulders—such as those that distinguish saprolite from shale or separate felsic subclasses—are preserved because the Hapke formulation enforces energy balance and scattering symmetry across the full band shape rather than allowing the model to exploit spurious pixel-level artefacts. Thus, projecting the raw cube onto this "Hapke manifold" sharpens class boundaries in spectrally ambiguous regions and improves cross-scene transfer. On the Tinto benchmark we obtain a reproducible 4–5pt gain in mean-IoU with few additional parameters, confirming that radiative-transfer priors complement attention-based hyperspectral segmentation.

4.4. Segmentation Loss Functions

Our MLP baseline utilizes cross-entropy loss, which is given by the equation

$$L = -\sum_{i=1}^{N} \log\left(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}}\right), \tag{4}$$

where N is the number of pixels, s_{y_i} is the score of the correct class for pixel i, and s_j is the score for the jth class.

Our U-Net models, including the U-Nets modified with spectral attention and the Hapke layer, use Dice loss [7], defined in terms of the Dice coefficient for each class c,

$$\text{Dice}_c = \frac{2TP_c}{2TP_c + FP_c + FN_c}, \tag{5}$$

where TP_c is the number of true positives (pixels correctly predicted as class c), FP_c is the number of false positives (pixels incorrectly predicted as class c), and FN_c is the number of false negatives (pixels labeled class c that were not correctly predicted). Dice loss is then calculated as

$$L = 1 - \frac{1}{C}\sum_{c=1}^{C} \text{Dice}_c. \tag{6}$$

Dice loss seeks to maximize Dice coefficients by maximizing TP_c and minimizing FP_c and FN_c for each class. It aims to increase the overlap between predicted and ground-truth masks for each class while decreasing the rate of false positives and false negatives. We chose Dice loss over cross-entropy loss for our deeper models to minimize false positives and false negatives while addressing class imbalances in our data [7]. Also, Dice loss is closely related to the IoU accuracy metric that we use for evaluation, optimizing for overlap between predicted and ground-truth labels instead of per-pixel accuracy.

4.5. Model Architectures

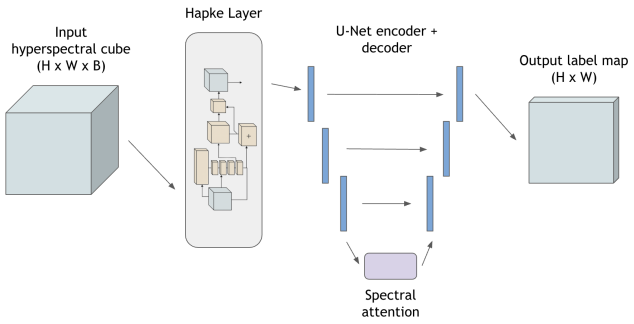

Figure 5. Overall architecture of the Hapke U-Net, including the Hapke layer and the spectral attention block.

Our vanilla U-Net is built from a typical 3-layer U-Net architecture. It first applies a 1×1 convolution layer to project the input to have the desired number of bands, then passes the projected input through an input convolution block, a DoubleConv layer (two convolution layers with batchnorm and ReLU), to extract initial features. Next, the features are passed through the encoder, consisting of three downsampling blocks, each applying a max pooling and a DoubleConv layer. Finally, the encoded features are processed by a decoder consisting of three symmetric upsampling blocks. The decoded features are then fed into a 1×1 convolution layer to predict class labels for each pixel. Due to the limited size of our dataset, we decided not to use a deeper network so that we could mitigate overfitting.

Our spectral U-Net uses a similar backbone as the vanilla U-Net, with the same projection, input convolution, downsampling, and upsampling layers. In addition, it uses a "squeeze-and-excitation"–style spectral attention block at the bottleneck to recalibrate channel-wise features. In order to prevent overfitting and make our model easier to train, we decided to add only a single spectral attention block. Inserting a squeeze-and-excitation block after a DoubleConv layer is consistent with [6], and putting it in the final encoder layer allows it to recalibrate the deepest, most meaningful features.

Our Hapke U-Net extends the spectral U-Net by inserting a lightweight Hapke layer immediately after the initial 1×1 projection. The Hapke layer takes the raw hyperspectral cube $X \in \mathbb{R}^{H\times W\times B}$ as its input and outputs a physics-regularized cube $X_{\text{out}} \in \mathbb{R}^{H\times W\times B}$ via pixel-wise

abundance estimation, per-mineral reflectance computation, grain-size attenuation, and a gated residual fusion. Specifically, a 1×1 convolution + softmax produces abundances $A \in \mathbb{R}^{H\times W\times M}$, the Hapke reflectance formulas compute X_{hapke}, and a small 1×1 enhancement conv merges $[X, X_{\text{hapke}}]$ before adding back to X. The resulting X_{out} replaces the original input to the first DoubleConv block of the encoder. Beyond this insertion, the downsampling, upsampling, and squeeze-and-excitation attention blocks remain unchanged, enabling spatial filters to operate on spectra already constrained by radiative-transfer physics without adding significant parameter overhead.

4.6. Spectral Masked Autoencoders

To address the problem of data scarcity in mineral-related machine learning tasks, we experiment with spectral masked autoencoders, a self-supervised training framework for hyperspectral images proposed in [4], similar to masked autoencoders for normal images. Our algorithm samples 3D spatial patches from the dataset (with all hyperspectral bands), randomly masking out a subset of bands in the patch and training the encoder to reconstruct masked bands. We use our spectral U-Net as an encoder, using mean-squared error (MSE) to train it to reconstruct masked bands. MSE loss is given by the equation

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{7}$$

where N is the number of masked voxels in a 3D patch (spatial dimensions and hyperspectral bands), y_i is the true value of a masked voxel, and $\hat{y}_i$ is the reconstructed value of the voxel.

By training the encoder to perform this band reconstruction task, we aim to produce features that incorporate both spatial and spectral context. Predicting missing bands encourages the model to learn associations between spectral bands, as well as associations between pixels that are spatially close to each other. We pretrained our spectral U-Net encoder on the Cuprite dataset, before finetuning a lightweight 1x1 convolution layer as a classification head to assess segmentation performance on the Tinto dataset.

5. Experiments

5.1. Evaluation Methods

To evaluate our models, we use the intersection over union (IoU) and mean intersection over union (mIoU) metrics, which are standard metrics for semantic segmentation [10]. For each class c, IoU is given by the equation

$$\text{IoU}_c = \frac{TP_c}{TP_c + FP_c + FN_c}, \tag{8}$$

where TP_c, FP_c and FN_c are defined as in equation (2). The numerator is the number of predicted pixels for class c, while the denominator is the number of pixels in the union of the predicted and ground-truth masks for class c. IoU score ranges from 0 to 1, with a higher score indicating more overlap between the predicted and ground-truth masks for a class.

The mIoU is the average of the IoUs for each class, calculated as

$$\text{mIoU} = \frac{1}{C}\sum_{c=1}^{C}\text{IoU}_c. \tag{9}$$

We chose mIoU as our main metric for overall accuracy over simple pixel-wise accuracy to handle class imbalances and to detect false positives and false negatives [10]. Since the proportions of each class in our data differ significantly, using pixel-wise accuracy would inflate performance while allowing models to neglect classes with fewer examples.

Additionally, we also keep track of IoU scores for each class to check whether our models are better at predicting certain classes than others. Taking both mIoU and IoU per class into account gives us a general idea of the holistic performance of each model as well as specific strengths and weaknesses.

5.2. Training

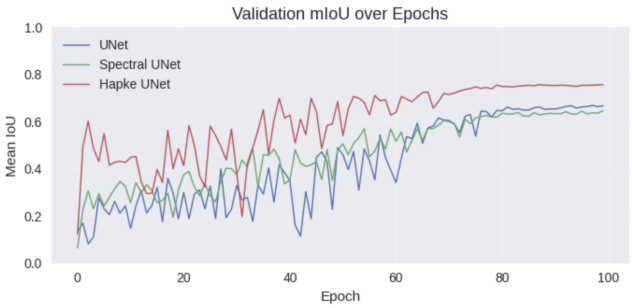

Figure 6. Validation mIoU over 100 epochs for U-Net based models trained on LWIR data.

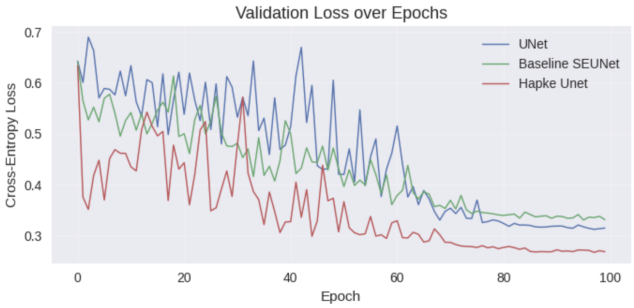

Figure 7. Validation loss over 100 epochs for U-Net based models trained on LWIR data.

We trained an MLP, a vanilla U-Net, a spectral U-Net (with a squeeze-and-excitation "spectral attention" layer),

and a spectral U-Net with an additional Hapke layer. In addition, we pretrained a spectral U-Net using masked spectral autoencoders. All supervised models were trained on the view 2 Tinto data, which consist of 512 x 1024 x B hyperspectral cubes, where B = 126, 141, and 51 for LWIR, SWIR, and VNIR respectively. We used the Adam optimizer with a learning rate of 0.001, and train all U-Net based models for 100 epochs. To prevent overfitting, we use L2 regularization, and only train the MLP for 10 epochs. For U-Net based models, we use a cosine annealing learning rate scheduler over 100 epochs reaching a minimum learning rate of 1e-6, which was empirically found to offer the most stable convergence.

For self-supervised pretraining using spectral masked autoencoders, we use our spectral U-Net as an encoder, and train it to reconstruct masked bands in the Cuprite dataset. The learned weights were then transferred to a segmentation model, which was then finetuned on the view 2 Tinto data.

To train all of our models, we sample batches of 64 pixel x 64 pixel patches from our dataset, where the center of each patch is in the train split. Using patches allows the model to incorporate spatial context into its pixel-wise predictions, while reducing the time needed for training. For validation and testing, we similarly sample patches centered at pixels in the validation and test splits.

5.3. Results

Model	LWIR	SWIR	VNIR
MLP	0.2947	0.5583	0.3428
U-Net	0.6899	0.7595	**0.7432**
Spectral U-Net	0.6648	0.7518	0.7053
Hapke U-Net	**0.7811**	**0.8014**	0.7275

Table 2. mIoU for all models trained on LWIR, SWIR, and VNIR view 2 Tinto data.

Table 2 gives results for our MLP baseline, and the three U-Net based models. All U-Net based models perform significantly better than the MLP baseline, illustrating the importance of spatial context in mineral segmentation. The Hapke U-Net outperformed other U-Net models by large margins on the LWIR and SWIR data, indicating that the physics-informed context provided by the Hapke layer is useful in predicting geological class. The spectral U-Net performed worse than we expected, producing a lower mIoU than even the vanilla U-Net on all three hyperspectral ranges. However, looking at the validation mIoU and loss curves in Figure 6 and Figure 7, we see that the spectral U-Net has slightly faster convergence than the vanilla U-Net, which may indicate the ability of the spectral attention block to help the model learn associations between bands.

Notably, the vanilla U-Net outperformed both enhanced U-Net models on the VNIR data, which we suspect is due to the relatively small size of the VNIR dataset compared to LWIR and SWIR (only 51 bands vs. 126 and 141). The additional complexity of the enhanced models likely means that these models require a larger amount of training data, which is why they perform better on the data with more bands.

For our self-supervised learning framework, we did not observe any nonnegligible increases in performance after finetuning our pretrained model, so results are omitted here. We believe that one of the main factors behind the failure of our pretrained weights to generalize was the numerous disparities between our pretraining dataset, Cuprite, and the Tinto dataset. Differences in sensor angle, scale (Cuprite covers a much larger area), and units used to store data were all obstacles when we were implementing our framework, which mostly likely made it difficult for pretrained weights to generalize. In a future project, we would like to attempt this pretraining framework with two datasets that are more similar to these respects.

5.4. Discussion

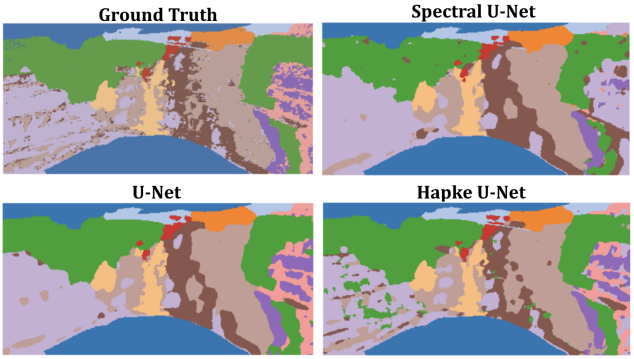

Figure 8. Labels predicted by our models trained on LWIR data vs. ground-truth labels for view 2 of the Tinto dataset.

More detailed IoU per class for each of our models is given in Table 3 below. These results reveal that the IoU per class often varies for different geological classes. For example, while the mIoU of the spectral U-Net is lower than that of the vanilla U-Net for LWIR and SWIR data, it offers an increase in IoU for classes 4 and 10. Further investigation is required to determine the causes of these variations, but we suspect they relate to the properties of the geological classes in our dataset and the properties of the hyperspectral ranges included in our data.

One instance of this phenomenon we investigated is the clear failure of the Spectral U-Net trained on LWIR data to detect class 7 (represented as pink in Figure 8 above). Class 7 is MaficA, a geological class of mafic volcanic lithologies rich in iron- and magnesium-bearing silicates [16]. In

	Model	2	3	4	5	6	7	8	9	10	11	mIoU
LWIR	U	0.7442	0.8517	0.7300	0.7778	0.6722	0.4111	0.6626	0.6711	0.4673	0.6044	0.6899
	S	0.7123	0.8012	0.7545	0.7207	0.6046	0.3631	0.6326	0.6551	0.5138	0.5578	0.6648
	H	**0.8793**	**0.9087**	**0.8078**	**0.8877**	**0.8372**	**0.5310**	**0.7429**	**0.7282**	**0.5922**	**0.6801**	**0.7811**
SWIR	U	0.8046	**0.9321**	0.6824	0.8597	0.7730	0.4127	0.7511	0.7740	0.6463	0.7364	0.7595
	S	0.7198	0.8903	0.6926	0.8462	0.8008	0.4334	0.7519	0.7703	0.6480	0.7323	0.7518
	H	**0.8248**	0.8911	**0.7941**	**0.9147**	**0.8347**	**0.5381**	**0.7814**	**0.8002**	**0.6789**	**0.7665**	**0.8014**
VNIR	U	0.7807	**0.8849**	**0.7749**	**0.8912**	0.8110	**0.4708**	**0.7100**	0.6943	0.5406	**0.6275**	**0.7432**
	S	0.7930	0.8133	0.6554	0.8615	0.8065	0.4537	0.6128	0.6966	0.4975	0.5743	0.7053
	H	**0.8142**	0.8751	0.6935	0.8660	**0.8145**	0.4498	0.6747	**0.6979**	**0.5427**	0.5814	0.7275

Table 3. IoU per class and mIoU for U-Net based models. "U" is the vanilla U-Net, "S" is the Spectral U-Net, and "H" is the Hapke U-Net. Column labels correspond to class labels in the Tinto dataset. LWIR, SWIR, VNIR have 126, 141, 51 bands respectively.

the LWIR range, this geological class exhibits diagnostic absorption features that are often subtle and easily confounded by illumination [15]. The Spectral U-Net relies on a squeeze-and-excitation block that pools spatial information into global channel weights, which consistently dilutes MaficA's subtle LWIR troughs when pixels are mixed or under differing angles, causing misclassification. This shows why the Spectral U-Net performed poorly on class 7.

In contrast, the vanilla U-Net doesn't have the squeeze-and-excitation block and the Hapke U-Net incorporates a physics-informed layer that enforces per-mineral bidirectional reflectance constraints. Those factors enables these models to preserve and sharpen MaficA's LWIR signatures, so they detect the pink regions well.

Another noticeable trend in Figure 8, and in the predicted labels for models trained on SWIR and VNIR data (see Appendices), is the difference in the shapes of class boundaries. We see that the class boundaries predicted by the U-Net are more regular and rounded compare to the more complicated models, especially the Hapke U-Net. This would make it difficult for the vanilla U-Net to capture nuances in more complicated data, although in our dataset it does not seem to penalize the mIoU too harshly. The ground-truth labels in the Tinto dataset have relatively simple class boundaries, but in a future extension of our project we would be interested in comparing the vanilla U-Net to our other U-Net models on a dataset with more complex class boundaries. In contrast to the vanilla U-Net, the Hapke U-Net produces much more nuanced class boundaries, which would most likely give it better performance on a dataset with more fine-grained class boundaries. As seen in the left half of the Hapke U-Net predicted label map in Figure 8, these nuanced boundaries closely match the ground truth labels and are completely overlooked by the vanilla U-Net and the spectral U-Net. This provides strong evidence that our Hapke layer provides valuable context for geological class prediction.

6. Conclusion

Deep learning on hyperspectral images holds significant promise in mineral prospection. The large quantity of information encoded in hyperspectral images holds key insights into mineral mapping, and segmentation is uses hyperspectral data to produce accurate mineral maps at minimal cost.

We built several segmentation models to produce pixel-wise labels on the Tinto dataset based on LWIR (long-wave infrared), SWIR (short-wave infrared), and VNIR (visible and near-infrared) hyperspectral data. In addition to an MLP, we implemented a U-Net baseline that provided the architectural backbone for several enhancements, including a spectral U-Net with a squeeze-and-excitation style spectral attention block, a physics-informed Hapke layer, and an attempt at spectral masked autoencoders, a self-supervised pretraining framework. The Hapke U-Net significantly outperformed the other U-Net models, providing strong evidence that incorporating a layer a models bidirectional reflectance improves segmentation of minerals.

In a possible extension of our project, we would like to investigate and compare our models on a hyperspectral dataset that is larger and has higher complexity in class boundaries. We believe this would more clearly show the strengths and weaknesses of our models and point towards possible paths to improving them. In addition, we would like to continue experimenting with a self-supervised framework to train a mineral segmentation model, as the scarcity of data remains a key obstacle in the deployment of deep learning in mineral prospection.

7. Contributions & Acknowledgements

Ryan worked on data processing functions, MLP baseline, vanilla U-Net, spectral masked autoencoders, literature review, training, and producing visualizations. Chandra worked on literature review, spectral attention U-Net, creation of the Hapke layer, training, and performance met-

rics.

We would like to thank our mentor Iris Xia for her overall guidance and help. We also reached out and had a meeting with Stanford Mineral-X [13] to learn more about the mineral detection industry. We would like thank them for the information they provided on their current work towards creating a resilient & sustainable critical mineral supply chain.

8. Appendices

8.1. Chandrasekhar H-Function

The Chandrasekhar H-function appears in the Hapke model as an approximation of multiple scattering. We used the following common analytic approximation:

$$H(\mu, \omega) \;=\; \frac{1 + 2\,\mu}{1 + 2\,\mu\sqrt{1 - \omega + \varepsilon}} \tag{10}$$

where:

- μ is the cosine of the viewing or illumination angle (i.e. $\mu = \cos e$ or $\cos i$).
- ω is the single-scattering albedo at that band: $0 \leq \omega \leq 1$.
- ε is a small positive constant (e.g. $\varepsilon = 10^{-6}$) to avoid division by zero.

In more exact form, the true Chandrasekhar H-function is defined via the integral equation

$$H(\mu, \omega) \;=\; \exp\left(-\frac{\omega}{2\mu}\int_0^1 H(\mu', \omega)\, G(\mu, \mu')\, d\mu'\right), \tag{11}$$

but we used the above rational approximation for computational efficiency and differentiability.

8.2. Henyey Greenstein Phase Function $p(g)$

The Henyey–Greenstein phase function models the angular scattering distribution as:

$$p(g, \theta) \;=\; \frac{1 - g^2}{\left(1 + 2\,g\cos\theta + g^2\right)^{3/2}}, \qquad g \in [-1, 1], \tag{12}$$

where:

- g is the asymmetry parameter: $g > 0$ biases forward scattering, $g < 0$ biases backward scattering.
- θ is the scattering angle between incident and emergent directions.

In our implementation, we use the single-term Henyey–Greenstein approximation, evaluated at the phase-angle $\theta = 0$, yielding:

$$p(g) \;=\; \frac{1 - g^2}{(1 + 2\,g + g^2)^{3/2}}.$$

8.3. Backscatter Function $B(g)$

The opposition-effect (backscatter) function is given by:

$$B(g) \;=\; \frac{B_0}{1 + \frac{1}{h}\,\tan\frac{\theta}{2}}, \tag{13}$$

where:

- B_0 is the peak amplitude at exact opposition (phase angle approaching zero).
- h is the half-width parameter controlling how rapidly the surge decays with increasing phase angle.
- θ is the phase angle ($\theta = i + e$ for narrow-angle approximations).

Since most airborne imagery uses a small phase-angle, we approximated it as $\tan(\theta/2) \approx \theta/2$ in radians.

8.4. Grain-Size Attenuation

Grain-size modulation enters the Hapke mixture as an exponential attenuation:

$$\mathrm{e}^{-d_m \lambda}, \tag{14}$$

where:

- $d_m \geq 0$ is a learnable grain-size parameter for mineral m. Larger d_m yields stronger attenuation (deeper absorption features).
- λ is the normalized wavelength (e.g. $\lambda \in [0, 1]$ after rescaling).

8.5. Mineral Abundance $A_m(i, j)$

The pixel-wise mineral abundances $A_m(i, j)$ are estimated by a 1×1 convolution followed by a soft-max over the M minerals:

$$A_m(i, j) \;=\; \frac{\exp\big(z_m(i, j)\big)}{\sum_{n=1}^{M} \exp\big(z_n(i, j)\big)},$$

where $z_m(i, j)$ is the raw score (logit) for mineral m at pixel (i, j). This enforces $\sum_m A_m(i, j) = 1$, making $A_m(i, j)$ interpretable as mixing fractions.

8.6. Additional Validation mIoU and Loss Curves

References

[1] A. J. A. et al. Tinto: Multisensor benchmark for 3-d hyperspectral point cloud segmentation in the geosciences. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. 1, 2, 3

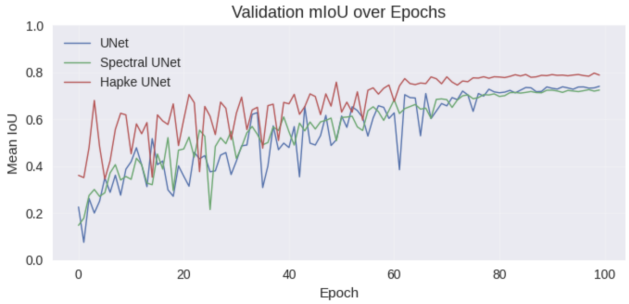

Figure 9. Validation mIoU over 100 epochs for U-Net based models trained on SWIR data.

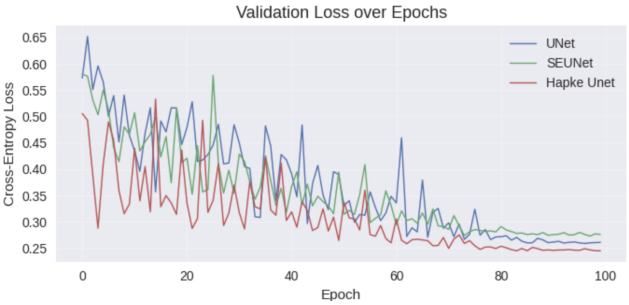

Figure 10. Validation loss over 100 epochs for U-Net based models trained on SWIR data.

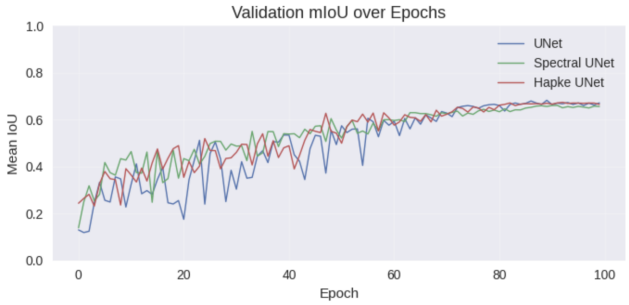

Figure 11. Validation mIoU over 100 epochs for U-Net based models trained on VNIR data.

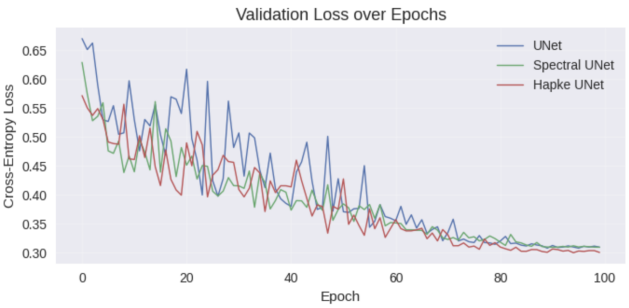

Figure 12. Validation loss over 100 epochs for U-Net based models trained on VNIR data.

[2] K. Q. et al. Hapke data augmentation for deep learning-based hyperspectral data analysis with limited samples. *IEEE Geoscience and Remote Sensing Letters*, 18(5):886–890, 2020. 2, 4

[3] O. O. et al. Attention u-net: Learning where to look for the pancreas, 2018. 4

[4] P. Feng, K. Wang, J. Guan, and G. He. Spectral masked autoencoder for few-shot hyperspectral image classification, 2023. 2, 6

[5] A. F. Goetz, G. Vane, J. E. Solomon, and B. N. Rock. Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147–1153, 1985. 1

[6] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 3, 4, 5

[7] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016. 5

[8] A. Perez and S. Prasad. Layer optimized spatial spectral masked autoencoder for semantic segmentation of hyperspectral imagery, 2025. 2

[9] A. Rangnekar, N. Mokashi, E. J. Lentilucci, C. Kanan, and M. J. Hoffman. Aerorit: A new scene for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11):8116–8124, 2020. 2

[10] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese. Generalized intersection over union: A metric and a loss for bounding box regression, 2019. 6

[11] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, 9351, 2015. 3

[12] S. Roy, S. R. Dubey, S. Chatterjee, and B. Chaudhuri. Fusenet: Fused squeeze-and-excitation network for spectral-spatial hyperspectral image classification. *IET Image Processing*, 14, 2020. 2

[13] Stanford University MineralX Lab. Mineralx: Hyperspectral imaging for mineral exploration, 2025. Accessed: 2025-06-04. 9

[14] F. D. van der Meer et al. Multi- and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*, 14(1):112–128, 2012. 1

[15] D. B. Williams and M. S. Ramsey. Infrared spectroscopy of volcanoes: from laboratory to orbital scale. *Frontiers in Earth Science*, 12:Article 1308103, 2024. Section: Volcanology; Published 24 January 2024. 8

[16] J. D. Winter. *An Introduction to Igneous and Metamorphic Petrology*. Prentice Hall, Essex, England, 2nd edition, 2014. See Chapter 5 (“Mafic and Ultramafic Rocks”), which describes mafic volcanic lithologies dominated by iron- and magnesium-bearing silicates such as pyroxene and olivine. 7